

Understanding "gardem bench" : studies on the perception of assimilated word forms

Citation for published version (APA):

Mitterer, H. A. (2003). *Understanding "gardem bench" : studies on the perception of assimilated word forms*. [Doctoral Thesis, Maastricht University]. Drukkerij Groeneveldt.
<https://doi.org/10.26481/dis.20031128hm>

Document status and date:

Published: 01/01/2003

DOI:

[10.26481/dis.20031128hm](https://doi.org/10.26481/dis.20031128hm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Understanding “gardem bench”

Studies on the perception of assimilated word forms

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de
Rector Magnificus, Prof. dr. A.C. Nieuwenhuizen Kruseman,
volgens het besluit
van het College van Decanen
in het openbaar te verdedigen

op vrijdag 28 november 2003 om 12.00

door Holger Mitterer
geboren te Hanau, Duitsland, 4 januari 1973

Promotor:

Prof. dr. Rainer Goebel

Co-promotor:

dr. Leo Blomert

Beoordelingscommissie:

Prof. dr. Jaap van Heerden (voorzitter)

Prof. dr. Anne Cutler (Katholieke Universiteit Nijmegen)

dr. Elia Formisano

dr. Bernhadette Jansma

Prof. dr. Sieb Nooteboom (Universiteit Utrecht)

ISBN 90-9017304-8

Copyright © 2003 by Holger Mitterer. All rights reserved.

Design: Udo Hanrieder, www.komfortbureau.de

Cover: Udo Hanrieder & Holger Mitterer

Printed by Drukkerij Groeneveldt BV te Landgraaf, The Netherlands

Acknowledgements

First of all I need to thank all the participants who took part in my experiments. Without their patience and co-operation in seemingly pointless experiments, and, for those poor bastards taking part in the electrophysiological experiments, their willingness to have someone put slimy electro-gel into their hair, this endeavor would be nowhere. A special thank you is in order for the Regionaal Instituut voor Dyslexie en de Maastrichtse basisscholen De Perroen, De Spiegel, de Basisscholen te Scharn en Wyck, en de Petrus en Paulus School for their co-operation.

I sincerely thank my supervisor Leo Blomert, who created many possibilities for me. The help of Csépe Valéria was crucial in order to get what I consider the most exciting part of this dissertation done in Budapest (and the possibility to skate in the historical city centre in Budapest was also greatly appreciated). I also thank Honbolygó Ferenc and Szücs Dénes for their knowledge of the Hungarian language, customs, and Tokaj! A special warm word of thanks is in order for Baliga Gabi whose conscientiousness is at a level I only can look up to in awe. In Maastricht, Bernhadette Jansma, Lisa Jonkman, and Fren Smulders helped me discover the ERP country, while Ron Hellenbrand, Johan Gielissen, and Huub Hamers solved any technical problems. Christaan Paffen, Petra Vlamings (I still don't know how you get any lead below 5 k Ω in the first pass), and Nina Davis helped in running some of the experiments in Maastricht.

An indispensable help were some e-mails from Paul Boersma, which taught me how to use PRAAT. My activities during my Ph.D. period were, and hopefully continue to be, based on an embodied and action-related perspective, for which I have to thank Lex van der Heijden and Odmар Neumann, who introduced me to theoretical psychology. Wido La Heij also took part in this in addition to lending a lot of moral support. Helma Kruse deserves a more than honorable mention in this last category as well.

Here in Maastricht, the atmosphere got merrier as more people arrived. Nienke van Atteveldt, Milene Bonte, Elia Formisano, Henk Jansma, Amanda Kaas, Niko Kriegeskorte, Angelika Lingnau, Judith Peters, Joel Reithler, Alard Roebroek, Alexander Sack, Sandra Smeets, and Anton de Vries helped making the time on and off University not only scientifically worthwhile. I will miss your patience with my jokes. Finally, Udo Hanrieder's graphical and Nicola Gundt's language talents helped making the final version much better than I ever could have achieved.

Contents

Chapter 1

Introduction	9
The subject matter	10
How can the input [tæym] be mapped on the lexical representation /tæym/?	12
An outline of the thesis	19

Chapter 2

Testing an Underspecification Account Using the MMN	21
Method	24
Results	25
Discussion	27

Chapter 3

Evidence for Early Compensation	29
A regressive-inference account	31
The present experiments	33
Experiment 1	34
Experiment 2	38
Experiment 3	41
Experiment 4	45
General Discussion	49

Chapter 4

Place Assimilations and the Interplay between Speech Perception and Phonology	53
Method	57
Results	58
Discussion	59

Length does matter	63
Experiment 1.....	68
A corpus study.....	71
Experiment 2.....	73
General Discussion	78
Appendix.....	80

Chapter 6

Perceptual Integration Drives the Recognition of Assimilated Word Forms	81
The present experiments.....	86
Experiment 1: Identification	88
Experiment 2: Discrimination.....	96
General Discussion	99

Chapter 7

The Automatic Processing of Assimilated Word Forms Indexed with the MMN	105
Experiment 1: Hungarian listeners	110
Experiment 2: Dutch listeners.....	116
Experiment 3a	120
Experiment 3b	122
General Discussion	123

Chapter 8

Compensation can be Triggered by Non-Speech Sounds	129
Experiment 1.....	134
Experiment 2.....	139
Experiment 3.....	142
General Discussion	144

Chapter 9

In Search of the Auditory, Phonetic and/or Phonological Problems in Dyslexia	147
Experiment 1.....	154
Experiment 2.....	160
Experiment 3.....	165
General Discussion	168

Chapter 10

Epilogue	171
Questions answered.....	172
Questions unanswered.....	175

Summary/Samenvatting	177
References.....	184
Curriculum Vitae.....	191

Introduction

The subject matter

Every time a sentence is being spoken, a *war* is going on: A war between speaker and listener. But behold: this is not going to be a modern version of ‘*homo homini lupus est*’ (Hobbes, 1983). This is just a drastic formulation that introduces the contradictory interests of speaker and listener with regard to speech production. The listener wants the speaker to speak as clearly as possible in order to make his task in comprehension easier. The speaker, on the other hand, is lazy and wants to put as little energy as possible into speech production (see Tobin, 1997, for an overview of the history of similar conceptions in phonology). You may have witnessed an indication of this inherent laziness of the speaker yourself: Sometimes, if very tired, speakers just say something like ‘mmmmh...mmmmhmmhmm.’ This is an example of extremely speaker-oriented speech in which there is nearly no lip movement and no tongue movement.

However, the listener does not like this type of speech, as it is very difficult if not impossible to understand. What he or she likes is the kind of speech one hears from the Dutch queen when opening a new parliamentary term (the so called ‘troonrede,’ the queen’s speech). It seems that the everyday speech we use to communicate is somewhere in the middle between these two extremes. Everyday speech seems to be a compromise between these two extremes, which is, to stick to the political metaphor, also typically Dutch (cf. Woltjer, 2001). However, the speaker gets more out of this compromise than one might naïvely think. Most listeners, for instance, are unaware of the fact that word-final and word-initial consonants may be strongly influenced by the surrounding speech sounds, a process called phonological assimilation. Due to such assimilations, explained in more detail below, the Dutch word ‘tuinbank’—just as its English translation ‘garden bench’—may be pronounced as ‘tuinbank’—in English ‘gardembench.’ This makes it easier for the speaker as an /mb/ sequence is easier to pronounce than an /nb/ sequence. However, the task for the listener has become more difficult; he has to understand ‘garden’ in ‘gardembench’ and in ‘gardenchair.’ At the same time, the listener must still differentiate ‘gun’ and ‘gum.’ This dissertation is about how listeners deal with this problem.

So, what is a phonological assimilation anyway? Let’s start with a *prima vista* understandable conception developed within the paradigm of Generative Linguistics (Chomsky & Halle, 1968). This framework describes phonological assimilations as rules. These rules are different for each language, and one of the two rules we are concerned with here is the rule of nasal place assimilation. Basically, this rule says that a word-final /n/ may take over the *place of*

*articulation*¹ of the onset of the following word if this word starts with an obstruent.² To give an example, consider the Dutch compound noun /tæyn/ + /banjk/ (Engl. 'garden' + 'bench'). The first noun ends on an /n/ and the second noun starts with a /b/, which is an obstruent. Therefore, the conditions for the rules of nasal place assimilations are met, and the compound noun may be pronounced /tæymbanjk/. Given that the subject matter is assimilation *rules*, the "may be pronounced" in the previous sentence seems odd. One might have expected a "must be pronounced." While this rule approach has its merits, it does not awfully well in explaining these kinds of assimilations. First of all, the rule is applied in only 60% of the cases (van Heuven & van Berg, 1982). Thus the compound is sometimes pronounced [tæymbanjk] and sometimes pronounced [tæynbanjk]. The generative approach solved this by calling rules like the Dutch nasal place assimilation optional. However, things got even worse for this rule approach when evidence started to accumulate that /tæynbanjk/ may not only be pronounced [tæymbanjk] or [tæynbanjk], but that all kinds of half-castes between /m/ and /n/ may be found at the end of /tæyn/ in /tæynbanjk/ (Gow, 2001, 2002a; Nolan, 1992; see also chapter 5). While it is still bearable that a rule may be applied just sometimes, it is impossible to see how a rule can be applied 'a little' in order to produce the half-caste segments that are often found in context in which (nasal) place assimilation is possible. In contrast to the rule-based approach, assimilations may be described more aptly as an overlap of articulatory gestures (Browman & Goldstein, 1992). Assimilation in /tæynbanjk/ is then due to an overlap of the lip-closure gesture for the /b/ with the tongue-tip closing gesture for the /n/. This overlap may be complete, then the speaker produces a full assimilation [tæymbanjk]. There may be no overlap at all, which leads to the unassimilated [tæynbanjk]. Half-castes can arise as a consequence of an incomplete gestural overlap between the /n/ and the /b/.

Even if conceived as arising from gestural overlap, phonological assimilations still lead to a problem for the perceiver. The assimilated form [tæymbanjk] often resembles a form with a labial nasal more than it resembles a form with the canonical pronunciation with an /n/. So, how

¹ The Dutch nasals /n/, /m/, and /ng/ differ in where the airflow through the mouth is cut off in order to let all air escape through the nasal passage. This is called the place of articulation. For the /m/ the air passage through the mouth is obstructed by a closure of the lips, therefore this place of articulation is called 'labial.' For further information, see Rietveld and van Heuven (1997) or Ladefoged and Maddieson (1996).

² Obstruent refers to a class of speech sounds (phonemes) that are, first consonants, and second, quite opposite to vowels. In Dutch, these are the stops (/b/, /p/, /t/, /d/, /k/) and the fricatives (/f/, /v/, /s/, /z/, /g/, /ʃ/). For further information, see Rietveld and van Heuven (1997) or Booij (1995).

can the listener then recognize the word /tæyn/ in [tæymbaŋk]? Recognition here means that an input can be linked to an internal representation through which previously acquired knowledge may be accessed. In psycholinguistic terms, the question then is: How can the input [tæym] be mapped on the lexical representation /tæyn/?

How can the input [tæym] be mapped on the lexical representation /tæyn/?

The laymen's answer to the previous question seems obvious: Well, the listener understands the [baŋk] part as indicating the concept 'bench' and then the small difference between an /n/ and an /m/ is overcome with a reference to semantics: [tæym] is not a word and it resembles the Dutch words [dæym] (Engl. 'thumb') and [tæyn] (Engl. 'garden'). As it is more likely that the speaker refers to a 'garden bench' than to a 'thumb bench,' the listener can infer that the speaker wants to communicate the concept 'garden bench.' We will argue that this is not a workable solution. However, first of all, we deal with a more sophisticated version of such a 'just-neglect-the-difference' account. Lahiri and Reetz (2002) argued that assimilated word forms are recognized, because lexical representations are *underspecified*. This underspecification account assumes that features that may vary between utterances as a consequence of place assimilation are not specified in the lexicon. This means that the lexical entry for the concept 'garden' in Dutch would be /tæy + {any nasal}/. [tæym] matches this entry and is thus recognized as an instance of the concept 'garden.' In contrast with this prediction, our data indicate that [tæym] is not generally recognized as an instance of the concept 'garden.'

Discarding any account that argues that [tæym] is similar enough to /tæyn/ in order to be recognized anyway leaves us with the problem how recognition is achieved after all. One possibility is that the context may play an important role. It is safe to say that speech perception, or even perception in general, is context-sensitive (Warren, 1999). Therefore, it is not unlikely that the context plays a role in the process of mapping the input [tæym] onto the lexical representation /tæyn/, given that the assimilation occurs only in certain contexts. [tæym] may only be recognized as an instance of /tæyn/ if it occurs in a context that allows the assimilation (as in [...baŋk]), but not if it occurs in the context [...stu:l], in which assimilation may not occur. To this end, we presented utterances of the original, unchanged [tæyn] and the changed version [tæym] to listeners in contexts in which the assimilation may occur or not. These contexts were [...baŋk] and [...stu:l]. We call the [...baŋk] context 'viable,' because [tæymbaŋk] is a possible pronunciation of /tæynbaŋk/. In contrast, the [...stu:l] context is called 'unviable,' because

*[tæymstu:l]³ is not allowed as an alternative pronunciation of /tæynstu:l/. This leads to the basic factorial structure used in the majority of the experiments described here, which is depicted in Table 1.1. The crucial feature indexing compensation for assimilation in this paradigm is an interaction between the factors Change and Context. The changed form should be treated differently in the viable context than in the unviable context: It should be accepted as an instance of the canonical form in the viable context, but rejected in the unviable context. This implies that the listener uses the context in order to compensate for the phonological assimilation in perception. The basic design is first used in a word-identification task, in which participants have to decide between two given alternatives, that is, two-alternative forced choice (2AFC). The instruction then is to indicate, upon hearing one of these stimuli, whether the first word of the compound noun ends on an /n/ or an /m/.

Table 1.1 Basic factorial design of the experiments

Factors		Context	
		viable	unviable
Change	unchanged	[tæynbʌŋk]	[tæynstu:l]
	changed	[tæymbʌŋk]	*[tæymstu:l]

Secondly, the same stimuli are used in a passive-listening task. At first glance, it may seem odd that any relevant evidence can be collected if participants just listen passively to these utterances. However, if the electro-encephalogram (EEG) is recorded during passive listening, the EEG can be averaged over all presentations of the same stimulus in order to obtain an event-related potential (ERP). This method has proven to be particularly fruitful when the listener hears a series of sounds at a steady rate (e.g., 0.2-1.5 Hz) and the series consists of only two stimuli of which one is more frequent (0.9) than the other (0.1). Then, a negativity arises in the ERP to the 'deviant' stimulus ($p = 0.1$) if compared to the standard stimulus ($p = 0.9$). The extent of this negativity, the mismatch negativity (MMN) can be used as a measure of the extent to which the listener makes a difference between the standard and the deviant (Näätänen & Winkler, 1999).

The results obtained with these methods support the assumption that the perception of phonological assimilation is context-sensitive. In the identification task, participants often

³The *star is used in Linguistics to mark an unacceptable form.

indicated that [tæym] ... [tæymbaŋk] contains an /n/, while this was not the case for the [tæym] in [tæymstu:l]. That is, the word /tæyn/ is recognized in [tæymbaŋk] but not in *[tæymstu:l]. In addition, in the passive-listening task, a MMN was only observed for the deviant *[tæymstu:l] compared to a standard [tæynstu:l] but not to a deviant [tæymbaŋk] compared with a standard [tæynbaŋk]. This indicated that listeners made a difference between *[tæymstu:l] and [tæynstu:l], but not between [tæymbaŋk] and [tæynbaŋk], although the acoustical difference was identical. It should be mentioned that this constitutes the first report that the MMN reflects context sensitivities in auditory speech perception.

The first two empirical chapters hence clearly show that the perception of assimilated word forms is dependent on the phonological context. It is not the case that [tæym] is similar enough to /tæyn/ to be recognized anyway. This result sets the stage for the rest of the studies. The proof of some form of context sensitivity, that is, ‘compensation for assimilation,’ raises several questions:

1. Is compensation for assimilation a general phenomenon or restricted to nasal place assimilation (NPA) in Dutch?
2. At what kind of processing level does the compensation for assimilation arise?
3. Is compensation for assimilation acquired during language acquisition?
4. If this compensation for assimilation is an important process in speech comprehension, is it impaired in populations which have a (subtle) speech-comprehension deficit?

One of the critiques that may be raised—and was raised—against our conclusion that the perception of assimilated word forms is context-sensitive, is that we generalize from, first, the stimulus pair used in this study to all Dutch phrases in which NPA was applied; and, second, from Dutch phrases with NPA to all assimilated utterances in the world’s languages. In order to counter this critique, we tried to replicate the results obtained with material from a different language with a different assimilation rule and, trivially, different stimulus material. Optimizing the ratio of linguistic distance to geographical distance led us to our second quadruple of stimuli, displayed in table 2.1.

Table 1.2 Basic factorial design with Hungarian stimuli

Factors		Context	
		viable	unviable
Change	changed	[bɔlro:l]	[bɔlna:l]
	unchanged	[bɔrro:l]	*[bɔrna:l]

These utterances are based on the Hungarian language. Hungarian is one of the non-Indo-European languages in Europe; another non-Indo-European language would be Basque, which is, as it seems the only surviving truly European language (see Cavalli-Sforza, 2001). Hungarian belongs to the Finno-Ugric languages with Finnish, Estonian, and Lapp as other prominent examples. These languages seem to have evolved from a common ancestor spoken in the Asian Southern Ural area. Given the probably poor expertise in Hungarian of the average reader, some explanation with regard to these stimuli is at its place. /bɔl/—pronounced as ‘bol’ with a vowel similar to a short version of the vowel in ‘port’— is the Hungarian word for ‘left.’ Now a feature of the Hungarian language comes into play which gives second-language learners serious headaches. What constitutes a word in Hungarian usually contains a number of morphemes and, if translated into English or Dutch, will correspond to a number of words. /bɔlro:l/, for instance, contains the morphemes /bɔl/ (Engl. ‘left’) and the Delative case suffix /ro:l/.⁴ The Delative indicates a movement from a surface, as from a table. This contrasts, for instance, with the Elative suffix /bɔl/, which indicates a movement from within something, as, for instance from within a house.

After this little excursion, we have learned that /bɔlro:l/ means ‘from the left’ (but only if the movement comes from a surface). As the reader might suspect after inspecting Table 1.2, Hungarian knows a rule of /l/-to-/r/ assimilation which is triggered by a /lr/ sequence. Therefore [bɔrro:l] is a viable alternative pronunciation of /bɔlro:l/. However, as Table 1.2 indicates, *[bɔrna:l] is an unviable alternative pronunciation of /bɔlna:l/. The observant reader has already guessed that /na:l/ also is a Hungarian case suffix, of which Hungarian has so many to offer. Indeed, /na:l/ is the case suffix for the Adessive, which would be used in order to

⁴The reader should also note that Hungarian not only has 16 different case suffixes, but that these suffixes also exist in different forms depending on the vowels in the root words.

answer a 'where' question if the answer is on the side of something, as in 'at the Balaton Lake.' In contrast to the Delative suffix /ro:l/ the Adessive suffix /na:l/ cannot trigger liquid assimilation, so that *[bɔrna:l] is an unviable alternative pronunciation of /bɔlna:l/.

So, we have another quadruple of stimuli that we can use in order to investigate compensation for assimilation. But is this not just more of the same, a rather boring replication? Not quite. All previous studies, including our own, examined the case of place assimilation rules in highly related languages. The new stimulus quadruple represents a manner assimilation rule in a language completely different from the Germanic languages that formed the basis of the earlier studies.⁵ In addition, the acoustics of the phonemic difference that is superficially neutralized is quite different. Therefore, replicating our earlier results with these stimuli may determine whether both our generalizations hold: The context-sensitive perception of assimilated forms is neither restricted to our specific stimulus pairs nor to NPA in Dutch. This answers the first of the four questions stated above.

This leads us to the *second* question: At what kind of processing level does the context sensitivity arise? In order to evaluate this question, we took two routes. First of all, we examined to what extent compensation for assimilation is influenced by the acoustic details of the utterances. The logic behind this is that on higher processing levels, abstraction has led to the loss of information about acoustic details. If acoustic details influence compensation for assimilation, this indicates a processing level at which auditory or phonetic details of the utterance are still available. However, we figured that this is only one part of a conclusive story. It is one thing to argue that acoustic details may influence compensation for assimilation. It is quite another thing why this should be the case. Compensation for assimilation should occur on a level at which acoustic detail is available, if acoustic detail contains relevant information. Consider a phrase like 'gun production.' Due to NPA, this may be pronounced 'gum production.' This seems to generate a homophone, and the possibility for ambiguity, especially in American English. However, as mentioned above, there may be small acoustic differences between utterances of 'gum production,' when the speaker intends to refer to the production of chewing gum, and an utterance of 'gun production,' when the speaker intends to refer to the production of guns. If this is the case, compensating for assimilation after abstracting from the acoustic signal actually means that information is disregarded. Therefore, we also conducted production studies in order to show that acoustic detail bears information.

⁵ This is not entirely correct. Hungarian words like 'paradiscom,' 'semkli,' 'kiffli,' and 'kavchaz' may seem familiar to German, and even more so to Viennese, readers.

Our second strategy to investigate at which level compensation for assimilation occurs was to enlarge our task repertoire beyond the 2AFC and passive-oddball task used previously. Following the lead of Beddor and Krakow (1999) and Kingston and Macmillan (1995), we used a discrimination task. If context influences discrimination performance, this indicates an integration of target and context on early auditory levels. If, in contrast, there is a dissociation of the effect of context between identification in a 2AFC task and performance in a discrimination task, this would indicate a higher-level process.

The question, at which level compensation for assimilation occurs, brings with it the question of the 'pre-lexicality' of the context effects. If we want to argue that compensation for assimilation happens at an auditory processing level, it would be odd if compensation for assimilation were influenced by the wordness of the assimilated forms (especially given the results of Pitt and McQueen, 1998). That is, is compensation stronger when it yields a word? If this is the case, the /m/ in 'leam bacon' should be more likely to be perceived as an /n/ than the /m/ in 'heam bacon.' We investigated this for both types of assimilation, NPA in Germanic languages and the liquid assimilation rule in Hungarian. To sum up, looking at the possible role of acoustic detail, the influence of lexical status, and contrasting performance on an identification task and a discrimination task allows us to answer our second question: At what kind of processing level does the compensation for assimilation arise?

This leads us to the *third* question. Is compensation for assimilation acquired during language acquisition, or is it innate? This is investigated by cross-linguistic comparisons. We probed the perception of Dutch utterances with nasal place assimilation by Portuguese listeners. The Portuguese language does not have a rule of nasal place assimilation. Therefore, the behavior of Portuguese listeners may indicate how naïve listeners perceive nasal place assimilations.

Furthermore, we were in the fortunate position to be able to probe the perception of the Hungarian stimuli by Dutch and Hungarian listeners. In addition, we managed to find a sample of native speakers of Dutch who do not speak Hungarian. These listeners have no experience with Hungarian liquid assimilation—or any similar rule—despite employing an /l/-/r/ contrast in their native language. This allows us to extend the cross-linguistic study undertaken with Portuguese and Dutch listeners. The results show in effect that we are not dealing with a single mechanism, but that both innate perceptual abilities and language learning contribute to compensation for assimilation.

The fact that innate perceptual abilities seem to contribute to the perception of assimilated forms is somewhat counterintuitive. Given that assimilation rules are to some degree

language specific, which makes the cross-linguistic design powerful in the first place, it seems odd that general perceptual abilities seem to be tuned for compensation for assimilation in any language. These results somehow backfire on our initial question. We started out on this endeavor by investigating how the listener solves the problem caused by phonological assimilations. It seems, however, that compensation for assimilation taps into basic perceptual abilities of the listener. This leads us back to the metaphor of a war between speaker and listener. Our starting point was to consider assimilations as a triumph of the speaker over the listener, who now has to deal with assimilated forms. Now it turns out that the balance of power may not be so one-sided after all. It seems possible that the listener may also exert some power over the speaker, and constrain him so that his assimilations have to conform to the listener's initial preferences. Although rather metaphysical at first glance, this idea seems to gain some credibility, as attested by the publication of a book by Hume and Johnson (2001) entitled "The role of speech perception in phonology." This idea of interplay between speaker (phonology) and listener (speech perception) will recur throughout this dissertation.

The framework of an 'interplay between perception and phonology' raises one particular question: If the general perceptual abilities of the listener influence the kind of assimilations a speaker may produce, independent of an ontogenetic or phylogenetic adaptation of the listener to assimilation rules, then non-speech sounds should in principle also be able to trigger context effects in the perception of changed word forms. To test this hypothesis, we created non-speech analogues of the [ro:l] and [na:l] case suffixes and evaluated the context effects caused by non-speech context sounds. This study is crucial for bringing home a central statement of this endeavor: Compensation for assimilation partly depends on general perceptual principles. This, in turn, indicates that assimilation rules are to some extent shaped by the perceptual preferences of a blue-print listener. That is, there is an influence of perception on phonology.

Having dealt with the third question, whether the context-dependency is acquired during language acquisition, we come to the *fourth* and final question: If compensation for assimilation is an important process in speech comprehension, is it impaired in populations, which have a (subtle) speech-comprehension deficit? One important group for which a subtle speech-perception deficit has been indicated is the group of people suffering from developmental dyslexia. Therefore, we investigated compensation for assimilation, and other context effects in speech perception, in a dyslexic sample compared to a control group. This study tries to solve a paradox that previous research has raised. A number of studies indicated that dyslexics do not perceive speech categorically. Although the strong claim of categorical perception of speech sound is practically dead (Schouten, Gerrits, & van Hesson, 2003), it has been a standard finding

that speech sound continua are not perceived as continuous. That is, if a speech sound (e.g., such as the /t/ in 'ta') is morphed with some intermediate steps into another speech sounds (e.g. such as the /k/ in 'ka'), the perception of this continuum is not continuous. Instead, listeners hear most sounds as predominantly 'ta' or 'ka,' while only a small subset of sounds is perceived as ambiguous. Dyslexics deviate from this pattern and perceive a larger range of speech sounds as potentially ambiguous. A straightforward explanation of this behavior is that dyslexics do not perceive the underlying acoustic differences reliably, and therefore, show more uncertainty in the categorization of speech sounds. However, Mody and Studdert-Kennedy (1997) showed that dyslexics perceive the underlying acoustic dimensions of speech sounds without impairment. This leads to the paradox that dyslexics show more uncertainty in the categorization of speech sounds, even though the perception of the underlying acoustic dimensions is not impaired. Therefore, I constructed a mathematical model that simulated the effects of insufficient context sensitivity on performance in a categorical-perception task. The model shows that the paradox arising from intact auditory perception but impaired speech perception by dyslexics can be resolved: If it is assumed that dyslexics show insufficient context sensitivity in speech perception, a categorical-perception deficit arises despite intact auditory perception. Therefore, we tested whether dyslexics would indeed show less context sensitivity than normal-reading children in speech perception. Giving away the punch line, the results did not confirm our hypothesis. Instead, context sensitivity in speech perception seems to be similar in dyslexic and control groups.

An outline of the thesis

Following this introduction, chapters two and three establish the fact that assimilated forms are perceived in a context-sensitive way. This indicates some 'compensation for assimilation.' The rest of the dissertation is then dedicated to illuminating the mechanisms that drive 'compensation for assimilation.' Chapter four investigates the perception of NPA in a Dutch utterance by Portuguese listeners, who do not have NPA in their native language. Chapter five then investigates the role of fine acoustic/phonetic detail in the perception and production of utterances with NPA.

Chapters six to eight are concerned with the perception of utterances to which Hungarian liquid assimilation has been applied (see Table 1.2). First, chapter six investigates the 'active' perception of these utterances in an identification and a discrimination task by Hungarian and Dutch listeners. Chapter seven then probes the 'passive' perception of these utterances in a passive-oddball task with concurrent EEG measurement, again by Hungarian and Dutch

listeners. Chapter eight then investigates the possibility of 'compensation for assimilation' triggered by non-speech sounds.

Finally, chapter nine compares context effects in the speech perception of a group of dyslexic children to the context effects in speech perception in a control group. An epilogue will then try to answer the questions stated here based on the obtained results.

Testing an Underspecification Account Using the MMN¹

Word recognition requires activation of the correct lexical representation by auditory input. Theories of word recognition differ in their assumptions regarding the amount of detail of lexical representation. According to an Underspecification account, segments that may change in some phonological context (e.g., 'garden' may sometimes be pronounced 'gardem') are not specified in the mental lexicon: According to this account, an input does not mismatch an internal representation, if the distinguishing feature may be subject to assimilation in running speech (i.e., 'gardem' does not mismatch 'garden'). We used the Mismatch Negativity in order to investigate this claim of Underspecification theory. Results failed to corroborate an Underspecification account for the recognition of assimilated word forms. This favors theories that assume detailed lexical representations with powerful pre-lexical mechanisms that compensate for systematic variance in connected speech.

¹ This is an adapted version of Mitterer, H. & Blomert, L. (submitted). How are assimilated word forms recognized? Testing an Underspecification account using the MMN.

For the recognition of a given spoken word, certain cues may be more informative than others. For instance, it has been argued that, in English and Dutch, the onset of a word is less influenced by circumstantial factors such as speech rate and phonetic context than word endings are (see, e.g., Stevens, 2002). The onset of a word is likely to be stressed in English and Dutch (Norris, McQueen, Cutler, & Butterfield, 1997) and segments in stressed syllables provide more stable cues for recognition than segments in unstressed syllables. However, word endings are often unstressed, may sometimes be deleted (Ernestus, 2000), and often lose some cues due to assimilation processes. For example, in the utterance 'garden bench,' the word-final alveolar nasal of the first noun may take over the labial place of articulation of the next segment to lead to the utterance 'gardem bench'.

Given that word-final segments do not provide solid cues for word recognition, it is not surprising that some have argued that word-final segments have only a limited role to play in the recognition of spoken words. One way to formalize such an assumption is based on Underspecification theory. Underspecification theory was originally developed in phonology and argued that some phonological features of words are not specified in the lexicon, but are computed online using default values and phonological context. In its adaptation to word-recognition theories (Lahiri & Marslen-Wilson, 1991), it is assumed that cues, which are relatively stable, are represented in the lexicon of the listeners. In contrast, unreliable cues are not specified in the lexicon. Therefore, the lexical representation is specified for only a subpopulation of the input features, hence underspecified. Consider the previous example of an utterance 'gardem bench.' How can the first word of this compound be recognized as an instance of 'garden'? According to an Underspecification account, the lexicon of the listener only specifies the reliable cues for the word 'garden.' This means that the lexical representation does not specify the place of articulation of the word-final nasal, because this cue can be modified by assimilation as in 'gardem bench,' or the Dutch translation of this phrase, [tœy(n/m)baŋk]. The lexical representation of the Dutch word for 'garden' is thus [tœy+ (any nasal)], and the assimilated input [tœym] matches this representation.

A number of studies corroborated such an Underspecification account. Gaskell and Marslen-Wilson (1996, Exp. 1) found that changed word forms such as [gardəm] activate the respective lexical representation 'garden' as strongly as a canonical pronunciation [gardən]. Lahiri and Reetz (2002) present a more rigorous test of an Underspecification account for the recognition of assimilated words. This test is based on the assumption of the Underspecification account that word-final /n/s are not specified as a consequence of the possibility to undergo

assimilation. In contrast, word-final /m/s are specified, because word-final /m/s are not assimilated to become alveolars by a process of place assimilation (at least in English, Dutch, and German). Therefore, a form like [gɑrdəm] should produce no mismatch with the lexical representation of 'garden,' due to underspecification of the word-final alveolar nasal. However, a form like [flain] produces a mismatch with the lexical representation of 'flame,' because this representation is specified for the place of articulation of the word-final nasal. Lahiri and Collier (to appear, cited in Lahiri & Reetz, 2002) tested this assumption using a cross-modal priming task. Participants heard spoken words and had to make a lexical decision to simultaneously presented visual words. Results showed that, in accordance with the Underspecification account, a spoken [gɑrdəm] primed a visual lexical decision to a semantically related word (e.g., 'flower') of the non-mismatching lexical entry 'garden.' However, a spoken [flain] did not prime words semantically related to the lexical entry 'flame.' This result corroborates the Underspecification account for the recognition of assimilated word forms.

However, other reports were not as favorable with regard to the Underspecification account. Gow (2001) devised a priming experiment using English tokens in the same experimental conditions as Lahiri and Collier (to appear) and failed to find results in support of Underspecification. In this experiments, tokens like [flain], derived from 'flame,' produced as much priming as tokens like [gɑrdəm] derived from 'garden.' The instability of results might be attributed to the fact that cross-modal priming and visual lexical decision are tasks that may lead to a strong involvement of experiment-specific strategies (cf. Balota & Chumbley, 1984; Stone & Van Orden, 1993). Therefore, we aimed to test the Underspecification account in a way that is free of such possible alternative interpretations. To this end we made use of electrophysiological measurements while participants listened passively to spoken words. With this method, it should be possible to collect data, which are unbiased by contamination from task-related aspects. Given the equivocal results of behavioral measures, such measures are needed in order to judge the adequacy of the Underspecification account for the recognition of assimilated word forms.

The Underspecification account argues that an input with a word-final /n/ *mismatches* an internal representation with a word-final /m/, while an input with a word-final /m/ *does not mismatch* an internal representation of a word with a word-final /n/ due to underspecification. Auditory mismatch processes have been widely studied with event-related potentials (ERPs): If a stimulus sequence consisting of two stimuli (AABABA...) is presented, and the two stimuli differ in their frequency of occurrence (e.g., 0.9 vs. 0.1), then the odd, or deviant, stimulus gives rise to a negative deflection in the ERP, called the Mismatch negativity (MMN, see, e.g.,

Näätänen & Winkler, 1999; Schröger, 1998). This MMN does not reflect differences in the acoustic properties of the two stimuli, but rather perceptual differences. It has been shown that the MMN is influenced by language experience (Näätänen et al., 1997; Winkler et al., 1999), and that the MMN accesses phonological categories (Phillips et al., 2000) and lexical representations (Pulvermüller et al. 2001; but also see Wunderlich & Cone-Wesson, 2001).

Given the sensitivity of the MMN to the linguistic processing of auditory input, a straightforward test of the Underspecification account for the recognition of assimilated words can be designed. According to this account, a deviant with a word-final /m/ should not mismatch a standard with a word-final /n/, because the /n/ in the standard is not specified for place of articulation. Nevertheless, one may expect a small auditory MMN: Winkler et al. (1999) showed that a MMN can be recorded even if the difference between standard and deviant is not linguistically relevant. However, a deviant with a word-final /n/ not only leads to an auditory mismatch with a standard with a word-final /m/, but also to a language-specific mismatch because of a mismatch in phoneme identity with the fully specified /m/. Therefore, this latter case may be expected to lead to a larger MMN than is the case for the deviant with a word-final /m/ and a standard with a word-final /n/.

In order to test this prediction, we used the Dutch word for 'garden,' which is [tæyn] and the possible alternative pronunciation [tæym], which may arise due to place assimilation in a compound such as [tæymbanj] (Engl. 'garden bench'). According to an Underspecification account for the recognition of assimilated forms, the deviant [tæyn] leads to a mismatch with the standard [tæym], while the deviant [tæym] does match the standard [tæyn]. That is, the MMN should not be the same when the roles of standard and deviant are reversed. As participants only listened passively and were diverted by watching a silent video, this constitutes a test of the Underspecification account for the recognition of assimilated utterances that is free from experiment-specific strategies.

Method

Participants

Nine students of the Maastricht University took part in the experiment, six females and three males. Participants were paid for participation. None of the participants reported a history of hearing problems. All were native speakers of Dutch.

Design

ERPs were recorded in two passive-listening oddball series. In one series, the standard was [tæyn] and the deviant was [tæym]. In the other series the standard was [tæym] and the deviant was [tæyn].

Materials

The stimulus materials were the cross-spliced utterances [tæyn] and [tæym] also used in chapter 3. To create this stimuli, [tæy] was spliced from an utterance of [tæynstu:l]. An alveolar and a labial nasal murmur were spliced off from two other utterances of [tæynstu:l] and [tæymbaŋk]. The nasal murmurs were equated in length and energy (using a root-mean square measure) using the 62 ms /n/ from the [tæynstu:l] utterance as template. Due to the constraint of splicing at zero-crossings, the nasal murmurs differed slightly in length. However, this was less than 1 ms. The edited nasal murmurs were concatenated with [tæy], resulting in the stimuli [tæyn] and [tæym]. These stimuli led to a clear perception of a word-final /n/ and /m/ (see chapter three, Exp. 1).

Apparatus

The stimuli were played to the participants by a DOS-operated computer using the software package ERTS (Behringer, 1996). The output of the SoundBlaster soundcard was redirected to 4 loudspeakers (JBL Control 25) within a sound-attenuated booth with a Spirit Folio mixer and amplified by a Yamaha P4050 power amplifier. The maximal sound-pressure level was 70 dB(A).

Procedure

After preparation for EEG-recording was finished, participants were seated in a sound attenuated room and instructed to relax and watch a silent movie. There were four blocks of measurement, each consisting of 800 stimuli (85% standards and 15% deviants, trial onset asynchrony 0.9s). A block lasted about twelve minutes. Each of the two oddball series (see Design) was presented in two blocks. Presentation order of blocks was counterbalanced across participants.

Electrophysiological recording and data reduction

The nose-referenced electroencephalogram (EEG) (0.1-27 Hz, sampling rate 256 Hz) was recorded with a 32-channel electrode cap covering frontal, central, temporal, and parietal scalp areas. Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supra-orbital ridge of the right eye. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. All electrode impedances (EEG and EOG) were kept below 5 kOhm. Event-related potentials (ERP) were obtained by averaging the EEG time-locked to the onset of the sound. The ERPs were band-pass filtered digitally from 1 to 30 Hz. The signal was normalized using a baseline that was calculated using the mean amplitude from 30 ms before stimulus onset to the onset of the mismatching nasal murmur, which was 170 ms post-stimulus onset (cf. Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993). Eye blink artefacts were reduced using linear regression. Epochs exceeding $|75| \mu\text{V}$ on any of the EEG channels or the horizontal eye channel were rejected, leading to a rejection of about 30 % of the trials. The individual ERP averages were obtained from 270 raw epochs in each condition by using all deviant stimuli and the same number of standard stimuli randomly drawn from all standard stimuli, excluding standards directly following deviants. The MMN was quantified in the following way (cf. Schröger, 1998). An area measure was obtained from the electrode location Fz. Mean amplitude was calculated for a 100 ms window starting 100 ms after the onset of the mismatching nasal murmur. These values were used for statistical analysis.

Results

Figure 2.1 shows the grand averages for the four ERPs, [tæyn] and [tæym] as standard and deviant for the Fz electrode as well as both mastoids. The mean amplitudes in the MMN window were $-2.49 \mu\text{V}$ for [tæyn] as standard, $-3.72 \mu\text{V}$ for [tæyn] as deviant, $-2.57 \mu\text{V}$ for [tæym] as standard, $-3.47 \mu\text{V}$ for [tæym] as deviant. These values were used in a repeated-measures ANOVA with Nasal Murmur ([tæyn] vs. [tæym]) and Role (standard vs. deviant) as factors. This analysis revealed that the factor Role was significant ($F(1, 8) = 9.7, p < 0.025$), that is, the deviant stimuli produced overall a more negative ERP than the standards. However, neither Nasal Murmur nor the interaction of the two factors was significant ($F_s < 1$).

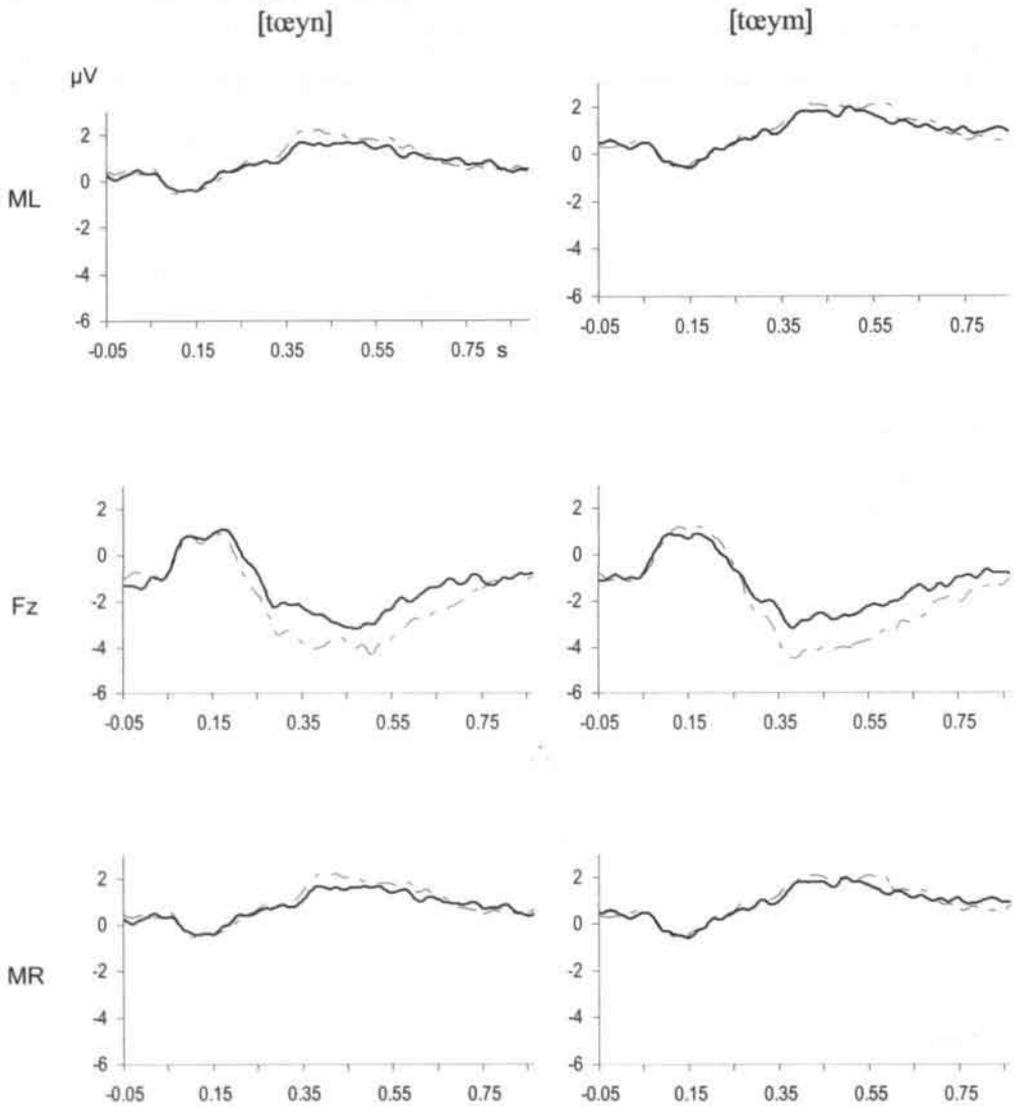


Figure 2.1 ERPs for the two stimuli [tœyn] and [tœym] at the left and right mastoid and Fz. The thick lines show the ERPs to the standards while the thin lines show the ERPs to the deviants.

Discussion

In this study, we tested a prediction of the Underspecification account for the recognition of assimilated word forms. According to such an account, inputs with a word-final /m/ should match internal representations with a word-final /n/, while an input with a word-final /n/ should mismatch an internal representation with a word-final /m/. This was tested using a passive-oddball paradigm. An Underspecification account predicts that the deviant [tæym] should not produce a mismatch with the standard [tæyn]. In contrast, the deviant elicited a MMN that was equivalent to the deviant [tæyn] with the standard [tæym]. This may indicate that representations for word recognition are not underspecified for the place of articulation of word-final consonants.

If representations are fully specified, however, the problem arises how an assimilated form like [gardəm] can match a fully specified lexical representation of 'garden.' Proponents of fully specified lexical representations (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2001, 2002a) assume that pre-lexical mechanisms exist that are able to compensate for the variance in the signal caused by place assimilations. Although the precise implementation of this pre-lexical mechanism is still under debate, it is generally assumed that a pre-lexical mechanism evaluates the word-final /m/ by taking into account the following context before the input [gardəm] is matched to a lexical representation of 'garden.' If the following context is a labial consonant—which may trigger place assimilation—the input is adjusted, so that it matches a full representation of the word 'garden.' If, however, the next segment after [gardəm] is a vowel or an alveolar consonant—which cannot trigger assimilation—no adjustment takes place and the word is perceived as a speech error. Evidence for such a context-sensitive processing of assimilated word forms is growing (Darcy, 2002; Gaskell & Marslen, 1996, 1998; Gow, 2002a).

One might argue that the present experiment does not represent a fair assay of the Underspecification account, if one assumes that the MMN probably reflects auditory processing more strongly than linguistic processing. However, previous research has shown that the MMN reflects linguistic processes (Näätänen et al., 1997; Phillips et al., 2000; Pulvermüller et al., 2001). In addition, the next chapter shows that the MMN *can* be used to investigate the mechanisms that underlie the perception of phonological assimilation in running speech. Based on the assumption of a pre-lexical compensation mechanism as described above, we predicted that the MMN to a viable assimilation (e.g., [tæynbɔŋk] vs. [tæymbɔŋk]) should be smaller than the MMN to an unviable assimilation (e.g., [tæynstu:l] vs. *[tæymstu:l]). Indeed, we only

observed a reliable MMN for the unviable assimilation, although the acoustic difference between standard and deviant was identical in viable and unviable conditions. This supports the assumption that assimilations in running speech are compensated for pre-lexically. This, in turn, allows a full representation at a lexical level.

In sum, these results show that the MMN is a powerful tool for investigating the question how assimilated word forms are recognized. The available evidence with the MMN speaks against an Underspecification account and favors the assumption of fully specified lexical representations and a pre-lexical mechanism.

Evidence for Early Compensation¹

The pronunciation of the same word may vary considerably as a consequence of its context. The Dutch word "tuin" (Engl. 'garden') may be pronounced "tuim" if followed by "bank" (Engl. 'bench') but not if followed by "stoel" (Engl. 'chair'). A series of four experiments examined how Dutch listeners cope with this context sensitivity in their native language. A first word-identification experiment showed that the perception of a word-final nasal depends on the following phonological context. Viable assimilations but not unviable assimilations were often confused perceptually with canonical word forms in a word-identification task. Two control experiments ruled out the possibility that this effect was caused by perceptual masking or influenced by lexical top-down effects. A passive-listening study using electrophysiological measurements showed that only unviable, but not viable, phonological changes elicited a significant mismatch negativity. The results indicate that phonological assimilations are dealt with by an early, pre-lexical mechanism.

¹ This is an adapted version of Mitterer, H. & Blomert, L. (in press) Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*.

One of the biggest problems for human and machine speech recognizers is to deal with the enormous amount of variation between different instances of the same spoken word. Even within a single speaker, the same word may be pronounced quite differently depending on the phonetic context. In Dutch, for instance, the word 'tuin' [garden] may be pronounced [tœym] if the segment following the intended /n/ has a labial place of articulation, as the /b/ in 'tui(n/m)bank' [garden bench] (cf. Booij, 1995). This process is called nasal place assimilation. The matter is further complicated by the fact that this assimilation process is not obligatory and occurs in approximately 60% of the cases (Van Heuven & Van Berg, 1982). This leads to the problem that, on the one hand, listeners must be able to map the different forms [tœyn] and [tœym] onto the same meaning. On the other hand, listeners have to make a difference between other forms that also vary only in the place of articulation of a final nasal but differ in meaning. Take, for instance, the Dutch minimal pair 'duin' [dune] and 'duim' [thumb] that differs only in the place of articulation of a final nasal, which is precisely the feature changed by assimilation.

Early psycholinguistic research paid little attention to this aspect of the invariance problem (cf. Huckvale, 1999). It was assumed that word recognition processes might treat such variation as noise (Norris, 1982). Such an assumption was fostered by the finding that listeners often fail to notice small mispronunciations (Cole, 1973). Furthermore, variation may be corrected by top-down influences (McClelland & Elman, 1986). In a recent review of these matters, however, Norris, McQueen, and Cutler (2000) argued that the spoken-word recognition system is quite sensitive to acoustic variation (see, e.g., Marslen-Wilson, Moss, & van Halen, 1996). In addition, Norris et al. argued that top-down influences are neither necessary for explaining existing word-recognition data nor beneficial for more efficient recognition. Although the existence of top-down influences is still a matter of debate, results indicate that possible top-down influences are rather small (Samuel, 1996). The sensitivity of the word-recognition system is nicely illustrated by the fact that even sub-phonetic variation can hinder lexical access. Utman, Blumstein, and Burton (2000) altered words with initial unvoiced plosives by reducing the voice-onset time (VOT). These alterations were sub-phonetic, because these words were still identified as starting with an unvoiced plosive. However, in an identity-priming paradigm, these altered words produced less priming than words with a canonical VOT.

This shows that the word-recognition system does not generally tolerate variation. Therefore, phonological assimilations pose a problem for the spoken-word recognition system. Due to the sensitivity of the word-recognition system, the acoustic consequences of assimilation will be picked up and cannot be ignored, because this would blur the distinction between other words ('duin' [dune] and 'duim' [thumb]).

A regressive-inference account

Place assimilations do not occur at random, but are constrained by the following phonological context. If a word-final /m/ is followed by a non-labial segment, the /m/ has to be ascribed to the underlying word form. If, however, a word-final /m/ is followed by a labial segment (/b/, /p/, or /m/), the pronounced /m/ might correspond to an underlying /n/ that has been assimilated. Given this rule-like behavior, assimilations may be compensated for in perception by a mechanism that uses the phonological context in order to regressively infer the underlying place of articulation of a word-final nasal. Henceforth, we will use the term 'perceptual compensation for phonological assimilation' in order to refer to a mechanism that helps phonologically assimilated utterances to activate a lexical entry. That is, such a mechanism makes it possible for [tæym] to activate 'twin.' Gaskell, Hare, and Marslen-Wilson (1995) introduced a computational model for such a mechanism. The input to this model is composed of traditional phonological features. At the input level, phonological features are perceived as articulated. This implies that the consequences of assimilation pass unaltered through all acoustic and phonetic processing stages. That is, the utterance 'lean bacon' is perceived as containing an /m/ up to a phonological level. Only then, a hidden layer connected to a single recurrent network influences the activation of phonological feature nodes and corrects for assimilations. The activations of coronal segments are not corrected by this structure (coronal is the phonological feature corresponding to an alveolar place of articulation), reflecting the fact that coronal segments do not occur as the product of regressive English place assimilations (cf. Collins & Mees, 1996). This is due to the asymmetric nature of place assimilation. Although an intended /n/ can be pronounced as /m/, an intended /m/ can never be pronounced as /n/. Therefore, a pronounced /n/ has to correspond to an intended /n/. If, however, a pronounced /m/ is followed by another labial segment, this segment activates the unit corresponding to a coronal place of articulation. This form of phonological inference reflects the fact that the utterance 'lean_bacon' is a possible assimilated pronunciation of 'lean bacon.' Thus, the mechanism proposed by Gaskell et al. compensates for place assimilation by a process of regressive inference. That is, the interpretation of segment *n* is influenced by segment *n+1*.

Gaskell and Marslen-Wilson (1996) provided experimental evidence for this regressive-inference account. They used assimilated tokens like 'lean' pronounced with a final /m/ in a cross-modal lexical-decision task. While listening to sentences, participants performed a lexical-decision task to visually presented words. If the visual target was preceded by the same word in the auditory stream, reactions were faster, a so-called identity-priming effect. In the second experiment, changed tokens ('lean' -> 'leam') were presented in two conditions. In one condition,

the change of the final nasal was viable ('leam bacon'), while in the other condition, the change was unviable ('leam gammon'). If the change was viable, both changed and unchanged tokens would lead to a comparable priming effect. In the unviable-context condition, however, the changed tokens ('leam gammon') yielded reduced priming effects compared to the unchanged tokens. This speaks for the existence of a phonological-inference mechanism that infers the intended 'deep form' from the perceived 'surface' utterance. However, these data do not allow us to pinpoint the locus of these effects, because priming in a lexical-decision task is a measure that leaves open many possibilities for interpretation, including strategic effects (cf. Neely, 1991).

The strongest evidence for a regressive-inference account stems from a phoneme monitoring study by Gaskell and Marslen-Wilson (1998). They used phrases containing phonologically viable and unviable changes. While listening to these phrases, participants performed a phoneme-monitoring task, monitoring for a segment with an alveolar place of articulation. Participants were more likely to (mis)detect a target with an alveolar place of articulation when the context allowed for an assimilation of the segment in question. This effect was present in words and in nonwords. That is, participants were more likely to report hearing a /t/ in the phrase 'prayp bearer' than in the phrase 'prayp carrier.' In the phrase 'prayp bearer,' the final /p/ in the nonword 'prayp' could have a 'deep form' with a final /t/. (A /t/ can be assimilated and become a /p/ if the following segment is a /b/.) This effect was moderated by the lexical status of the carrier words. Misdetections of coronal segments, indicative of regressive phonological inference, were more likely in word targets than in nonword targets. This led Gaskell and Marslen-Wilson to extend the model of Gaskell et al. (1995) by allowing lexical influences on phonological inference. In the earlier model, phonological inference was assumed to be purely pre-lexical. The presence of an effect in nonword targets, however, points in the direction of a pre-lexical nature of the phonological-inference mechanism.

This interpretation needs further consideration. Warren (1971, 1999, 2000; Warren, Bashford, & Gardner, 1990) has argued that "listeners do not, and indeed cannot, perceive phonemes in running speech directly, but that their presence is inferred following prior identification of larger units" (Warren, 1999, p. 172). That is, according to this view, phoneme detection in a phoneme monitoring task occurs following and not prior to phonotactic and lexical processing of the stimulus. Indeed, evidence from event-related-potential studies has shown that lexical information is available before phoneme monitoring can be performed (Thierry, Doyon, & Démonet, 1998). Therefore, it needs further clarification to what extent the lexical influences on the perceptual compensation for phonological assimilation were genuine or task-related in the study of Gaskell and Marslen-Wilson (1998).

The present experiments

The evidence in favor of a regressive-inference account is mainly based on studies employing cross-modal priming and phoneme-monitoring tasks. Gow (2001) argued that the pattern of results in favor of a regressive-inference mechanism is due to a disruption of lexical processing: On hearing, 'leam,' participants may activate the lexical candidate 'lean' and expect a following word to start with a labial. As this expectation is not fulfilled, lexical processing is disrupted with increasing reaction times as a consequence.

To avoid the possibility of such alternative interpretations we used a two-alternative forced-choice (2AFC) task with feedback. In this task, participants had to decide whether a Dutch word ('tuin' [garden]) was pronounced canonically or changed with a word-final /m/. It may be argued that this 2AFC task is less open to alternative interpretations such as 'disruption of lexical processing' (Gow, 2001). Although the 2AFC task does not exclude the possibility of lexical influences (e.g., Ganong, 1980), it may be argued that neither phoneme awareness nor lexical processing are necessary in order to perform the task. This can be derived from the fact that the 2AFC is easily adaptable for non-literate pre-school children (e.g., Nittrouer, 1996) and even non-human species (e.g., Lotto, Kluender, & Holt, 1997; Schulze & Scheich, 1999; Sinott & Sapirota 2000).

In order to investigate the perception of assimilation tokens with a 2AFC task, words were presented in isolation or in a phrase that allowed or did not allow assimilation. According to a regressive inference account, the change ('tuin' → 'tuim') should be easily detected in isolation, but would be difficult to detect in a viable context. This hypothesis was investigated in Experiment 1. Experiments 2 and 3 also used a 2AFC task in order to test alternative interpretations of the results of Experiment 1. Experiment 2 tested whether masking contributes to the perceptual compensation for place assimilation. Experiment 3 tested whether lexical processes influence the perception of place assimilation.

Finally, we tested whether evidence for perceptual compensation for phonological assimilation can be found when participants are only listening passively to viably and unviably changed phrases. In Experiment 4, we therefore used a paradigm that does not involve an active decision by the participant: the passive-oddball paradigm. In this paradigm, participants hear a train of stimuli that consists of a standard stimulus (e.g., 85%) and a deviant stimulus (e.g., 15%). Participants do not have to react to the stimuli and are either reading a book or watching a silent videotape. When the electroencephalogram is measured and averaged—time-locked to the stimulus onset—to obtain event-related potentials (ERPs), the ERP to the deviant stimulus is more negative than the ERP to the standard stimulus starting around 100 ms after stimulus

onset. This negativity is called mismatch negativity (MMN) (see, e.g., Näätänen, 1992, 1995; Schröger, 1998). The MMN is interesting for the present purposes, because the MMN seems to reflect the perceptual and not the physical difference between the standard-deviant stimulus pair (Näätänen & Winkler, 1999). Näätänen, Schröger, Karakas, Tervaniemi, and Paavilainen (1993) showed that a stimulus pair consisting of two complex tone stimuli induces a MMN only when participants are able to discriminate between the two stimuli with a high level of accuracy. In this study, a group of participants learned to discriminate between two stimuli over the course of the experiment. These participants did not show a MMN in the first block of passive listening, but did so in the last block after their discrimination performance had increased. This shows that the MMN does not reflect the physical but rather the perceptual difference between the standard and the deviant. Along similar lines, Näätänen et al. (1997) obtained a comparable result with vowels as stimuli. They measured the MMN to two vowel stimuli in Finnish and Estonian participants. The main result of this study showed that the MMN was enlarged, if both standard and deviant were vowel prototypes in the native language of the listener. Again, the same physical difference between a standard and a deviant gave rise to different MMNs depending on the native language of the listener (see also Winkler et al., 1999).

These findings show that the MMN is an interesting tool for our purposes, because it is sensitive to the perceptual distance between a stimulus pair. If there is a regressive compensation mechanism that compensates for phonological variation, this should reduce the perceptual distance between a canonical and a changed utterance. That is, the perceptual distance between 'tuinbank' [garden bench] and 'tuimbank' should be smaller than the distance between 'tuinstoel' [garden chair] and 'tuimstoel.' In the first example, the change of the /n/ to a /m/ is viable according to the assimilation rule. In the second example, however, the change is not viable. Accordingly, the MMN should be smaller to a stimulus pair 'tuinbank - tuimbank' (viable condition) than to a stimulus pair 'tuinstoel - tuimstoel' (unviable condition). This was the hypothesis we set out to test in Experiment 4.

Experiment 1

In the first experiment, we used edited natural speech to investigate how participants handle phonological assimilation in the 2AFC task. Participants had to indicate whether the Dutch word 'tuin' [garden] was pronounced canonically or with a changed place of articulation for the final nasal, that is, 'tuim.' This was tested by presenting the target words in isolation, or in a context that did ('...bank') or did not allow ('...stoel') for the assimilation of a word-final nasal.

The fact that the stimulus materials for the present experiment were also to be used in a passive-oddball paradigm later on puts certain constraints on the stimulus material. A MMN may be observed not only if two stimuli differ in their phonological content, but also if they differ in pitch, intensity, or length. This implies that one cannot use arbitrary utterances with and without a change of the final phoneme and use them in a passive-oddball paradigm. As a result one might find the MMN to be smaller in the viable condition, not because of phonological content, but because the two viable utterances 'tuinbank - tuimbank' differed less in pitch contour, length, or intensity than the unviable stimulus pair 'tuinstoel - tuimstoel'. To avoid this potential confounding we used cross-splicing to generate the stimulus materials. Thus, this first experiment also served to test whether cross-spliced materials lead to a stable perception of the place of articulation of a final nasal.

Method

Participants

Eight students of Maastricht University participated in the experiment. Six were female, two were male. None reported any history of hearing loss.

Design

A two-factorial design was constructed, with the first factor Nasal Murmur, measured at two levels: /n/ and /m/ and the second factor Context measured at three levels: no context; a viable context ('bank'), in which assimilation ('tuimbank') may occur in natural speech; and an unviable context ('stoel') in which assimilation does not occur. The three Context conditions were blocked. The no-context condition ([tœyn] 'tuin' and [tœym] 'tuim') was always presented first. The order of presentation of the two following context blocks (viable-context block and unviable-context block) was counterbalanced between subjects. Each block consisted of 40 trials, that is, 20 repetitions of the two stimuli presented in a random order. In order to focus participants' attention on the phonetic details of the stimuli, participants received explicit feedback after every trial.

Stimulus material and apparatus

A male native speaker of Dutch was selected, because of his low F0 (75-80 Hz), which did not vary greatly between utterances. The sonority of the speaker's voice contributed to the naturalness of the cross-spliced stimuli. Multiple pronunciations of 'tuinbank' [tœynbɑŋk], 'tuimbank' [tœymbɑŋk], 'tuinstoel' [tœynstu:l], and 'tuimstoel' [tœymstu:l] were digitally recorded with a sampling rate of 44.1 kHz. Utterances were digitally band-pass filtered (40 - 5200 Hz) and re-sampled at 11025 Hz. Experimental stimuli were created by cross-splicing. First, [tœy] was spliced from an utterance of [tœynstu:l]. From two other utterances of [tœynstu:l] and [tœynbɑŋk], the nasal murmurs were spliced. The nasal murmurs were equated in length and energy (using a root-mean square measure). Due to the constraint of splicing at zero-crossings, the nasal murmurs differed slightly in length. However, this was less than 1 ms. The edited nasal murmurs were concatenated with [tœy], resulting in the experimental stimuli [tœyn] and [tœym] used in the isolation condition.

The Context words 'stoel' [stu:l] and 'bank' [bɑŋk] were each spliced from two other utterances. The [bɑŋk] utterance was produced without pre-voicing. In order to sound natural, a 25-ms period of silence was used to emulate a closure. By cutting 5 ms of the friction noise of the /s/ in [stu:l], the length of this sound was made equivalent to the length of the [bɑŋk] sound, including the silent period (see Table 3.1 for a detailed description of the nasal murmurs and the context sounds). The stimuli were played to the participants by a DOS-operated computer using the software package ERTS (Behringer, 1996). The output of the SoundBlaster soundcard was redirected to 4 loudspeakers (JBL Control 25) within a sound-attenuated booth with a Spirit Folio mixer and amplified by a Yamaha P4050 power amplifier. The maximal sound-pressure level was 70 dB(A).

Table 3.1 Acoustic properties of the stimuli used in the experiments

	Formant 1			Formant 2			Formant 3			Formant 4		
	C	B	A	C	B	A	C	B	A	C	B	A
/n/	360	180	34.7	1500	470	25.3	2170	40	17.1	3460	180	18.7
/m/	357	160	33.9	1308	420	21.4	1960	100	20.5	3010	355	9.3
/s/	1870	520	4.9	2550	285	5.7	3095	270	15.0	3892	290	11.1
/b/	593	77	41.1	1130	175	30.4	1995	60	20.2	3379	30	18.4

Note. Values for the nasal murmurs and the context sounds were derived from formant estimations provided by Praat (Boersma & Weenink, 1999). Number of assumed formants was adjusted manually to achieve a stable solution for each speech sound. C = center frequency in Hz, B = bandwidth in Hz, A = relative amplitude in dB.

Procedure

Participants entered a sound-attenuated booth and were seated facing a monitor and an array of buttons. Two of the buttons were labeled 'tuin' and 'tuim.' Participants were instructed to press one of two buttons upon hearing a target. Instructions stressed accuracy without special emphasis on speed of response. After each trial, participants received visual feedback in the form of the words "right" and "wrong" on the monitor indicating whether their identification had been correct.

Results

For each participant, the percentage of correct identifications was calculated for each condition. The mean percentages of correct identifications are given in Table 3.2. The descriptive data show that performance was near ceiling in the no-context and in the unviable-context condition. In the viable-context condition, however, performance levels dropped. A two factor-repeated measurement analysis of variance (ANOVA) on the percentages of correct identifications with the factors Nasal Murmur (/n/ or /m/) and Context (no context, unviable context: 'stoel,' and viable context: 'bank') revealed a significant influence of Context ($F(2, 14) = 38.49, p < 0.001$). The factor Nasal Murmur was close to significance ($F(1, 7) = 4.79, p = 0.065$) and the interaction between the two factors was significant ($F(2, 14) = 8.19, p < 0.01$). Given the significant interaction, we examined the effect of Context for each of the Nasal Murmurs separately. For the alveolar nasal murmur, Context produced only a trend ($F(2, 14) = 3.1, p < 0.1$). For the labial nasal murmur, Context influenced percentages of correct responses significantly ($F(1, 14) = 56.8, p < 0.001$). Post-hoc Newman Keuls ($p < 0.05$) revealed that performance was worse in the viable-context condition (69.3% correct) than in the two other conditions (98.8% correct for no context and 96.3% correct for the unviable-context condition). These latter conditions did not differ significantly. In sum, the interaction indicates that the participants had an overall tendency to (mis)perceive the changed stimulus in the viable-change condition ('tuimbank') as a canonical pronunciation ('tuinbank').

Table 3.2 Percentages of correct identifications in Experiment 1

Stimulus	Context		
	No Context % correct	Unviable ('stoel') % correct	Viable ('bank') % correct
'tuin'	96.3	96.9	87.5
'tuim'	98.8	96.3	69.3

Discussion

The first point to be stressed is that our stimulus generation method was justified: cross-splicing the nasal murmur led to a stable perception of the place of articulation. In the no-context condition, performance was near ceiling level. This result is worth noting, because it is not self-evident whether cross-splicing nasal murmurs is sufficient for creating a stable percept of the place of articulation. Usually, formant transitions within the vowel part of the signal contribute to the perception of place of articulation in nasals. However, Repp and Svestikula (1988) found that nasal murmur is a more important cue for post-vocalic than for prevocalic nasals. In addition, we purposefully chose the word for this study in order to minimize the formant transition cue. Formant-transitions cues are expected to be weaker in case of high vowels and in case of diphthongs. Formant transitions into the nasal murmur are more salient for low vowels, because this affords more movement of the articulators to achieve the closure necessary for the nasal. Based on, for instance, the window model of speech-sound production (Keating, 1990), formant transitions in diphthongs are constrained in order to convey the diphthongal nature of the sound and leave less room for co-articulatory anticipation of the following segment's place of articulation. Given that the to-be-identified nasal in this experiment was post-vocalic and preceded by a diphthong with a high target, it is therefore not completely unexpected that the nasal murmur turned out to be a sufficient cue for a stable perception of place of articulation.

The stable perception of place of articulation also did not suffer from the introduction of an arbitrary context. Performance was equivalent and close to ceiling level in both, the no-context and the unviable-context condition. However, in the viable context, participants not only made more errors, but also showed a significant tendency to (mis)perceive the stimulus 'tuimbank' as 'tuinbank.' This result supports the assumption that phonological variation is compensated for regressively: An /m/ followed by a stimulus with a labial place of articulation is interpreted as a /n/. As a consequence the stimulus 'tuimbank' is perceived as 'tuinbank.' These

results are consistent with the assumption of a regressive compensation mechanism dealing with place assimilation. It is difficult to see how strategic higher-level processes could influence the current results obtained with a comparably simple task (as Gow, 2001, assumed for the data of Gaskell and Marslen-Wilson, 1998).

One might argue that the current result is compromised by the fact that only two stimuli were used. However, the current result extends earlier similar findings (Gaskell & Marslen-Wilson, 1996, 1998), which employed a large number of stimuli. Therefore, it seems unlikely that the context sensitivity evidenced in the present experiment is due to different reasons than the context sensitivity in these earlier studies. However, it seems worthwhile to consider two alternative interpretations of the presented results. First, amplitude differences between the context words might cause the results. The /b/ in the viable context (bank) has a larger amplitude than the /s/ in the unviable context (stoel), as evident from Table 3.1 and the oscillograms in Figure 3.2. The results might thus not be due to the viability of the assimilation in the different contexts, but may be caused by backward masking or unspecific distraction caused by the stimulus with the larger amplitude. Secondly, the results might be caused by lexical rather than phonological inference. Both utterances 'tuinbank' and 'tuimbank' are legal pronunciations of an existing word. If lexical knowledge about the possible pronunciations of 'tuinbank' [garden bench] influences identification performance, 'tuimbank' stimuli might trigger 'tuinbank' responses. To investigate these alternative interpretations, masking and lexical influence, we conducted another two identification experiments using the same stimuli.

Experiment 2

The purpose of this experiment was to test whether the results obtained in Experiment 1 can be attributed to a perceptual compensation mechanism for place assimilations or are due to masking. Masking could be responsible for the results of Experiment 1, because the context words differed in amplitude of the first segment. The stop in the viable context was higher in amplitude than the fricative in the unviable-context condition (see Table 3.1). The poor performance in the viable-context condition in Experiment 1 might be caused by the large amplitude of the context word's onset. If masking was responsible for the results in the first experiment, then participants should be less accurate when the stimuli 'tuin' and 'tuim' are followed by a context consisting of large-amplitude wideband noises.

To test this possibility, we created three band noises that had the same root-mean square (RMS) as the 25 ms of maximal amplitude of the 'bank' [bench] stimulus. A drop in performance might occur as a consequence of a frequency-specific masking or a general-distraction effect. In

the case of a frequency-specific effect, it is conceivable that the low second formant of /b/ is responsible for a masking effect. If energy in the nasal of the target is reduced in perception due to masking effects, participants might hear an /n/ instead of an /m/, only if the /m/ is followed by a low-frequency masker. On the other hand, it may also be possible that the result of Experiment 1 is due to an unspecific distraction by any large-amplitude context sound. In this case, any wideband noise should lead to a decrease in accuracy in identifying the nasal murmur in 'tuin' and 'tuim.'

Method

Participants

Eight students of Maastricht University participated in the experiment. None reported any history of hearing loss. None of the participants had participated in Experiment 1. Participants were paid for participation.

Design

A two-factorial design was constructed, with the first factor Nasal Murmur, measured at two levels: /n/ and /m/ and the second factor Context measured at five levels: This could be no context, the viable context from the previous experiment ('bank'), and three band noises with energy in 'labial', 'alveolar', and 's-frication' frequency areas. All conditions were blocked.

Stimulus material and apparatus

The speech stimuli and the apparatus were the same as in the previous experiment. Three wideband noises were generated with the help of the software package PRAAT 3.9 (Boersma & Weenink, 1999). A white noise of 100 ms duration was generated. This noise was used as a template and filtered with a Hanning-band window (smoothing 100 Hz) to generate three band noises. A 'labial' noise was generated with energy in the area of 0.5 to 1.5 kHz. (corresponding roughly to the distinctive second formant of /b/). An 'alveolar' noise was generated with energy in the area of 1.5 to 2.5 kHz. A 's-frication' noise was generated with energy in the area of 2.5 to 3.5 kHz. (The spectrum of the friction in the 'stoel' context revealed a maximum at 3 kHz.) After filtering, the band noises were equaled in RMS to the RMS of the maximum-amplitude area of the /b/ in the 'bank' context. Band-noises were concatenated with the target words with a 25-ms silent interval, mimicking the presentation of the 'bank' context.

Procedure

The procedure was similar to the one in Experiment 1. All participants first completed the 'no-context' condition. Then, half of the participants heard the stimuli in the 'bank' context, and continued with the s-frication noise, the alveolar-noise, and the labial-noise context. The other half of the participants heard the stimuli in the order labial-noise context, s-frication noise context, alveolar-noise context, and finally the 'bank' context. These two orders cross-balanced the position of the two probably most difficult conditions: the 'bank' context and the labial-noise context.

Results

The percentages of correct identifications are summarized in 3.3. As in Experiment 1, performance was worst in the 'bank'-context condition. In addition, there was no visible influence of the band-noises on the percentage of correct identifications. These observations were borne out by a repeated-measurement ANOVA with Nasal Murmur ('tuin' or 'tuim') and Context (none, 'labial' noise, 'alveolar' noise, 'friction' noise, and 'bank') as factors. The factor context did reach significance ($F(4, 28) = 17.9, p < 0.001$).

Table 3.3 Percentages of correct identifications in Experiment 2

Stimulus	Context				
	None	Labial Noise	Alveolar Noise	S-friction Noise	'bank'
	% correct	% correct	% correct	% correct	% correct
'tuin'	93.8	95.5	93.1	94.2	85.6
'tuim'	95.6	95.6	94.0	96.4	63.6

The factor nasal murmur was not significant ($F(1, 7) = 3.6, p < 0.1$). However, the interaction of the two factors was significant ($F(4, 28) = 36.1, p < 0.001$). In order to investigate the nature of the interaction, we examined the effect of Context on each level of the factor Nasal Murmur separately. For the alveolar nasal murmur, Context influenced performance significantly ($F(4, 28) = 3.8, p < 0.05$). Post-hoc Newman Keuls ($p < 0.05$) revealed that performance in the viable context condition (85.6% correct) is worse than in all other conditions (range: 93.1% - 95.5% correct). All other comparisons failed to reach significance. For the labial nasal murmur, Context also influenced performance, too ($F(4, 28) = 30.6, p < 0.001$). Post-hoc Newman Keuls ($p < 0.05$) revealed that performance was worse in the viable-context condition (63.6% correct) than in the all other conditions (range: 94.1% - 95.6% correct). These latter conditions did not differ significantly. For both nasal murmurs, performance was worse in the viable-context condition than in the other conditions. We then tested if context effect of the viable-context condition was larger for the labial nasal murmur than for the alveolar nasal murmur. To this end, we computed a 'viable-context effect' by subtracting the percentage of correct responses in the viable-context condition from the mean percentage of correct responses in all other conditions. The decrease of the percentage correct responses in the viable-context condition in comparison to all other conditions is larger ($t(7) = 10.2, p < 0.001$) for the labial nasal murmur (31.9%) than for the alveolar nasal murmur (8.6%). This reflects that participants made most errors in response to the changed stimulus in the viable condition, that is, they were inclined to hear 'tuimbank' as 'tuinbank.'

Discussion

These results revealed that the results obtained in Experiment 1 were not caused by amplitude differences between the viable and unviable-context sounds. None of the three band noises led to a decrease in accuracy. Neither unspecific distraction by large-amplitude sounds nor frequency-specific masking was responsible for the results obtained in Experiment 1. It may

therefore be concluded that the discrimination between final nasals with different places of articulation is not impaired by large-amplitude context sounds presented immediately after the target. The interpretation of the results of both experiments in terms of a perceptual compensation mechanism seems valid. Furthermore, the drop in accuracy in the viable-context condition as well as the tendency to (mis)perceive the stimulus 'tuimbank' as 'tuinbank' also replicated the results of Experiment 1. However, in the present experiment there also was a significant drop in percentages of correct responses for the alveolar nasal murmur. A similar trend was observed in Experiment 1. We refrain from giving an interpretation for this effect until the discussion of Experiment 4, when additional evidence is available. Nevertheless, this experiment shows that the results of Experiment 1 are not due to masking. In the next experiment, we investigated possible lexical contributions to the regressive context effect observed in the two previous experiments.

Experiment 3

In Experiment 1 as well as in Experiment 2, participants tended to (mis)perceive the stimulus 'tuimbank' as 'tuinbank.' The purpose of this third experiment was to investigate the possibility that lexical inference causes those misperceptions. If the lexical information that 'tuimbank' is a legal variation of 'tuinbank' has become available to the listener, then this might trigger a 'tuinbank' reaction in response to a 'tuimbank' stimulus. One possible strategy to rule out this possibility is to use nonwords (cf. Gaskell & Marslen-Wilson, 1998). We took, however, a different road and replicated Experiment 1 with German participants with no knowledge of Dutch. The reason to do so was twofold. First, this allows a direct comparison with Experiment 1 by using identical stimuli. Second, it has been argued that nonwords might access the lexicon by analogy with similar words (Marslen-Wilson & Warren, 1994). It is highly unlikely that German listeners are able to perceive the Dutch stimulus words by analogy with similar words. The German word for garden [gartən] is completely dissimilar to the Dutch word [tœyn]. Moreover, the diphthong in 'tuin' (/œy/) does not exist in the German language. An analogy with any German word is therefore difficult. Even if one assumes that German listeners first assimilate [tœyn] to their native phonological categories (following Best's perceptual assimilation model, 1994), they would arrive at [dɔm] or [tɔm], both being nonwords in German. Consulting a lexical database (CELEX, Baayen, Piepenbrock, & Gulikers 1995) revealed that the nearest neighbor is [nɔm] (Engl. nine), which is acoustically quite different. Thus, the use of the same stimuli as in Experiment 1, but now with German listeners, allows one to estimate possible

lexical involvement using identical stimuli as words with Dutch listeners and nonwords with German listeners. If the results of Experiment 1 will be replicated, it is unlikely that this is due to lexical inference. It has to be noted, however, that German has a rule for nasal place assimilation. Therefore, it may not be ruled out that knowledge of a rule of place assimilation as such may also play a role.

Method

Participants

Ten native speakers of German participated in the experiment. One participant had just started his first lesson in Dutch, all other participants had no knowledge of Dutch. It was verified that this one participant did not know that 'tuin' [tœyn] is the Dutch word for garden. The sample was quite heterogeneous with an age range from 20 to 58 years and quite different professional occupations.

Design, Material, and Procedure

The design and material was identical to the design of Experiment 1. Given that one might expect non-native speakers to show some initial problems in judging the, for them, foreign-language materials, we doubled the length of each block, so that in effect every stimulus was heard 40 times.

Apparatus

The experiments were done with a mobile lab. Stimulus presentation was controlled from a laptop by a DELPHI 5.0 (Borland, 1999) program written by the first author. The sound output of the computer was amplified (Radio Design Labs, ST-PH 1) and played to the participants via headphones (SONY MDR-V900). Experiments were done in a quiet room with only the experimenter and the participant present.

Results

The percentages of correct identifications are displayed in Table 3.4 (see also Figure 3.1). A repeated-measurement ANOVA showed that the percentages of correct identifications were significantly influenced by the context ($F(2, 18) = 11.3, p < 0.005$). In addition, the factor Nasal Murmur was significant ($F(1, 9) = 9.6, p < 0.05$), as was the interaction between the factors ($F(2, 18) = 7.6, p < 0.005$). In order to investigate the nature of the interaction, we examined the effect on Context on each level of the factor Nasal Murmur separately.

Table 3.4 Percentages of correct identifications in Experiment 3

Stimulus	Context		
	No Context	Unviable ('stoel')	Viable ('bank')
	% correct	% correct	% correct
'tuin'	97.0	98.3	92.8
'tuim'	97.5	97.0	72.0

For the alveolar nasal murmur, Context produced only a weak trend ($F(2, 14) = 2.1, p < 0.15$). For the labial nasal murmur, Context influenced percentages of correct responses significantly ($F(1, 14) = 11.4, p < 0.005$). Post-hoc Newman Keuls ($p < 0.05$) revealed that the level performance was lower in the viable-context condition (72.0% correct) than in the two other conditions (97.0% correct for no context and 97.5 % correct for the unviable-context condition). These latter conditions did not differ significantly. This interaction revealed that the German listeners replicated the performance of the Dutch subjects participating in Experiment 1; they also made most errors in (mis)perceiving 'tuimbank' as 'tuinbank.'

Post-hoc between-experiment comparisons

Combining the Dutch and German results

To allow a more detailed comparison, the results of Experiment 1 and 3 were combined in a repeated-measures ANOVA with Native Language as a between-subjects variable. The descriptive data are presented in Figure 3.1, showing similar trends in both samples. In the statistical analysis, the factor Context is significant ($F(2, 32) = 35.7, p < 0.001$). The factor Nasal Murmur was also significant ($F(1, 16) = 13.6, p < 0.005$), as was the interaction between Nasal Murmur and Context ($F(2, 32) = 15.0, p < 0.001$). However, neither the between-subject variable Native Language nor its interaction with any of the within-subject variables were significant (all $F_s < 1$). Familiarity with the stimulus words did not have any measurable influence on identification performance.

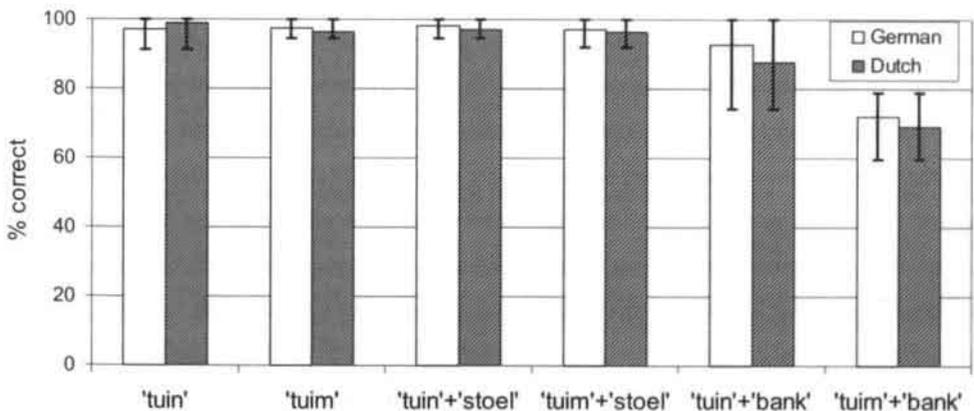


Figure 3.1 Mean percentages of correct identifications in Experiment 1 (Dutch sample) and Experiment 3 (German sample).

Testing for transitory effects

The results of the previous experiments indicated that participants tend to (mis)perceive the stimulus 'tuimbank' as 'tuinbank.' During the experiments, we often observed that participants were initially puzzled when negative feedback was provided in the viable-context block. This might have led to a change in performance over the course of the critical block. Therefore it was tested whether accuracy improved over the course of the critical viable-context block. The accuracy for the first and second half of this block were compared. Accuracy did not differ between the first and second half of the viable-context block in any of the experiments (Experiment 1: 77.5% vs. 79.4%, $t < 1$, Experiment 2: 78.9% vs. 70.0, $t(7) = 2.0$, $p = 0.09$, Experiment 3: 81.3% vs. 83.6%, $t < 1$). On average, accuracy is almost identical for the first (79.2%) and second (77.7%) half of the viable-context block over all three experiments ($t < 1$).

Discussion

The results render it unlikely that the relatively poorer identification in the viable-context condition and the bias towards perceiving a final /n/ in response to the stimulus 'tuimbank' result from lexical inference. The results of the German participants did not differ from the results of the Dutch participants. The results were not only qualitatively similar, but also quantitatively comparable. A combined statistical comparison of the present experiment with Experiment 1 showed that neither the main effect of Native Language nor any interactions involving this factor showed a trend towards significance. This is in slight contrast with the results of Gaskell and Marslen-Wilson (1998), who found a stronger context effect in the word than in the nonword condition. In our study, however, it did not matter whether the stimuli were perceived as words or nonwords.

In addition to testing for effects of lexical inference, we also evaluated the possibility that participants only needed to adjust their strategy in the viable-context condition. If this were the case, performance in the viable-context condition block should be confined to the first trials. However, in all three experiments an equivalent number of errors was made in the first and second half of the critical block. These data indicate that the effect was not transitory but was robust over the course of the experiment. Though participants clearly showed an awareness response to negative feedback in the viable-context block, they were not able to devise a strategy such that it led to a level of performance near ceiling level as in the other blocks of the experiment. An interpretation of the results of Experiment 1 favoring a regressive-compensation mechanism, free of perceptual masking and lexical inference effects, might be compromised by the fact that the experimental task also recruits attentive and decision making processes.

Experiment 4

In this experiment, a rather different method is used to provide evidence for regressive compensation for phonological assimilation. The evidence previously brought forward to substantiate this claim is based on data collected by means of experimental tasks that always require an active decision on the part of the subject (Gaskell & Marslen-Wilson 1996, 1998, Gow, 2002a). This also holds for our own Experiments 1 to 3. It may therefore be argued that compensation for phonological assimilation mainly occurs as a consequence of decision processes and not as a result of an automatic perceptual mechanism. Outside the laboratory, listeners do not actively decide on the place of articulation of word-final nasals. Therefore, this fourth experiment aims to find evidence for the perceptual compensation for phonological variation by means of a task that does not afford a decision by the participant. This can best be achieved by using a passive-oddball experiment while measuring the MMN (see the General Introduction for the rationale).

Two different trains of stimuli were used. One train of stimuli consisted of a canonical form and a viable variant ('tuinbank' and 'tuimbank'), the other train of stimuli of a canonical form and an unviable variant ('tuinstoel' and 'tuimstoel'). The acoustic difference between the standard and deviant stimulus within each pair was identical. If we find a difference in the size of the MMN elicited in a passive-oddball paradigm, such that the MMN to the viable-context pair is smaller than the unviable-context pair, this provides evidence for the assumption that there is a compensation mechanism for phonological assimilations that is perceptual in nature.

Method

Participants

Eight student of Maastricht University took part in the experiment, five females and three males. Participants were paid for participation. None of the participants reported a history of hearing problems. None of the participants had participated in any of the previous experiments.

Design

As in the previous experiments, a two-factorial design was used. The first factor was the identity of the nasal murmur (/n/ or /m/). The canonical pronunciation with an alveolar nasal murmur was used for the standard stimulus and the labial nasal murmur for the deviant stimulus of the oddball series. The second factor was the Context (viable: 'bank' vs. unviable: 'stoel'). The stimuli were presented in four blocks. Each block consisted of 800 stimuli (83% standards and 17% deviants) and lasted about 15 minutes. Blocks alternated between the viable context (standard: 'tuinbank', deviant: 'tuimbank') and the unviable context (standard: 'tuinstoel', deviant: 'tuimstoel'). Presentation order was counterbalanced across participants.

While the ERPs were being measured, participants did not react overtly to the stimuli and therefore did not receive any feedback. Therefore, we administered an identification post-test after the passive-oddball task. In this task participants listened to the target stimuli in the no-context as well as in the viable and unviable context. In this post-test, no feedback was given in order not to deviate from the ERP experimental design.

Material, Apparatus, and Procedure

The stimulus material and apparatus were the same as in Experiment 1. After preparation for EEG-recording was finished, participants were seated in a sound attenuated room and instructed to relax and watch a silent movie. Meanwhile, they heard a train of stimuli with 1.15 s between the onsets of consecutive stimuli. The stimuli had a duration of 617 ms. After completing the four blocks, participants judged the stimuli in a 2AFC task (see Experiment 1) without feedback.

Electrophysiological recording and data reduction

The nose-referenced electroencephalogram (EEG) (0.1-125 Hz, sampling rate 256 Hz) was recorded with a 32-channel electrode cap covering frontal, central, temporal, and parietal scalp areas. Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supra-orbital ridge of the right eye. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. All electrode impedances (EEG and EOG) were kept below 5 kOhm. Event-related potentials (ERP) were obtained by averaging the EEG time-locked to the onset of the sound. The ERPs were band-pass filtered digitally from 1 to 30 Hz. The signal was normalized using a baseline that was calculated using the mean amplitude from 30 ms before stimulus onset to the onset of the mismatching nasal murmur, which was 170 ms post-stimulus onset (cf. Näätänen et al., 1993). Epochs exceeding $75 \mu\text{V}$ on any of the channels (including the eye channels) were rejected, leading to a rejection of about 30 % of the trials. The individual ERP averages were obtained from 272 epochs in each condition by using all deviant stimuli and the same number of standard stimuli randomly drawn from all standard stimuli, excluding standards directly following deviants. The MMN was quantified in the following way (cf. Schröger, 1998). An area measure was obtained from the electrode location Fz, which is the most common electrode for MMN measurement. Mean amplitude was calculated for a 100 ms window starting 100 ms after the onset of the mismatching nasal murmur. These values were used for statistical analysis.

Results

Electrophysiological measurements

The grand-average waveforms from a selection of electrodes—giving an impression of the scalp distribution—are displayed in Figure 3.2. There is a clearly defined negative deflection for the deviant stimulus in the unviable-context condition, but not in the viable-context condition. In the viable-context condition, the mean-amplitude measurements were $-0.23 \mu\text{V}$ for the standard stimulus and $-0.38 \mu\text{V}$ for the deviant stimulus. This leads to a MMN of $-0.15 \mu\text{V}$, which was not significant ($t < 1$). In the unviable-context condition, the mean voltage values were $0.34 \mu\text{V}$ for the standard stimulus and $-0.75 \mu\text{V}$ for the deviant stimulus. This equals a MMN of $1.09 \mu\text{V}$, which was significant ($t(7) = 4.55, p < 0.005$). To test whether the two MMN waves differed from each other, a two-factor repeated measure ANOVA was performed with the factors Nasal Murmur (standard /n/ vs. deviant /m/) and Context (viable vs. unviable). The factor Nasal Murmur was significant ($F(1, 7) = 10.0, p < 0.025$), indicating an overall MMN. The factor Context did not reach significance ($F < 1$). Most importantly, the interaction between factors was significant ($F(1, 7) = 9.9, p < 0.025$), indicating that the change in nasal murmur affected the electrophysiological response differently depending on the viability of the change.

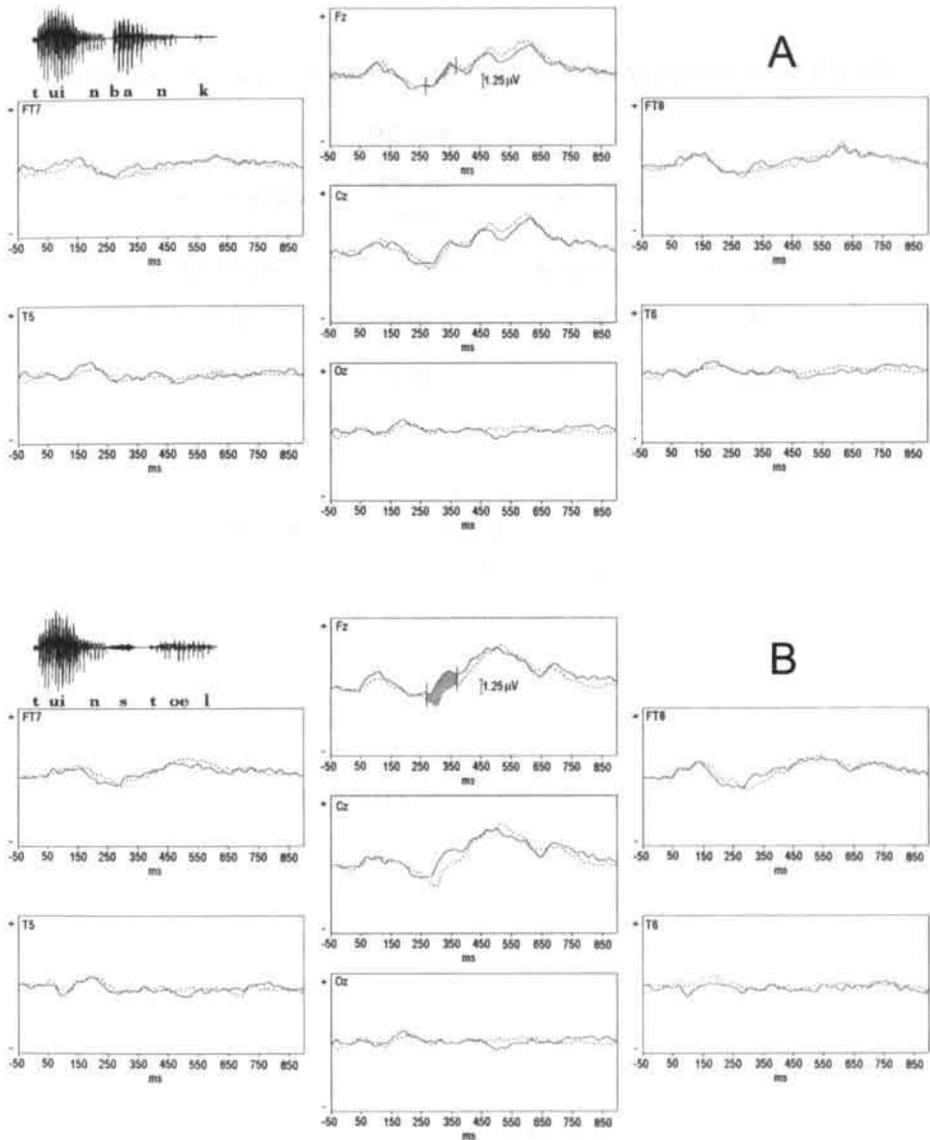


Figure 3.2 Grand average waveforms from Fz, Cz, Oz, FT7, FT8, T5, and T6. The electrodes are arranged analogously to their position on a head that is 'looking' towards the top of the page. The upper part (a) shows the ERPs in the viable-context ('bank') condition, the lower part (b) shows the ERPs to the unviable-context ('stoel') condition. The solid lines represent the standards ('tuinbank' and 'tuinstoel'), the dotted lines represent the deviants ('tuimbank' and 'tuimstoel'). In the upper left corners, an oscillogram of a sound used is displayed as it aligns in real-time with the ERPs. The gray shadings indicate the area in the Fz waveforms used for the estimation of the MMN.

Table 3.5 Percentages of correct identifications in the post-test in Experiment 4

Stimulus	Context		
	no context	unviable ('stoel')	viable ('bank')
	% correct	% correct	% correct
'tuin'	96.3	99.4	96.9
'tuim'	96.3	96.9	35.0

Behavioral post-test

The mean percentages of correct identifications are displayed in Table 3.5. A repeated-measurement ANOVA confirmed that Context did influence performance ($F(2, 14) = 92.7$, $p < 0.001$). The factor Nasal Murmur was also significant ($F(1, 7) = 68.3$, $p < 0.001$), as was the interaction between the two factors ($F(2, 14) = 57.9$, $p < 0.001$). In order to investigate the nature of the interaction, the effect of Context was evaluated separately for the alveolar and labial nasal murmur.

Context did not influence performance for the alveolar nasal murmur ($F < 1$), but did so for the labial nasal murmur ($F(2, 14) = 87.4$, $p < 0.001$). Post-hoc Newman Keuls ($p < 0.05$) revealed that performance in the viable context condition (35%) was worse than in the unviable-context condition (96.9%) and the no-context condition (96.3%). These latter two conditions did not differ significantly.

Discussion

In this experiment, we found evidence for a regressive compensation for phonological variation by using a passive-oddball paradigm in which participants are not required to make active decisions. The MMN was significant only for the unviable deviant stimulus, but not for the viable deviant, despite the same acoustical difference between the viable and unviable deviants and their respective standards. It should be noted here that this is the first report to show that the MMN reflects auditory context in speech perception.

The main result of this fourth experiment is that the MMN in the unviable-context condition is significantly larger than the MMN in the viable-context condition. If one interprets the MMN as a measure of perceptual distance (see General Introduction), the main result translates into the finding that a canonical phonological form and a viable change are separated by less perceptual distance than a canonical form and an unviable change. This result provides

strong evidence for the existence of regressive context effects in the perception of place assimilation.

On a general level, MMN results are ambiguous with regard to the level at which the observed context sensitivity is located. Previous studies indicated that the MMN is sensitive to phonological (Phillips et al., 2000) and lexical (Pulvermüller et al., 2001, but see Wunderlich & Cone-Wesson, 2001) levels. However, influences from lexical processes seem rather unlikely in our case given the timeframe of the MMN. The MMN is already measurable immediately at the onset of the context word. At this point in time, participants heard [tœyns...] or [tœynba...]. It is thus not yet clear that the complete stimulus is an existing compound noun. Therefore, the present experiment provides converging evidence for a regressive compensation mechanism for place assimilation that is perceptual in nature and very likely pre-lexical.

The behavioral post-test yielded two additional results. In contrast to the previous experiments, no feedback was provided here. First, the context effect in the viable condition for the labial nasal murmur was much larger than in the previous experiments. Second, there was no trend for a drop in accuracy for the alveolar nasal murmur. Such a drop was significant in Experiment 2 and was observed as a trend in Experiments 1 and 3. In the discussion of Experiment 2, we promised to discuss this here. Previously, we noted that the feedback initially elicited a puzzled reaction by participants when performing the viable-context block. It seems likely that feedback led participants to adopt a strategy in which they tried to 'balance' their answer over the possible answer categories; given that negative feedback is mostly a consequence of choosing the canonical-form response. Although this strategy neither obliterated the effect that most errors were made in reaction to the viable assimilated form nor led to an increase of the level of performance over the course of the block, it induces some errors for the stimulus with the alveolar nasal murmur, too. The assumption that the errors arising from responding 'tuim' in case of the [tœynban̩k] stimulus might be due to the attempt to prevent the 'opposite' error (responding 'tuin' in case of the [tœymban̩k]) is buttressed by the fact that most participants from Experiments 1 to 3 (23 out of 26, 88.5%) made their first error in reaction to the [tœymban̩k] stimulus. This is significantly different ($\chi^2 = 5.1$, $p < 0.025$) from the 72.7% that would be expected if all errors observed would be distributed equally over the blocks.

General Discussion

Four experiments were conducted to investigate how listeners cope with the variation caused by place assimilation in continuous spoken word recognition. In the first experiments,

participants had to indicate whether the Dutch word 'tuin' [Engl. garden] was pronounced canonically or with a changed place of articulation for the final nasal, that is, [tœym]. This was tested by presenting the target words in isolation, or in a context that did ('...bank', Engl. bench) or did not allow ('...stoel', Engl. chair) for the /n/ to /m/ change of a word-final nasal. Identification performance was near ceiling in both the no-context and the unviable-context condition. However, in the viable context, participants not only made more errors, but also showed a significant tendency to (mis)perceive the stimulus 'tuimbank' as 'tuinbank'. Two more behavioral experiments were conducted, to control whether these results were (partly) caused by masking or lexical top-down effects. Experiment 2 showed that large-amplitude noise bands presented directly after the target stimulus did not lead to a decrease in performance. This rules out the possibility that the differences in amplitude between the viable context ('bank') and the unviable context ('stoel') caused the difference in identification accuracy. In Experiment 3, German participants listened to the Dutch utterances, which sound unlike any German word, and the original pattern of results was replicated. This makes it unlikely that the results can be attributed to some kind of lexical inference.

Finally, Experiment 4 ruled out the possibility that the effects were attributable to attentive and/or decision processes. The MisMatch Negativity for the viable standard-deviant pair 'tuinbank'-'tuimbank' was smaller than for unviable pair 'tuinstoel'-'tuimstoel.' In addition, the fact that the MMN was measured in the same timeframe in which the context words were presented indicates that the lexical status of the stimulus hardly influences the context-sensitive compensation mechanism.

While both the electrophysiological and the behavioral results indicate a pre-lexical mechanism for perceptual compensation for phonological assimilation, the precise level at which the perception of place assimilations is vulnerable to context effects needs further clarification. Gaskell and Marslen-Wilson (1996, 1998) proposed a regressive phonological inference mechanism working on abstract phonological representations. This implies that an assimilated utterance such as [tœymbank] is perceived 'correctly'—that is indicating a labial place of articulation for the nasal—at auditory and/or phonetic levels of processing. Currently, however, there is no evidence indicating the supposed 'context-insensitivity' of earlier auditory and phonetic processing stages. It is noteworthy, that even in the present experiments, in which explicit feedback directed participants' attention on acoustic and phonetic details, participants were still influenced by context, and this effect did not dissipate over the course of the present experiments. Although this does not preclude that other measures may show early, context-insensitive processing of place assimilations, it seems worthwhile to consider the possibility of

context sensitivity at an auditory (e.g., Lotto & Kluender, 1998; Lotto, et al., 1997) or phonetic (Fowler, Brown, & Mann, 2000) level more closely.

At least two lines of reasoning can be put forth in favor of such an assumption. First, a compensation mechanism at a sub-symbolic level may be better suited to deal with some ecological facets of assimilations. In Dutch, nasal place assimilation happens in only 60% of the cases where phonology would allow it (Van Heuven & Van Berg, 1982). In addition, Gow and Hussami (1999) reported sub-categorical differences between intended and assimilated labials (see also Nolan, 1992). Hence, place assimilations are optional and often only partial in speech-sound production. In some respect, this is mirrored by our finding that participants were significantly beyond chance level in recognizing the 'm' in [tœymbaŋk], but far from perfect. This suggests that compensation for place assimilation is probably only partial, as observed in other cases of context-dependent speech perception (Beddor & Krakow, 1999; Fowler & Brown, 2000). Therefore, the equation might be that partial compensation by the listener complements partial assimilation by the speaker to achieve successful communication. In its current form, the model of phonological inference assumes that compensation for assimilation is an all-or-none phenomenon. As such, it is difficult to account for graded effects. Although graded effects may be accommodated by allowing phonological inference to be a graded instead of an all-or-none phenomenon, a sub-symbolic mechanism may still be more efficient. At a sub-symbolic level, auditory/phonetic detail, information about speech rate and the amount of hypo-articulation, is still available. Such variables may help to differentiate labial segments from coronal segments that have been assimilated by labials.

A second argument in favor of a sub-symbolic locus of compensation for assimilation stems from proposals in the field of phonology (Hura, Lindblom, & Diehl, 1992; Seo, 2001; Steriade, 2001). These linguistic approaches suggest that assimilation rules are 'perceptually licensed,' that is, attested assimilation rules tend to lead to perceptually inconspicuous changes from the canonical form. That is, the changes in speech production (assimilation rules) are not independent from the perceptual abilities of the listener. However, the invariance problem caused by place assimilation has almost without exception been treated as a purely perceptual problem, comparable to the problem of color constancy in spite of constantly changing illumination conditions during the day. An alternative account suggests that languages, and therefore assimilation rules, have an evolutionary trajectory of their own, during which they adapt to the properties of the human—in the present case—perceptual system (see, e.g., de Boer, 2000; Deacon, 1997; Steels, 1999). As context sensitivity is a general perceptual principle (Warren, 1999), it might be expected that the shaping of place assimilation rules made use of the

already present context-sensitive mechanisms for dealing with invariance problems (Lotto et al., 1997; Mitterer, 2000).

Very recently a related account of compensation for assimilation was proposed, which supports our suggestion of a sub-symbolic locus for the processing of assimilated speech (Gow, 2002a). This feature-parsing account builds on the evidence that the assimilated segment bears information for both a labial and the original alveolar place of articulation (see above). Thus, the /m/ in [tœymbaŋk] actually bears feature cues for an alveolar and a labial nasal. Compensation for articulation is then achieved by parsing the cues for the labial place of articulation from the assimilated segment (the /m/ in [tœym]), ascribing this information to the assimilating segment (the /b/ in [baŋk]). After this feature-parsing, the assimilated segment only bears cues for the intended alveolar place of articulation. Feature-parsing is assumed to rely on principles of auditory grouping (Bregman, 1990), which de-convolutes the overlapping feature cues.

In principle, this feature-parsing account seems compatible with the kind of sub-symbolic inference account we propose. Gow (2002a) assumes that the perception of an assimilated segment is not an all or none, but a graded phenomenon based on Gestalt principles. As Gow admits, his 'feature-parsing' account still leaves open the precise nature of the mechanisms that help us overcome the invariance problem caused by assimilatory processes. Therefore, in terms of Marr (1982), 'feature parsing' may be considered a description at a computational level, while the mechanisms proposed for context effects in speech perception can be considered descriptions at an algorithm level. Indeed, conceptions of general auditory contrast effects (Lotto & Kluender, 1998), direct perception of articulatory gestures (Fowler et al., 2000), or covariant learning on an acoustic-phonetic level (Holt et al., 2001) may all constitute algorithms that contribute to the computational goal of feature parsing. Future research should indicate in how far these conceptions of context sensitivities in speech perception may be able to explain context effects in the perception of place assimilations.

Place Assimilations and the Interplay between Speech Perception and Phonology¹

Several studies have indicated that phonological assimilations in fluent speech are processed in a context-sensitive way, similar to the finding of compensation for coarticulation. Psycholinguistic approaches attributed this perceptual compensation for phonological assimilation to a specific adaptation of the listener to the phonological assimilation rules of his native language. Some linguists, however, have argued that speech production, and hence the range of possible assimilation rules, is constrained by the general perceptual abilities of listeners. If one adopts this view, it follows that listeners should be able to compensate for phonological assimilations independent of experience with an assimilation rule. In order to evaluate the role of experience with a phonological assimilation rule in compensation for assimilation, we presented Dutch utterances with nasal place assimilation to Portuguese listeners. The Portuguese language has no rule of nasal place assimilation. Results indicate that general perceptual principles as well as specific language experience both contribute to compensation for phonological assimilation.

¹This is an enlarged version of Mitterer, H., Castelo-Branco, M., & Blomert, L. (submitted). The Interplay between speech perception and phonology: Evidence from the perception of assimilated word forms.

Speech perception in general is context sensitive. Recently, several studies have shown that context sensitivity is also critical in accounting for variance in the speech signal caused by (postlexical) place assimilation. In Dutch, for instance, the word for 'garden,' [tæyn] may be pronounced as [tæym], if the next segment is a labial obstruent (cf. Booij, 1995). The perceptual system cannot just disregard such variation, because it has to differentiate other forms based on the place of articulation of a word-final nasal, such as the Dutch words [dæyn] and [dæym], meaning 'dune' and 'thumb'. It has been shown that listeners are able to perceptually compensate for the variance in the speech signal caused by phonological assimilations by using the phonological context of the assimilated phonemes (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2002a).

The previous chapter provided a straightforward example for perceptual compensation for nasal place assimilation (NPA) in Dutch. In a word-identification experiment, *Dutch* participants had to indicate whether the Dutch word for 'garden' [tæyn] was pronounced canonically (with an alveolar word-final nasal) or, in a changed form, with a word-final labial nasal [tæym]. The tokens were presented in three conditions. In a control the target tokens [tæyn] and [tæym] were presented in isolation. In two context-conditions, these targets were presented in compound nouns, that is, concatenated with a context word. In one of these compound nouns, [tæynbank] (Engl. 'garden bench'), the form with an /m/—[tæymbank]—may arise as a consequence of the application of NPA. In the other compound noun, [tæynstul] (Engl. 'garden chair'), however, the form with /m/ is not allowed. When asked to identify the targets as either [tæyn] or [tæym], participants performed near ceiling level in the isolation condition as well as in the control condition with the compound noun [tæy?stoel]. However, participants often failed to correctly identify the viably changed form [tæymbank]. This form was often identified as the canonical form [tæynbank].

In a cross-linguistic study (see chapter three), we investigated whether lexical factors contributed to compensation for assimilation. With this aim, the *Dutch* stimuli were presented to *German* listeners. A query of a lexical data base (CELEX, Baayen, Piepenbrock, & Gullikers, 1995) showed that [tæyn] is quite dissimilar to any German word. German listeners were, like the Dutch listeners, asked to indicate whether this stimulus, a German nonword, ended with an alveolar nasal [tæyn] or a labial nasal [tæym]. The results obtained with German participants were equivalent to the result obtained with Dutch participants: Stimuli differing in place of

articulation of the final nasal condition were perceived as being quite different from each other in the unviable-context. The viably changed stimulus, however, was often perceived as the canonical form.

This cross-linguistic comparison indicated that lexical knowledge has only a minor role to play in perceptual compensation for phonological assimilation. However, it leaves open the question whether there is an influence of implicit phonological knowledge. Both the Dutch and the German language have a rule for NPA. Therefore, it remains an open question in how far experience with a rule of NPA may contribute to perceptual compensation for this assimilation. It was the purpose of the present study to investigate whether experience with NPA is necessary for perceptual compensation for NPA.

Two quite different approaches make different predictions with regard to the role of specific language experience in perceptual compensation for phonological assimilation. In a psycholinguistic tradition, Gaskell and Marslen-Wilson (1996, 1998) assume that experience with an assimilation rule is crucial for perceptual compensation. In their model of perceptual compensation for place assimilation, it was assumed that this compensation is based on the learning of feature co-occurrences. Through experience, listeners learn to map a perceived /m/ onto an underlying /n/ if the subsequent segment has a bilabial place of articulation, like in [tæymbaŋk]. In the beginning of language acquisition however, the perceived /m/ is mapped on a /m/ even though the next segment is a bilabial. According to this view, the listener is in principle well able to hear the difference between a canonical and an assimilated word form. However, during language acquisition, the listener is 'trained to neglect' the difference between a canonical and an assimilated form, just as listeners are trained to neglect differences that are allophonic in their native language. That is, the 'phonology' of the native language influences 'perception.' In this view, experience with NPA is necessary to develop adequate perceptual compensation for NPA.

It is important to note here that Gaskell and Marslen-Wilson (1996, 1998) treat phonological assimilation as a problem imposed on the perceptual system. The problem is analyzed similarly to the problem of color constancy; another instance of an invariance problem: During different times of the day, different lighting conditions lead to a different chromatic composition of the input to the visual system. The perceptual system, however, compensates for the variance caused by lighting conditions in such a way that we perceive objects as having a constant color. The view on phonological assimilations by Gaskell and Marslen-Wilson is in some respect similar. Phonological assimilations change word forms and thereby increase the variance in input conditions that have to be associated with the same referent object. However,

viewing phonological assimilations as a property of the environment to which the listener has to adapt means missing one point: In contrast to the varying lighting conditions over the course of a day, the variability of human speech-sound production is amenable to change. Therefore, it is possible that constraints of the perceptual system shape the phonological assimilations that are allowed in speech production. Several linguists indeed have argued that the observed assimilation rules are constrained by the perceptual abilities of the listener (Kohler, 1990; Steriade, 2001). Steriade (p. 222) put forth a similarity principle as a basic constraint on phonological assimilations: "The likelihood that a lexical representation R will be realized as modified R' is a function of the perceived similarity between R and R'." 'Perceived similarity' here refers to an 'a-priori' similarity that is independent from language experience. Such a view explains, for instance, why alveolar consonants tend to be assimilated regressively in post-vocalic position, but retroflex consonants tend to be assimilated progressively in pre-vocalic position. A retroflex consonant is perceptually more salient in post-vocalic position than in pre-vocalic position. Therefore, an assimilation of a retroflex consonant in pre-vocalic position is a more moderate change than in post-vocalic position. Vice versa, this applies to alveolar consonants, which are more salient in pre-vocalic position but are assimilated in post-vocalic position. This indicates that speakers only produce assimilations that lead to perceptually small changes. Then, no experience with an assimilation rule is necessary for the listener in order to be able to recognize assimilated forms as instances of the respective canonical forms, because in this view, it is the 'perception' which influences 'phonology.' There are different views on how the perceptual system may impose itself on speech production. Steriade (2001) and Lindblom (1990) assume that the speaker has an implicit knowledge of the perceptual abilities of the listener, while others (de Boer, 2001; Ohala, 1990) assume that the perceptual system shapes speech production in a quasi-evolutionary process. Nevertheless, according to both views perceptual compensation for phonological assimilation is based on innate properties of the perceptual system that are independent from language experience.

In the present study, we evaluated the role of experience with an assimilation rule for compensation for this rule. To this end, we tested Portuguese listeners with the same Dutch material previously used with German and Dutch listeners in which NPA has been applied (see chapter three). In the Portuguese language, word-final nasals are not pronounced and just lead to a nasalization of the preceding vowel (Mendeloff, 1969). Therefore, nasals do not surface in coda positions and the Portuguese language cannot have a rule of (postlexical) NPA, which would alter the place of articulation of word-final nasals. To allow a direct comparison with the behavior of listeners who do have experience with NPA, we performed the same word-

identification experiment as in the previous chapter with native speakers of German and Dutch as listeners. Using Dutch stimulus material, Dutch listeners had experience with the word and with the assimilation rule in question, while German listeners had no experience with the word but experience with the assimilation rule. By now probing the perception of these utterances by Portuguese listeners, we test a group of listeners without experience with the word and without experience with the assimilation rule.

Method

Participants

Eight Portuguese subjects participated in the study. Three were male and five were female. Six participants were aged 25 to 30, the other two were between 45 and 50 years old. Participants were paid for their participation. Participants were quite diverse in their schooling levels. The two oldest participants had four years of basic school and were mono-lingual. Three participants got schooling until tenth grade, including some basic English. Three participants were University students, but also had limited experience in English and had not been abroad for an extensive period of time. Five participants spoke some French. Note, however, that in French, nasals only surface before vowels and not in CC sequences in which NPA might apply. No participant had experience with the Dutch language.

Stimuli and Apparatus

As stimuli, we used the Dutch word for 'garden' in its canonical form [tœyn] and in a changed form [tœym]. Stimuli were generated by cross-splicing either an alveolar or a labial nasal murmur to a [tœyn] utterance from which the original nasal murmur had been removed. This yields the two stimuli [tœyn] and [tœym]. These stimuli were used in isolation and or concatenated with either [banjk] or [stu:l] to yield the compound nouns. In the [banjk] context, the changed form [tœynbanjk] is viable. In the [stu:l] context, the changed form *[tœymstu:l] is unviable. The context words were spliced from other utterances of the same speaker and concatenated with the [tœyn] and [tœym] stimuli.

Experiments were run on a standard PC. Stimulus presentation was controlled by a computer program written in the Delphi programming language. A SONY MDR-V900 headphone was attached to the headphone output of the soundcard.

Procedure

The Portuguese language has no minimal pair based on a word-final /m/ vs. /n/ contrast. Given that the current experiment asks the participants to identify a non-native contrast, we slightly adjusted the 2AFC procedure. In the first phase of the experiment, participants were trained to perceive the /n/-/m/ difference in the stimuli used in the present experiment. Participants were instructed that they were going to hear two Dutch words, one ending in an /n/ as in 'nariz' [nose] and one ending in /m/ as in 'maxila' [jaw]. They were instructed to press the 'm' key of the computer keyboard if they judged the word to be ending in an /m/ as in 'maxilla' and the 'n' key if they judged the word to be ending in an /n/ as in 'nariz.' Then they heard a series of 20 stimuli consisting of the stimuli [tœyn] and [tœym] in a randomized order. After responding to a stimulus, they received explicit feedback on whether their choice was correct. If 17 out of 20 responses were correct, the training was terminated. Otherwise, a second series of 20 stimuli was presented. There were maximally five training blocks. After that, training was terminated even if the criterion of 17 corrects responses out of 20 was not met.

After this training procedure, the same [tœyn] and [tœym] stimuli were presented concatenated with the contexts /banjk/ (Engl. bench) and /stu:l/ (Engl. chair) to yield the composites /tœynbanjk/ and /tœynstu:l/ in either the canonical form or with a word-final /m/. Each of the four stimuli arising from the two word-final nasals and two different contexts was presented 40 times. Context conditions were blocked and order of presentation of the blocks was counterbalanced over participants. Within blocks order of presentation was randomized.

Design

The independent variables were Nasal Murmur (two levels: labial and alveolar) and Context (two levels: viable—/baŋk/ [Engl. bench]—and unviable—/stu:l/ [Engl. chair]). The dependent variable was percentage of correct responses.

Results

In the pre-test, two participants reached the criterion of 17 correct responses out of 20 on the first 20 trials. Five participants reached that criterion on the second series of 20 trials, while one participant failed to reach the criterion even after five series of 20 stimuli. However, even this participant performed overall above chance (64 correct answers out of 100, $\chi^2 = 3.97$, $p < 0.05$).

The mean percentage correct responses are displayed in Table 4.1. An ANOVA using the individual percentages correct responses revealed a significant effect of Context ($F(1, 6) = 6.1$, $p < 0.05$). The effect of Nasal Murmur and the Interaction between the two factors was not significant ($F_s < 1$).

We then compared these results to the pooled result of German and Dutch speakers ($n = 18$, see Table 4.1) from chapter three. In a combined analysis of both samples, it is evident that Portuguese listeners overall make more errors ($F(1, 24) = 6.6$, $p < 0.05$). The main effect of Context is significant ($F(1, 23) = 28.8$, $p < 0.001$) and is not moderated by an interaction with Listener Group ($F(1, 23) = 1.0$, $p > 0.1$). The interaction between Nasal and Context is not significant ($F < 1$), but is moderated by an interaction with Listener group ($F(1, 23) = 4.7$, $p < 0.05$). This is due to a significant interaction between Context and Nasal Murmur for the German and Dutch listeners ($F(1, 16) = 14.5$, $p < 0.001$), which was not significant for the Portuguese listeners.

Table 4.1 Percentages of correct identifications

Group	Nasal murmur	Context	
		/baŋk/	/stu:l/
		% correct	% correct
Portuguese	'tuin'	70.6	85.8
	'tuim'	70.5	77.8
Dutch and German	'tuin'	90.4	97.6
	'tuim'	70.8	96.7

Discussion

Chapter three showed that German and Dutch listeners compensate for NPA in a Dutch utterance: They performed near ceiling level in the unviable context condition; but made a sizeable number of errors in the viable context condition. Most of these errors occurred because participants perceived the assimilated word form [tæymbaŋk] as the canonical [tæynbaŋk]. Both listener groups have experience with a rule of NPA in their native language.

Portuguese listeners, who do not have experience with NPA, show a significantly different, although not completely dissimilar, pattern. They make more errors in the viable-context condition than in the unviable-context condition. In this respect, they behave similarly to Dutch and German listeners. However, in contrast with the Dutch and German listeners, these errors in the viable context condition were evenly distributed over the different conditions of the factor Nasal Murmur. That is, Portuguese listeners made a similar number of errors in response to the [tæynbaŋk] and the [tæymbaŋk] stimulus. However, Dutch and German listeners made more errors in perceiving [tæymbaŋk] as [tæynbaŋk] than vice versa. Thus, experience with a rule of nasal place assimilation generates a bias towards hearing alveolar nasals in the viable context, while the sensitivity for different places of articulation in viable and unviable contexts seems not to be affected. How do these results relate to the opposing viewpoints of 'phonology influences perception' and 'perception influences phonology'?

The comparable main effect of Context in all samples shows that a canonical word form and an assimilated version are indeed perceptually similar, independent of experience with the applied assimilation rule. This corroborates the 'similarity principle' proposed by Steriade (2001, p. 222). Even to a naïve listener, the modified form R' ([tæymbaŋk]) is similar to the canonical form R (/tæynbaŋk/). In the original formulation, the similarity principle was based on the perceptual saliency of a segment 'n' that was likely to be assimilated. The current result extends the similarity principle by showing that not only the perceptual salience for the assimilated segment 'n' seems to play a role. In addition, the perceptual salience seems to be influenced by the assimilating segment 'n+1.'

However, the data also partly support the assumption that experience with a given assimilation rule influences the way listeners deal with assimilated utterances. Dutch and German listeners have a bias towards hearing the canonical form in the viable context, which was not evident in the Portuguese data. Without experience with NPA, a listener seems to perceive a nasal as ambiguous with regard to the place of articulation if the nasal occurs in a context in which assimilation might be applied. If the listener, however, has experience with NPA, he or

she is able to perceive the—to the naïve listener—ambiguous segment as the intended alveolar nasal. Therefore, it seems that specific language experience works in concert with general perceptual constraints to achieve adequate perceptual compensation for NPA is achieved.

Assuming a combination of language-specific and general perceptual factors may not only account for our data but also resolve an apparent paradox in the field of phonological-assimilation research. Quite recently, two other researchers also explored perceptual compensation for phonological assimilation cross-linguistically. Gow (2002b) showed that context effects may be independent of experience with an assimilation rule. He tested how Hungarian and English listeners perceived Hungarian phrases, to which a regressive voicing assimilation had been applied. The Hungarian assimilation rule of regressive voicing has no counterpart in the phonology of English. According to his feature-parsing account, listeners must *parse* the cues for the voicing *feature* from the assimilated segment and ascribe it to the post-assimilation context. Successful feature-parsing is then indicated by faster phoneme-monitoring for post-assimilation context, because the voicing value of the post-assimilation context was already signaled by the assimilated segment. This was indeed the case. Both English and Hungarian participants performed better in case of viable assimilations, in which the voicing of the target segment is already signaled by the assimilated segment. Therefore, Gow concluded that language experience is not necessary for successful feature-parsing in assimilated utterances. In contrast, Darcy (2002) found that only native-language listeners or highly proficient second-language listeners compensate for phonological assimilation in a word-monitoring task. Based on this data set, she concluded that language experience is crucial in order to compensate for phonological assimilation.

The interpretation given for our own data may help to resolve the apparent contradiction between these results. The current data show that, even without experience with a given assimilation rule, the perception of assimilated utterances is context sensitive. This may explain the finding of Gow (2002b), who used phoneme-monitoring for post-assimilation context as dependent variable. In a phoneme-monitoring task, participants do not need to recover the original canonical word form. However, the word-monitoring task used by Darcy (2002) forces the listeners to recover the canonical word form. Thus, the apparent contradictions between the data sets from Gow and Darcy may be due to the different nature of the tasks they used. In some respect, we replicated the data of Darcy by showing that Portuguese listeners are less well able to infer the canonical word, if that word form has been altered by Dutch NPA, an assimilation rule without counterpart in Portuguese phonology. However, our data also indicate that the perception of assimilated utterances is context-sensitive, independent of specific

language experience. This general context sensitivity may also lead to a faster recognition of post-assimilation context, as evident in the data of Gow.

Nevertheless, it is still unclear *how* language experience influences the perception of phonological assimilations. Why are Dutch and German listeners able to infer the canonical word form, while Portuguese listeners are not? Lexical knowledge could explain the Dutch results, but not the German results. Another possibility is that the learning of statistical properties of the language may be instrumental in perceptual compensation for phonological assimilation. A survey of the corpus of spontaneous speech collected by Ernestus (2000) indicated that, in casual spoken Dutch, /n/ is much more likely to occur in positions in which NPA may apply than /m/ (see chapter five)². Therefore, the listener may infer the identity of the ambiguous nasal segment using his knowledge of *phoneme transition probabilities*. Note that phoneme transition probabilities are widely acknowledged as an important factor in (speech) perception, that is easily learnable by human (Saffran, 2001) and other primate (Hauser, Newport, & Aslin, 2001) learners. From a psycholinguistic point of view, Norris, McQueen, and Cutler (2000) allow phoneme-transition probabilities to influence perception on an early level, although they do not allow any other higher-level knowledge, such as word knowledge, to influence speech perception at early levels. Therefore, it is not too great a leap to argue that the learning of phoneme transition probabilities may play a role in the learning of perceptual compensation for phonological assimilation.

To conclude: the purpose of the present study was to investigate whether experience with a certain assimilation rule is important in order to be able to compensate for it. Based on the proposals of, for instance, Steriade (2001) it was argued that general perceptual principles may be sufficient in order to account for perceptual compensation for phonological assimilation. The present data indicate that general perceptual principles as well as experience with specific assimilation rules contribute to the perceptual compensation for phonological assimilation. In accordance with the similarity principle proposed by Steriade, assimilated forms are similar to their respective canonical forms independent of language experience. This similarity evokes ambiguity, and specific language experience seems necessary to resolve the ambiguity so that the intended word form can be inferred in spite of variance added by phonological assimilations. This indicates that both 'phonology influences perception' and 'perception influences phonology.'

² An analysis of the German CELEX corpus showed that /n/ is more likely than /m/ in German, too. The wordforms corpus contained 22,871 entries ending on /m/ and 84,243 entries ending on /n/. This difference was enlarged when the frequencies were taken into account. There were 133,360 words ending on /m/ but 1,242,686 ending on /n/. Note, though, that these figures are not based on spontaneous speech.

Length does matter

In speech production, words are altered by phonological assimilation rules. Hence, 'garden' may be pronounced with a word-final /m/ if the segment in the utterance is a labial obstruent (as in 'gardem bench'). The inverse effect is observed in speech perception: a word-final /m/ before a /b/ is interpreted as possibly arising from an /n/. In this chapter, we investigate the role of phonetic detail in this 'perceptual compensation for phonological assimilation.' First, we show that phonetic detail influences the degree of perceptual compensation. Segments with strong cues with regard to their place of articulation are not influenced by context while segments with weaker cues are. Second, we show that the influence of phonetic detail in perceptual compensation is mirrored by actual productions of place assimilations. Intended non-alveolar segments tend to carry stronger cues with regard to their place of articulation than alveolar segments, which may be assimilated to non-alveolar. The results indicate that the mechanisms driving perceptual compensation for phonological assimilations take phonetic detail into account.

In speech perception, quite different acoustic stimuli must be mapped onto one and the same lexical representation. One process that adds to this surface variance is (postlexical) place assimilation. In Dutch, for instance the word for ‘garden,’ [tævn], may be pronounced as [tævm] if the next segment is a labial obstruent, such as in [tævmbaŋk] (Engl. garden bench). The perceptual system cannot just disregard such variation, because it has to differentiate other forms based on the place of articulation of a word-final nasal, such as [dævn] and [dævm], meaning ‘dune’ and ‘thumb.’

Given that the application of assimilations (such as, $n \rightarrow m$) is only allowed in certain contexts (e.g., if the next segment is a labial obstruent), the perceptual system may compensate for such variation in a context-sensitive way. The listener may use implicit knowledge of the context dependencies of assimilation rules in order to recognize the assimilated form as instances of the intended word. Consider the following Dutch examples: The /m/ in [ɑrmstu:l]—meaning ‘armchair’—is interpreted as corresponding to an intended /m/, because the labial place of articulation cannot be ascribed to an assimilation triggered by a subsequent labial. This is because the alveolar /s/ cannot trigger place assimilation to a labial nasal (see Booij, 1995). However, the /m/ in [tævmbaŋk]—meaning ‘garden bench’—is interpreted as possibly corresponding to an /n/, because it is followed by a labial obstruent, which may trigger place assimilation.

Gaskell and Marslen-Wilson (1996, 1998) were the first to show that context does indeed play a role in the perception of phonological assimilations. Their results, obtained with cross-modal identification priming and phoneme-monitoring tasks, indicated that the word “lean” is recognized in ‘*lean bacon*’ but not in ‘*lean salami*.’ That is, in a context that allows the assimilation from ‘*lean*’ to ‘*leam*,’ ‘*leam*’ is recognized as an instance of the lexical representation “lean,” while this is not the case if the context does not allow the assimilation. Given that a similar effect was found for pseudowords, Gaskell and Marslen-Wilson argued that their results indicate that this context sensitivity is the result of a pre-lexical mechanism.

Chapter three provided additional evidence for the context-sensitive perception of place assimilations. In a word-identification experiment using a two-alternative forced-choice (2AFC) task, participants had to identify an utterance as the Dutch word [tævn] (Engl. ‘garden’) or the Dutch nonword [tævm]. While [tævm] is in principle a nonword, it is a possible alternative pronunciation of /tævn/ in the context of a subsequent labial segment, which may trigger nasal place assimilation (NPA). These two targets were presented in three conditions: in isolation, in a compound noun [...baŋk] (Engl. bench) that may trigger NPA, and in a compound noun

[...stʊ:l] (Engl. chair) that cannot trigger NPA. Performance was near ceiling when [tœvn] and [tœvm] were presented in isolation or in compounds in which assimilation was not allowed ([tœvnstʊ:l] versus the unviably changed *[tœvmstʊ:l]). If the context allowed the assimilation, a sizeable number of errors were made than in cases the context did not allow the assimilation. Most of these errors were due to responding with 'tuin' when the stimulus was the viably changed form [tœvmbaŋk]. In addition, chapter three reported electrophysiological responses to viably and unviably changed forms in a passive-listening situation in which participants watched a silent movie and paid no attention to the auditory stimuli. Participants heard two oddball series in which the frequent stimulus (83 %) was either [tœvnbəŋk] or [tœvnstʊ:l] (meaning 'garden bench' and 'garden chair' respectively) and the deviant stimuli (17%) were [tœvmbaŋk] and *[tœvmstʊ:l]. In such an experiment, one can expect to find a MisMatch Negativity in the ERP to the deviant stimulus. The MMN is assumed to reflect the perceptual distance between standard and deviant based on automatic, pre-attentive mechanisms (Näätänen & Winkler, 1999). Such a mismatch component was only observed for the unviably changed *[tœvmstʊ:l] deviant, but not for the viably changed [tœvmbaŋk] deviant. This indicates that compensation for phonological assimilation is an automatic and pre-attentive mechanism.

Thus, several studies provided converging evidence for the assumption that phonological assimilations are processed in a context-sensitive way. Therefore, the question arises how this context sensitivity can be accounted for. Gaskell, Hare, and Marslen-Wilson (1995) proposed a specialized, speech-specific mechanism for handling perceptual compensation for phonological assimilation. The most relevant assumption of Gaskell et al. for the present purpose is that auditory input is categorized phonologically before perceptual compensation for phonological assimilation is initiated. In case of the assimilated utterance [tœvmbaŋk], the last phoneme of the first noun is first categorized as "m" (henceforth, we will refer to this as the 'first-pass analysis'). When the next segment is categorized as a labial obstruent, perceptual compensation for phonological assimilation is initiated. This corrects the place of articulation of the preceding phoneme so that it can lead to the activation of the intended word in the lexicon of the listener. Therefore, this account is coined a 'regressive-inference' account.

This regressive-inference account assumes that the input is first categorized phonologically before the context is taken into account. However, there is little evidence in favor of the assumption that the 'first-pass analysis' of an assimilated utterance is context-insensitive. The results of Gaskell and Marslen-Wilson (1996, 1998) indicate that at the level of lexical access, the influence of context on the perception of assimilated word forms has already been

processed. However, no evidence is provided, which indicates that earlier stages are context-insensitive. It should be noted that, in chapter three, a task was used that invites listeners to rely on auditory or phonetic representations, which should not be influenced by context. In the first experiment of chapter three, participants were asked to indicate whether a word-final nasal was pronounced as /n/ or as /m/ in a 2AFC task. Feedback was provided after every trial. The use of immediate feedback very likely discouraged the use of higher-level representations in doing the task, because this feedback ‘punished’ a compensatory response. That is, responding [tæyn] to the stimulus [tæymbaŋk] led to negative feedback. Despite this feedback, participants still responded in a context-sensitive way. They indicated hearing an /n/ more often in [tæymbaŋk] than in [tæymstu:l]. More importantly, this effect did not dissipate over the course of the experiment, showing that feedback did not improve performance over the course of the experiment. These results indicate that the context sensitivity in the perception of assimilated utterances probably arises at earlier levels of processing than at a dedicated phonological level of processing. It might be argued, though, that the 2AFC task is not sensitive enough to probe an auditory/phonetic first-pass analysis of assimilated utterances. Therefore, these experiments did not allow us to refute the assumption of a context-insensitive first-pass analysis. Nevertheless, the possibility that the perception of assimilated utterances is context-sensitive at auditory or phonetic stages of processing needs further consideration.

The assumption that phonological assimilations are dealt with after the auditory input has been transformed into a string of phonemes has also been criticized on other grounds. First of all, this assumption builds on a classical view of how phonological assimilations arise (Chomsky & Halle, 1968). According to this view, the outcome of place assimilation in the case of [tæynbaŋk] is [tæymbaŋk] (Engl. garden bench), with a full-blown /m/, that is indistinguishable from an intended /m/, as in [armstu:l] (Engl. arm chair). However, detailed measurements revealed subtle articulatory (Nolan, 1992) and acoustic (Gow & Hussami, 1999) differences between intended and assimilated segments from the same phonological category. These results are not compatible with the classical view that assimilation is a rule-like process at an abstract phonological processing stage. Instead, these results support the articulatory-phonology framework (Browman & Goldstein, 1992), which describes place assimilation as the result of a variable degree of gestural overlap between the assimilated and assimilating segment.

More importantly for current purposes, such subtle differences between intended and assimilated segments may be picked up by the perceptual system. Gow (2002a) recorded phrases like [raip beriz] arising from speakers intending to say ‘ripe berries’ or intending to say ‘right

berries.' (The utterance 'right berries' may be pronounced [raip beriz] if the optional rule of English coronal place assimilation is applied) Fisi, Gow showed that these tokens differ acoustically. That is, the formant endpoints in [raip] arising from 'ripe berries' differ from the formant endpoints in [raip] arising from 'right berries,' although both items are phonologically identical. Then, Gow used these utterances in a cross-modal identity priming task. Although the two types of utterances were phonologically identical, different patterns of priming were observed depending on the *intended* meaning and thus phonetic detail of the utterance. If the phrase [raip beriz] arose from an intended 'right berries' plus assimilation, it primed the visual target 'right' but not 'ripe' in a lexical-decision task. A phonologically identical phrase [raip beriz] resulting from an intended 'ripe berries,' however, primed the visual target 'ripe' and inhibited 'right.' That is, the auditory stimuli only primed the words the speaker intended to say. Thus, Gow showed that phrases that become identical after phonological categorization nevertheless produce different patterns of priming. Hence, the small acoustic differences between assimilated and intended segments arising in production seem to be functional in perceptual compensation for phonological assimilation.

The finding that phonetic detail matters in perceptual compensation for phonological assimilation is difficult to reconcile with the assumption that compensation for assimilation is only initiated after a first-pass phonological analysis of the input. Therefore, we pursue the issue of phonetic detail for the case of nasal place assimilation (NPA) in Dutch. According to the estimation of van Heuven and van Berg (1982), NPA is applied in 60% of the cases in which an alveolar nasal and an obstruent co-occur. Whether or not NPA applies is, however, most likely not a random process. Based on the model of articulatory phonology, the key variable determining whether assimilation will occur or not, is the gestural overlap between the movement of the tongue tip to an alveolar place of articulation for the word-final nasal and the labial closure gesture for the following labial obstruent. It follows that, the longer the duration of the nasal segment, the smaller the degree of gestural overlap, which, in turn, makes place assimilation less likely. According to the model of phonological inference (Gaskell et al., 1995; Gaskell & Marslen-Wilson, 1998), the length of the nasal should not influence the perception of possibly assimilated utterances, because length is not represented in categorized phonological input. However, if the disambiguation of assimilated utterances occurs at an auditory or phonetic level, length may naturally influence perception. It is unlikely that long and hyper-articulated segments arise as a consequence of assimilation. Therefore, one may conjecture that only short segments are subject to assimilation, and therefore, the perceptual system may not apply

compensation for assimilation to long segments. This hypothesis was tested in the first experiment.

However, even if perceptual compensation for NPA is sensitive to length variation, the case for a role of phonetic detail may be strengthened if it is possible to show that the same phonetic detail, which causes differences in the perception of possibly assimilated utterances, also differentiates intended alveolar segments from intended labial as well as assimilated labial segments in *production*. Therefore, we also investigated the role of phonetic detail in speech production in, both, free conversation and in an experimental task. This was done by, first, using samples of casual, free speech from the Dutch corpus collected by Ernestus (2000). Data from this corpus reflect the kind of speech that listeners have to deal with in everyday situations. A corpus study is limited to naturally occurring observations in which many variables (e.g., word frequency) cannot be controlled for. Therefore, we also conducted a production experiment, in which words were selected based on frequency. If both production studies show that phonetic detail, operationalized as the length of the segments, differentiates intended from assimilated nasals, while a perception experiment shows that the same phonetic detail influences perception, a strong case can be made that phonetic detail is functional in perceptual compensation for NPA.

Experiment 1

In this experiment, we tested whether the length of a word-final nasal influences its perception in an utterance in which NPA is allowed. The length of a nasal is not represented after phonological categorization. Therefore, the duration of a segment should not influence a mechanism for compensation for NPA that operates with categorized phonological features as input. In contrast, if it can be shown that compensation for NPA takes into account phonetic detail, it indicates that compensation may occur at auditory or phonetic levels of processing.

In order to probe the perception of possibly assimilated utterances, we used a two-alternative forced-choice task (2AFC), in which listeners have to decide whether the word-final nasal of a given word (i.e., the Dutch word 'tuin') has been produced canonically with an alveolar place of articulation or with a labial nasal. In this experiment, we used similar materials as in chapter three. However, the present experiment differs from our earlier experiment (Chapter three, Exp. 1) in two aspects: the length of the final nasal was manipulated and no feedback was supplied.

Method

Participants

9 (7 female/2 male) undergraduate students from Maastricht University participated in the study. They were paid for participation. All participants indicated to have no history of hearing problems.

Stimulus material and apparatus

A male native speaker of Dutch was selected for stimulus generation, because of his low f_0 (75–80 Hz), which did not vary greatly between utterances. The sonority of the speaker's voice contributed to the naturalness of the cross-spliced stimuli. Multiple pronunciations of 'tuinbank' [tœynbɑŋk], 'tuimbank' [tœymbɑŋk], 'tuinstoel' [tœynstu:l], and 'tuimstoel', [tœymstu:l] were digitally recorded. All sound editing and manipulation was done using PRAAT 3.9 (Boersma & Weenink, 1999). Experimental stimuli were created by cross-splicing and then lengthening the stimuli with the Pitch-Synchronous Overlap and Add (PSOLA) method as provided by PRAAT 3.9. First, [tœy] was spliced from an utterance of [tœynstu:l]. An alveolar and a labial nasal murmur were spliced off from two other utterances of [tœynstu:l] and [tœymbɑŋk]. The nasal murmurs were equated in length and energy (using a root-mean square measure) using the 62 ms /n/ from the [tœynstu:l] utterance as template. Due to the constraint of splicing at zero-crossings, the nasal murmurs differed slightly in length. However, this was less than 1 ms. The edited nasal murmurs were concatenated with [tœy], resulting in the stimuli [tœyn] and [tœym]. The context word 'bank' [bɑŋk] was spliced from another utterance. The [bɑŋk] utterance was produced without pre-voicing. In order to sound natural, a 25 ms period of silence was used to emulate a closure. The two resulting stimuli were then manipulated using PSOLA. The length tier was manipulated from 10 ms before onset of the nasal murmur to 10 ms after the offset of the nasal murmur. Within the first 10 ms, the length tier increased linearly from a factor 1 to a factor of 1.5 or 2, remained stable up to the end of the nasal murmur and then returned to a factor 1 during the 10 ms after nasal murmur offset. This procedure created six experimental stimuli: the three versions of the utterances 'tuinbank' [tœynbɑŋk], 'tuimbank' [tœymbɑŋk] with nasal murmurs with a length of respectively 62, 93, and 124 ms.

The stimuli were played to the participants by a DOS-operated computer using the software package ERTS (Behringer, 1996). The output of the SoundBlaster soundcard was redirected to 4 loudspeakers (JBL Control 25) within a sound-attenuated booth with a Spirit Folio mixer and amplified by a Yamaha P4050 power amplifier. The maximal sound-pressure level for all stimuli was 70 dB(A).

Procedure

Participants sat in a sound-attenuated booth facing a monitor and an array of buttons. Two of the buttons were labeled 'tuin' and 'tuim.' Participants were instructed to press one of the buttons upon hearing a target. Instructions stressed accuracy without special emphasis on speed of response. First, there was a short practice period in which participants judged the stimulus words in isolation (i.e., [tœyn] and [tœym]). During this training, participants received visual feedback in the form of the words "right" and "wrong" on the monitor indicating whether their identification had been correct. After practice, the experiment started. Stimuli were now presented with the context word [bɑŋk]. In these trials, no feedback was given. The three different length conditions were blocked and three different orders were created on the basis of a Latin square. Each of these orders was used for three participants. Within each block the order of presentation was randomized.

Design

The experiment entailed two factors. The first factor, Nasal Murmur, refers to the nasal murmur of the target word which was either alveolar or labial. These nasal murmurs were presented in three lengths, the original length, or lengthened by the factor 1.5 or 2. This Length factor constitutes the second factor of the design. The dependent variable is the percentage of correct responses.

Results

The results were transformed into percentages correct responses per condition (see Table 5.1). A repeated measure ANOVA revealed that the factor Nasal Murmur was significant ($F(2, 16) = 14.48, p < 0.005$), as was the factor Length ($F(1, 8) = 14.73, p < 0.01$) and the interaction ($F(2, 16) = 8.50, p < 0.005$).

Table 5.1 Percentages of correct identifications in Experiment 1

	Length		
	62 ms	93 ms	124 ms
Nasal	% correct	% correct	% correct
'tuinbank'	95.0	97.8	99.6
'tuimbank'	66.2	92.0	98.7
Difference	28.2	5.8	0.9

In order to investigate this interaction, we calculated the effect of Nasal Murmur (percentage correct 'tuinbank' minus percentage correct 'tuimbank') for each participant and each length condition. This measure shows to what extent a participant tended to perceive the possible alternative pronunciation [tæymbaŋk] as an instance of the canonical pronunciation [tæynbaŋk]. As such, this effect-size measure reflects the degree of compensation for NPA. These effect-size measures were subjected to a one factor repeated-measure ANOVA (identical to the test for the interaction in the previous ANOVA). Post-hoc Newman Keuls comparison were used to evaluate the differences in the mean effect sizes (%correct [tæynbaŋk] minus %correct [tæymbaŋk], see Table 5.1). It turned out that the degree of compensation for NPA for the stimuli in the original length condition (28.2%) was significantly different from the two other conditions (medium-length condition, 5.2%, $p < 0.05$; maximum-length condition, 0.9%, $p < 0.01$), while the two conditions with longer nasal murmurs did not differ significantly.

Discussion

The results showed that the length of a nasal murmur did influence the perception of the word-final nasal. A compensatory response was more likely for short nasals than for longer nasals. This indicates that participants took phonetic detail into account in perceptual compensation for phonological assimilation. Listeners only tended to interpret short labial nasal murmurs as potentially coronal. Longer nasal murmurs were immune to any influence exerted by the concatenated context. Such segments were not subjected to compensation for assimilation. Hence, long nasal murmurs were perceived as intended long nasal murmurs. In agreement with Gow (2002a), this shows that phonetic detail is critical in compensation for phonological assimilation.

However, it might be argued that the results of the current experiment show a perceptual bias that may be due to the artificial lengthening of the nasal murmurs. As more information is

available, because the nasal murmurs are longer, participants show better discrimination of the two potentially ambiguous word forms. According to such an argument, the present results only show that the performance in a 2AFC task is influenced by the length of a nasal murmur; and this is due to specific task affordances. There is no obvious way to control for this alternative explanation within the context of the experimental design. Therefore, we followed a different route and tried to show the 'ecological validity' of the results obtained in this perceptual experiment. We examined whether the effect of length on perception reflects existing phonetic details in the actual productions of assimilated utterances and intended labial nasals. Concretely: do the results of Experiment 1 reflect an ecological difference in the intrinsic length of labial and alveolar nasal murmurs in samples of naturally produced speech?

A corpus study

In order to measure the distribution of length of nasals, we analyzed a corpus of casual spoken Dutch collected by Ernestus (2000). She collected and transcribed about 30 hours of conversations in standard, but casual Dutch. In order to gather recordings representative of standard Dutch, participants were selected from the Western regions of the Netherlands. In order to record samples of casual speech, two well acquainted speakers were recorded during free conversation and during a role play, in which speakers were asked to imitate a negotiation between supplier and retailer.

Data collection

Speech samples were selected by means of an automatic text search in the orthographic transcription of the corpus. In order to make it possible to link the data from this corpus study to the perception experiment, only nasal + /b/ combinations, such as instances of 'n#b,' 'm#b,' 'nb,' and 'mb,' were searched for. The selection of samples was further constrained by the following restrictions: Items ending in [ən]—mostly plural and infinitive suffixes—were disregarded, because, in Standard Dutch, this ending is mostly realized as a single schwa. In addition, word-internal nasal-plus-/b/ combinations were only used when the two segments were separated by a morpheme boundary, and the morphologically complex word was transparent. That is, words such as 'ambt' [office] were not used because the /mb/ combination is within the morpheme. However, words such as 'stemband' [vocal chord] were used, given that the two syllables correspond to the words 'stem' [voice] and 'band' [chord], each individually contributing to the meaning of the compound. Tokens were rejected when contaminated with background noise or a simultaneous utterance of another speaker or when the speaker made a pause between the two words.

The onset and offset of the nasal murmurs were determined on the basis of the simultaneous inspection of oscillogram, spectrogram, and formant traces calculated by PRAAT 4.0 (Boersma & Weenink, 2001). Cues to nasal onset were the local minima in the oscillogram, the onset of the nasal zero in the spectrogram, and a sudden change of the formant structure. Offset was determined by either an onset of pre-voicing or a sudden increase in amplitude co-occurring with the offset of the nasal zero.

Results

Searching the corpus yielded 274 utterances that were free of contaminating noise. There were 243 of these with an /n(#)b/ sequence and only 31 with a /m(#)b/ sequence. This shows that, in colloquial speech, word-final alveolar nasals are much more common than labial nasals.

The minimum length was 25 ms for an /n/ and 47 ms for an /m/. Maximal length was 150 ms for /n/ and 145 for /m/. Medians were 64 ms for /n/ and 80 ms for /m/. Length data were analyzed with an ANOVA using Speaker as a random factor and Target (/n/ or /m/) as a fixed factor. Word frequency as extracted from the lexical database CELEX (Baayen, Piepenbrock, & Gulikers 1995) was used as a co-variate. This analysis revealed that the factors Speaker and Target influenced length significantly ($F(12, 12) = 5.73, p < 0.005$ and $F(1, 32) = 21.94, p < 0.001$, respectively). Neither the co-variate nor the interaction between Target and Speaker attained significance ($F(1, 247) = 1.28, p > 0.25$ and $F < 1$, respectively). Figure 5.1 shows a boxplot of the length data for /n/ and /m/ targets. As is evident in this visualization, the overall difference in length between alveolar and labial targets does not arise as a consequence of a mean difference between similarly shaped distributions. Instead, it is evident that alveolar nasals and labial nasals mainly differ in the proportion of very short instances. For both kinds of nasals quite long instances (around 120 ms) are often found. However, only alveolar nasals are commonly found with a length of 60 ms or even less.

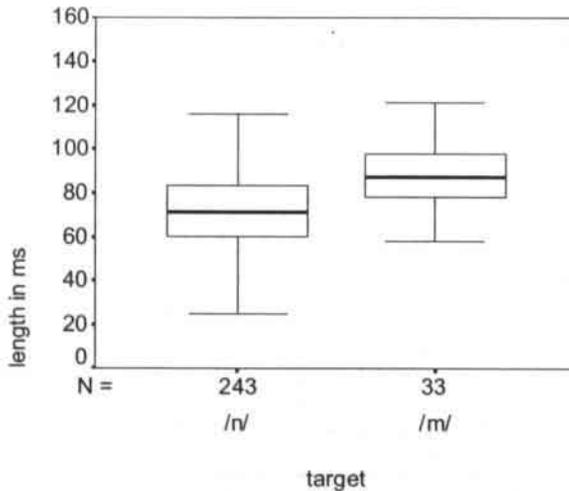


Figure 5.1 Length of labial and alveolar nasal murmurs after accounting for speaker differences.

Discussion

The corpus study shows that sub-categorical length differences exist between labial and alveolar nasals in casual conversational speech. Intended labial nasals seem to have a longer 'minimal length' than alveolar nasals. In some respect, this ecological fact mirrors the results of Experiment 1. In this experiment, long labial nasals were not subjected to compensation for assimilation. However, short nasals were perceived as more ambiguous, but only in context in which nasal place assimilation may apply.

The present study shows that labial nasals and alveolar nasals differ in their length distribution in contexts in which NPA might apply. This leads to the question whether this is restricted to these cases or whether labial nasals are generally longer than alveolar nasals, independent of phonological context. Based on articulatory evidence, it seems that closure gestures for labial and tongue-tip closures do not intrinsically differ in the amount of time needed to achieve closure (Adams, Weismer, & Kent, 1993). Nevertheless, it seems possible that labial nasals may generally be longer. For words it has been shown that more frequent words are shorter (Greenberg, Harvey, Hitchcock, & Chang, 2002). If a similar inverse relationship exists for phoneme frequency and phoneme length, word-final /m/s may generally be longer than word-final /n/s. This possibility was evaluated in Experiment 2, in order to achieve a more direct comparison between perception and production data.

Experiment 2

In this experiment, we pursued the question whether the length differences between alveolar and labial nasals are restricted to contexts in which NPA might apply. Alternatively, it may be that word-final alveolar nasals are generally shorter than word-final labial nasals. To this end, we asked participants to produce utterances with word-final nasals by means of a picture naming task. Word-final nasals were to be produced as either phrase-final segments in single-word utterances or phrase-medial segments in compound nouns. In the compounds, the second noun always started with a /b/. In this context, NPA might apply. If length differences between labial and alveolar nasals are restricted to contexts in which NPA might apply, we expect a length difference only in the compound nouns.

For the picture naming task, twenty compound nouns were selected from the CELEX database (Baayen, Piepenbrock, & Gulikers 1995). All compounds consisted of two nouns, of which the first either ended in /n/ or /m/, and the second noun started with a /b/. Participants produced the first noun either in isolation or in the compound noun. Single and compound

nouns were matched for lexeme frequency for words containing word-final labial and alveolar nasals. The dependent variable was the length of the nasal murmur.

Method

Participants

Six undergraduate students (four female/two male) of the Universiteit Maastricht participated in the study for a course credit. All participants were native speakers of Standard Dutch. Four subjects had a southern Dutch accent, two had a Flemish accent. None of them had any phonetic training and all were unaware of the existence of assimilation rules as verified at the end of the experiment.

Material

Twenty compound nouns were selected from the CELEX data base with a medium to low frequency (see Appendix). All compounds were completely transparent, that is, the meaning of the compound derives from the meaning of the constituent nouns. The second noun of the compound always started with /b/. Half of the compounds had a first noun ending in an /n/, the other half ended in /m/. For each word, a picture was selected from the ArtToday Picture database.

In order to conceal the purpose of the experiment, two lists were then prepared and a given participant named only the words on one list. On each list, half of the words were to be named as compound nouns and half of the words had to be named as single nouns. If a word was to be named as a single noun on the first list, it was to be named as a compound noun on the second list. Thus, each participant named a word as either a compound noun or a single noun.

Procedure

Given the constraints on the pool of possible words, it proved difficult to find pictures that were named as intended spontaneously. (It is, for instance, difficult to find a picture that is spontaneously named as 'vineyard.') Therefore, the picture-naming task was concealed as a 'picture-word association task.' Prior to picture naming, participants were handed a list with pictures and associated words. Participants had five minutes to memorize the lists.

After memorization, participants were seated in a sound-shielded booth facing a computer monitor. They were instructed that the pictures they had just learned to associate with the target words would now appear on the screen and that their task was to react as quickly as possible with the associated word. The stimuli were presented to the participants by a DOS-operated computer using the software package ERTS (Behringer, 1996). Each participant named each of the twenty pictures on his or her list twice.

Productions were recorded using a Sennheiser K6 Table Microphone and stored on a DAT-tape using a sampling frequency of 44.1 kHz. After recording, stimuli were rerouted to a PC, band-pass filtered (50-10400 Hz) and re-sampled to 22050 Hz. The length of the nasal segments was estimated as in the previous corpus study.

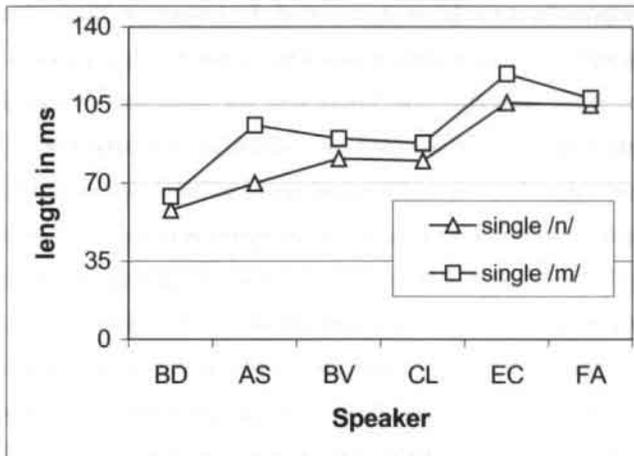
Design

The design entailed two factors with two levels each. The first factor is Target, referring to the place of articulation of the word-final nasal with 'alveolar' and 'labial' as levels. The second factor is the Phrase position factor with the levels 'final' and 'medial.' The dependent variable is the length of the nasals.

Results

The picture-word association did not pose a problem for the participants. Only 9 naming errors occurred over the 240 trials. The length of the nasal murmur in the 231 correct utterances were analyzed using an ANOVA with Target (/n/ or /m/) and Phrase position (medial vs. final) as fixed factors and Participant as random factor. This analysis revealed a significant interaction between Phrase position and Participant ($F(5,5) = 13.1, p < 0.01$) and a marginal interaction between Phrase position and Target ($F(1,5) = 4.6, p < 0.1$). No other effects approached significance.

A



B

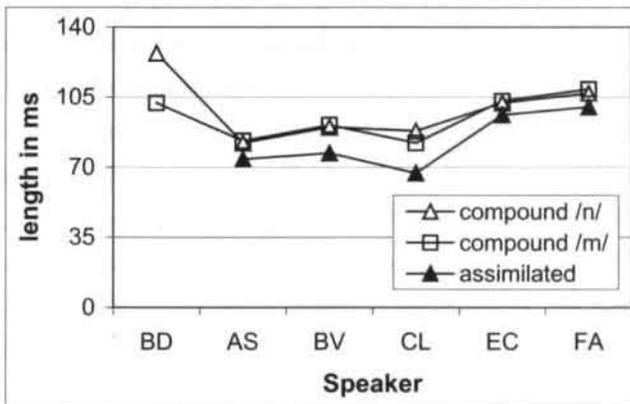


Figure 5.2 Mean length of intended labial, assimilated labial, and alveolar nasal murmurs for different speaker in a) single nouns and b) compound nouns in which assimilation of the word-final nasal of the first noun might occur.

Given the significant interaction, we ran separate analyses for single and compound nouns. The ANOVA for single nouns (see Figure 5.2) with Target (/n/ vs. /m/) and Participant as predictors revealed a significant effect of Target (/n/ vs. /m/, $F(1,5) = 11.6, p < 0.025$) and a significant effect of Participant $F(5,5) = 23.2, p < 0.005$). The interaction was not significant ($F < 1$). Overall, the labial nasals (mean: 94 ms) tended to be slightly longer than the alveolar nasals (mean: 85 ms, see Figure 5.2a).

For the compounds, a similar analysis tested the impact of Target (/n/ vs. /m/) and Participant on the length of the word final nasals. This revealed a significant effect of Participant ($F(5,5) = 6.0, p < 0.05$). The effect of Target was not significant in this analysis ($F < 1$). However, these simple effects were qualified by a significant interaction ($F(5,104) = 2.6, p < 0.05$). Figure 5.2 shows that the source of this interaction is mainly speaker BD, who, in contrast to the other participants, pronounced alveolar targets significantly longer than labial targets (127 ms vs. 102 ms, $t(17) = 3.6, p < 0.005$). To shed further light on the interaction between Participant and Target, we analyzed how often the different speakers produced an intended alveolar nasal as an assimilated labial. Participants varied greatly in the amount of assimilation they produced. Only participant (BD) produced all forms ‘canonically.’ The two participants FA and BV produced one assimilation (10%). Two other participants (AS and EL) produced a medium amount of assimilated utterances (30% and 40%, respectively), while one participant (CL) produced mostly (70%) assimilated utterances.

For the five speakers who did produce assimilated forms, the length-of-nasal data were then re-analyzed taking into account whether an alveolar nasal was produced canonically or in an assimilated form. That is, an ANOVA was performed using Participant and Type of Utterance as random factors (see Figure 5.2b). Type of Utterance was either ‘labial,’ ‘alveolar pronounced canonically,’ or ‘alveolar pronounced assimilated.’ The analysis revealed a significant effect of Type of Utterance ($F(2,48) = 6.4, p < 0.005$) and a significant effect of Participant ($F(8,74) = 16.2, p < 0.001$). The interaction was not significant ($F < 1$). A post-hoc Newman-Keuls procedure revealed that the assimilated nasals (mean: 78 ms) were significantly ($p < 0.05$) shorter than canonically pronounced alveolars (mean: 94 ms) and intended labial nasals (mean: 95ms), while these last two did not differ significantly.

Discussion

In this production study, we mainly asked whether length differences between alveolar and labial nasals are restricted to contexts in which NPA might apply or whether length differences generalize to utterance-final nasals. This last option turned out to be true. That is, the length differences between labial and alveolar nasals as found in the corpus study extend to single-word utterances and are not restricted to segments in which NPA might apply. How can this result be interpreted? A straightforward interpretation of this difference is to attribute the length differences to physical constraints. The tongue tip closure of alveolar segments might be easier to produce than the labial closure. However, both closures eventually involve jaw movement, and Stevens (1998) reports that tongue-tip and lip closure gestures do not strongly differ in movement velocity.

Given that articulatory constraints do not seem to be involved, another likely explanation is that the differences in length between alveolar and labial nasals reflect the predictability of the place of articulation of a final nasal. As the corpus study indicated, word-final labial nasals are less frequent in spoken Dutch than alveolar nasals. It might be the case that the highly likely word-final /n/ is produced with less articulatory effort in speech production than the less frequent /m/. This assumption is supported by a result obtained by Ernestus (2000) in her overview of phoneme realizations in casual Dutch. Ernestus reported that /n/ may be deleted in word-final position. In contrast, she did not report any /m/-deletions, although high frequency words with word-final /m/ exist (e.g. 'hem,' Engl. him).

It seems that, on the one hand, word-final alveolar nasals are prone to assimilation or deletion. On the other hand, alveolar nasals are quite predictable as the most common word-final nasal. Therefore, it is conceivable that a nasalization of the preceding vowel may be sufficient to signal the presence of a following nasal, because vowel nasalization is not phonological in Dutch. The listeners can therefore infer the presence of a nasal from vowel nasalization. If no clear place information is available, the listener may use the a-priori likelihoods in order to predict the place of articulation. Given this listener strategy, Dutch speakers need not put much effort in the production of an alveolar nasal murmur, which in turn, may lead to NPA and word-final deletions. However, the less predictable labial nasal murmur has to be produced with more articulatory effort in order to communicate the 'non-default' place of articulation of this word-final nasal. This 'division of labor' between speaker and listener may explain why length differences were observed between labial and alveolar nasals.

Another result obtained in the present study was that speakers differed in the frequency with which NPA was applied and these differences were related to length differences in the production of alveolar and labial nasals. Therefore, we categorized these productions in canonically produced and assimilated versions. This revealed length differences between assimilated and canonically produced word-final /n/s, thus complementing the findings of Experiment 1. In this perception experiment, long labial nasals were not subjected to compensation for NPA. The production data show that long labial nasals were unlikely to arise from NPA of intended alveolar nasals. They are more likely to result from intended labial nasals. It seems that the perceptual systems takes this into account.

An additional unforeseen result is related to the data of one speaker. Participant BD produced—opposite to the general trend—longer alveolar nasal murmurs than labial nasal murmurs. However, all of her nasal segments in contexts in which NPA might apply were produced canonically, that is without application of NPA. Although the participant was neither

aware of the structure of the stimuli nor the existence of the NPA rule, it seems that she was able to adjust her speaking strategy in order to prevent assimilation and thus spoke 'correctly.'

General Discussion

In the present studies we examined the role of phonetic detail in the perception and the production of NPA. Experiment 1 investigated whether listeners take into account the length of a nasal segment when identifying word forms in which NPA might apply. This was indeed the case. The longer a segment, the less likely it was interpreted as a possible result of NPA. Compensatory responses were only observed for nasals with a length of 62 ms and to a lesser degree for nasals with a length of 93 ms. A corpus study showed that a duration of 62 ms is within the lower tail of the duration distribution for alveolar nasals, but is an extreme value for the duration distribution of labial nasal murmurs. In addition, a picture-naming study showed that this length difference is not restricted to contexts in which NPA might apply, but generalizes to word-final nasals in phrase-final position. The picture-naming study also revealed that the shorter the alveolar nasal, the more likely it is that the nasal is subjected to NPA.

Together, these data show that phonetic detail bears useful information in order to differentiate between intended labial nasals, assimilated labial nasals, and canonically pronounced alveolar nasals; and the perceptual system makes use of this phonetic detail. If perceptual compensation for NPA was based on categorized phonological input, helpful phonetic-detail information would not be available.

Such effects of phonetic detail are difficult to capture in the phonological-inference model proposed by Gaskell and Marslen-Wilson (1996, 1998). In their pioneering work on perceptual compensation for phonological assimilation, they assumed that phonetic detail is not important. Very recently, however, Gaskell (in press) elaborated his model of phonological inference in order to accommodate anticipatory effects in the perception of phonological assimilations. Gow (2001) had shown that assimilated segments are not only subject to regressive context effects, but also allow the listener to anticipate the next segment that actually causes the assimilation. The modified model of phonological inference now incorporates phonetic details by allowing graded input. That is, a word-final nasal at the input level may be 40% alveolar and 60% labial. Given such graded input, the modified model is able to correctly predict anticipatory effects. This is due to the fact that ambiguous segments arise mostly due to assimilation. Therefore, the model learns that ambiguous segments with cues for both an alveolar and a labial place of articulation are, first, probably assimilated segments with an underlying alveolar place of

articulation, and, second, that the cues for a labial place of articulation are probably a consequence of a following labial segment.

In its current form, the model still predicts that fully labial segments followed by another labial are subject to phonological inference triggered by the context. The data of Experiment 1 showed that this is not always the case. Hyper-articulated labial nasals are not prone to regressive context effects. Therefore, further refinements of the model still seem necessary. In principle, however, this seems possible within the framework of phonological inference. Consider what happens if articulatory effort is added as an input dimension. Our results show that labial nasals with strong articulatory effort always correspond to underlying labial nasals in perception. As the amount of under-articulation increases, the likelihood of an assimilated labial nasal also increases. As such, the learning problem presents itself as linearly separable in a vector space with the dimensions place of articulation and articulatory effort. Hence, the current data could in principle be accounted for by a co-variant learning mechanism as assumed for models of phonological inference (Gaskell, *in press*; Gaskell, Hare & Marslen-Wilson, 1995).

This shows that a learning mechanism as proposed by Gaskell et al. (1995) is in principle able to account for the results as reported here. Independent from the question whether compensation for assimilation is learned or not, the question arises whether compensation for assimilation is aptly described by a model working on phonological features. In this chapter, we investigated the role of length in the perception and production of NPA. Results showed that length distinguishes assimilated and intended labial nasals in production, and the perception experiment (Experiment 1) shows that the perceptual system indeed uses this information. This may be accounted for in a phonological model by allowing length to influence the amount of activation of phonological features. However, length is not the only variable that differentiates assimilated from intended segments. Gow (2002a) has shown subtle differences in the formant structure of intended and assimilated labial stops, with similar repercussions in perception as has been shown for length in the present experiments. Again, a phonological model could accommodate this finding by allowing graded activation of phonological features. However, although a phonological model may in principle incorporate the roles of length and formant structure separately, it does so by the same means: the amount of activation of phonological features. Hence, formant structure and length are indistinguishable at a phonological level. If, in contrast, compensation arises at auditory or phonetic levels of processing, variables such as length and formant structure can influence the compensation-for-assimilation process, as independent acoustic variables. In the phonological model, they can only influence the compensation-for-assimilation process by means of their projections on the activation of (the

same) phonological features. In contrast, an acoustic/phonetic mechanism for compensation for assimilation allows different sub-symbolic variables to interact and, therefore, has more degrees of freedom to perceptually compensate for phonological assimilation. Given that different types of assimilation lead to quite different acoustic manifestations of assimilation, acoustic/phonetic mechanisms as proposed here may be more apt to compensate for assimilation than a mechanism that relies on graded phonological-feature input.

Appendix

Table A5.1 Words and compounds used in Experiment 2

Dutch word	Engl. translation	single noun log fr	compound noun log fr
azijn#bacterie	vinegar bacteria	2.59	.30
bazuin#blazer	trombone player	1.58	.00
beton#blok	concrete block	2.61	1.00
banaan#boom	banana tree	2.41	.30
bouillon#blokje	beef cube	2.56	.85
graan#bouw	grain growing	2.41	.30
neon#buis	neon tube	1.46	1.61
salon#boot	saloon boat	2.97	.78
tuin#bank	garden bench	3.70	1.00
wijn#berg	vineyard	3.99	1.90
arm#band	bracelet	3.90	2.36
bliksem#bezoek	lightning visit	2.73	.70
boterham#beleg	sandwich filling	2.89	.30
dam#bord	checkerboard	2.35	1.11
drum band	marching band	1.41	.95
maximum#bedrag	maximum amount	2.57	.48
natriumbicarbonat	sodabarbonate	1.75	1.18
rem#blok	brake block	2.49	.48
schuim#bad	bubble bath	2.62	.95
totem#beeld	totem statue	1.48	.30

Perceptual Integration Drives the Recognition of Assimilated Word Forms¹

We investigated how spoken words are recognized when they have been altered by phonological assimilation. Previous research has shown that there is a process of perceptual compensation for phonological assimilations. Three recently formulated proposals regarding the mechanisms for compensation for assimilation make different predictions with regard to the level at which compensation is supposed to occur as well as regarding the role of specific language experience. In the present study, Hungarian words and nonwords, in which a viable and an unviable liquid assimilation was applied, were presented to Hungarian and Dutch listeners in an identification task and a discrimination task. Results indicate that viably changed forms are difficult to distinguish from canonical forms independent of experience with the assimilation rule applied in the utterances. This reveals that auditory processing contributes to the perceptual compensation for assimilation, while language experience has a role to play when identification is required.

¹ This is an adapted version of Mitterer, H., Csépe, V., & Blomert, L. (submitted).

The invariance problem in speech perception presents itself to the observer as a ‘variance problem.’ Different utterances of the same word vary dramatically depending on the speaker, the emphasis in the sentence, and the immediate phonological context. The immediate phonological context of a word can in some circumstances lead to phonological assimilation, that is, word-final or word-initial segments may be modified by onsets or codas of the surrounding words. Assimilation is a rather strong form of context-dependent variation that, superficially, neutralizes a phonemic contrast. This raises the question of how words can still be recognized in their assimilated form.

Studies concerned with the recognition of assimilated word forms mostly investigated cases of place assimilation in Germanic languages as English (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2001, 2002a, 2003), Dutch (see chapter three), and German (Coenen, Zwitserlood, & Bölte, 2001). This type of assimilation typically applies in *C#C* sequences in which the word-final consonant takes over the place of articulation of the second, word-initial consonant (e.g., ‘lean bacon’ → ‘leam bacon’). Studies concerned with the recognition of assimilated word forms (e.g. ‘leam’) showed that the variance due to assimilatory processes is compensated for in a context-sensitive way. A changed form (‘leam’) is accepted as an instance of the intended word (‘lean’), only if the context allows the change to occur as the result of a phonological assimilation (... ‘bacon’). Assimilated forms are perceived as a violation, if the context does not allow assimilation (... ‘salami’): Gaskell and Marslen-Wilson (1996, see also Gow, 2002a) showed that assimilated forms such as ‘leam’ prime a lexical decision to a visual target, that is, the written word denoting the original word (i.e., lean), only if the phonological context allows the assimilation, but not if the context does not allow the assimilation. Similarly, they (Gaskell & Marslen-Wilson, 1998) showed in a phoneme-monitoring task that the /m/ in ‘leam’ is more likely to be—falsely—recognized as an /n/ if the phonological context allows an assimilation. Using event-related potentials, we showed in chapter three that compensation is a fast and automatic process. In a passive oddball paradigm, participants were exposed to an oddball series while watching a silent movie. The oddball series consisted of a Dutch word pronounced canonically ([tœyn...], Engl. ‘garden’) as standard and deviants, which were either viable alternative pronunciations ([tœymbank], Engl. ‘garden bench’) or unviable alternatives (*[tœymstu:l], Engl. ‘garden bench’). A deviation-elicited MisMatch-Negativity component could only be observed in case of the unviable alternative. This shows that pre-attentive perceptual mechanisms make a difference between a canonical pronunciation and an unviable alternative, whereas viable alternatives are ‘perceived’ as not significantly different from the canonical pronunciations.

Two different accounts for perceptual compensation for phonological assimilation have been proposed until now. First, Gaskell and Marslen-Wilson (1998) put forward a model of phonological inference. According to this view, the compensation is located at a phonological processing level. First, the input is categorized phonologically in a context-insensitive way. That is, an assimilated utterance such as 'leam bacon' is first categorized as having a labial nasal. Only when the following segment is phonologically categorized as a labial plosive, a process of regressive phonological inference compensates for the alveolar-to-labial assimilation of the previous segment. This regressive-inference mechanism is also assumed to be influenced by the lexicon in a top-down manner. That is, the regressive-inference is stronger if it changes a nonword into a word (as is the case in 'leam' → 'lean').

In addition, Gaskell and Marslen-Wilson (1996, 1998) assume that phonological inference is based on implicit knowledge of the assimilation rules of the listener's native language. Being exposed to the rule of coronal place assimilation, an English listener learns that a labial nasal followed by a labial obstruent may actually correspond to an intended alveolar nasal that has been assimilated by the labial obstruent (see Gaskell, Hare, & Marslen-Wilson, 1995).

A second account of compensation for phonological assimilation has been proposed by Gow (2001, 2002a, 2003). According to Gow's feature-parsing account, the assimilated segment, for instance the /m/ in 'leam bacon,' should be considered as bearing information for both a labial and the original alveolar place of articulation. That is, assimilated segments differ from intended non-alveolar segments (as in 'arm chair') and are hybrids in terms of place of articulation. This assumption has been corroborated by acoustic measurements of assimilated segments (Gow & Hussami, 1999; Nolan, 1992) and is now widely accepted (see Gaskell, 2002). Thus, the /m/ in 'leam bacon' actually carries feature cues for the original alveolar and the assimilated labial place of articulation. Compensation for articulation is then achieved by parsing the cues for the labial place of articulation from the assimilated segment (the /m/ in 'leam'), and ascribing this information to the assimilating segment (the /b/ in 'bacon'). After this feature-parsing, the assimilated segment only carries cues for the intended alveolar place of articulation. Similar to the phonological-inference account (Gaskell & Marslen-Wilson, 1996, 1998), the feature-parsing account assumes that compensation for assimilation is based on a representation of phonological features. In contrast to the phonological-inference account, however, the features are not assumed to be already organized in discrete segments, so that the segment $n+1$ can regressively influence segment n . Instead, features are activated in a continuous fashion over time, leading to partially overlapping feature activations. Feature parsing disentangles the overlapping features in order to associate them with segmental representations. Hence, both the

phonological-inference account and the feature-parsing account assume that compensation for assimilation occurs at a phonological processing level. The difference between the two accounts is that segmentation of the continuous input precedes compensation for assimilation in the framework of phonological inference, while segmentation follows compensation for assimilation in the framework of feature parsing.

Most importantly, Gow (2003) assumes that feature parsing is a general process in speech perception that leads to a de-convolution of overlapping features. Feature parsing is supposed to build on general grouping processes as described by, for instance, Bregman (1990). Therefore, compensation for assimilation does not depend on experience with the assimilated rule in question, but on Gestalt principles working at a phonological-feature level. A lexical influence on compensation for assimilation does not follow directly from this point of view. Nevertheless, it is possible to account for a lexical influence by appealing to the notion that higher-level schemata are allowed to influence grouping in the framework of Bregman. Lexical items as such may be viewed as such higher schemata that influence the grouping process.

These abovementioned accounts assume that compensation for assimilation occurs at a level at which the acoustic input has already been transformed into continuous or discrete phonological features. However, it is conceivable that compensation for assimilation arises at earlier auditory processing levels. As Gow (2003) also stated, the present evidence with regard to compensation for phonological assimilation does not rule out such a possibility. In this chapter, we investigate the possibility of such an account. Our proposal is based on the framework developed by Kingston and Macmillan (Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickey, & Bartels, 1999). In this framework it is assumed that speech perception involves at least two processing stages. First, the acoustic input is converted into a multi-dimensional perceptual space. Within this perceptual space, decision-rules associate regions of the perceptual space with certain phonological categories. Context sensitivity can arise at both processing stages. If context sensitivity arises as a consequence of *perceptual integration* of target and context, then the position of the target in the perceptual space is already influenced by the context. This may in turn lead to a different phonological categorization by a—context-independent—decision rule. Such a mechanism may, for instance, be assumed for context effects as ‘auditory enhancement’ (Summerfield & Assmann, 1989), ‘spectral contrast’ (Holt & Lotto, 2002; Lotto, Kluender, & Holt, 1997), or other context effects arising at early auditory processing stages (see Delgutte, 1997). However, a context effect can also arise at the level of phonological categorization, without an influence of context on the representation of the input in the perceptual space. In the latter case, the context modifies the decision rule. Then, the

decision rules associate different regions of the perceptual space with different phonological categories depending on the context.

Here, we want to argue that perceptual compensation for phonological assimilation may arise as a consequence of *perceptual integration* of target and context at early auditory levels of processing. Such a perceptual-integration account differs in multiple ways from the other two accounts. First, both the phonological-inference and the feature-parsing account assume that the context sensitivity arises after the acoustic input has been transformed into—continuous or discrete—phonological features. This implies that there is no perceptual integration of target and context at early auditory levels. In contrast, the perceptual-integration account proposes that context effects in the perception of phonologically assimilated words arise precisely at early, language-independent stages of processing.

Because compensation is assumed to arise at early auditory stages, a perceptual-integration account predicts that there is no lexical influence on compensation for assimilation. One would have to assume an influence of the lexicon on auditory processing. However, available evidence (Pitt & McQueen, 1998) indicated that there is no such top-down influence on earlier, phonetic or auditory processing stages. Therefore, it is unlikely that the lexicon influences the assumed auditory locus of the compensation process.

The evidence with regard to the question whether there is an influence of lexical processing on compensation for assimilation is equivocal. Using a phoneme monitoring task, Gaskell and Marslen-Wilson (1998) found that the erroneous detection of a word-final alveolar was more likely in case the erroneous detection yielded a word ('freigh**h** bearer' → 'freight') than in case this yielded a nonword ('preigh**h** bearer' → 'preight'). That is, the /p#b/ sequence was more likely to undergo compensation for assimilation if a /t#b/ interpretation yielded a word. The results reported in chapter three failed to replicate such a 'word-superiority effect.' In this study, German participants were as likely to compensate for place assimilation in a Dutch word as were Dutch participants. A query of the lexical database CELEX (Baayen, Piepenbrock, & Gullikers, 1995) showed that the Dutch stimulus in question was not similar to any German word. Therefore, the evidence with regard to a lexical influence on compensation for phonological assimilation is equivocal.

In addition, the phonological-inference account differs from the two other accounts with regard to the role of learning in establishing perceptual compensation for assimilation. The phonological-inference account assumes that experience with a given assimilation rule is crucial in order to be able to compensate for this assimilation rule. Both the feature-parsing account and the perceptual-integration account assume that such specific language experience is not

necessary. According to the feature-parsing account, compensation for assimilation arises from the application of Gestalt principles to a phonological-feature level. The perceptual-integration account assumes that compensation for assimilation arises from general perceptual principles such as Warren's (1999) criterion shift rule.

The assumption that compensation for assimilation does not need to be learned may seem to propose that a very specific problem in speech recognition—compensation for assimilation—is solved by innate mechanisms. It has been argued that such specific innate dispositions are neurobiologically quite unlikely (Elman, Bates, Johnson, Karmiloff-Smith, & Plunkett, 1996). However, we do not assume that compensation for assimilation may be such a specific problem for the listener, because our assumption is based on the view that phonological rules may be shaped by existing perceptual preferences, that is, 'perception influences phonology' (see, e.g., Hume & Johnson, 2001; Hura, Lindblom, & Diehl, 1992; Kohler, 1990; Ohala, 1990; Seo, 2001; Steriade, 2001). In this view, assimilations are not a problem the speaker imposes on the listener. Instead, assimilations follow the 'a-priori' perceptual preferences of the listener. This link between perception and production may arise in interplay between speaker and listener (de Boer, 2000): The speaker accidentally produces 'ad-hoc' assimilations when reducing articulatory effort. Depending on the perceptual saliency of the assimilation, the listener will 'license' the assimilation or not. If the change is perceptually salient, an articulatory simplification is forbidden; however, if the perceptual saliency of the change is small, the listener will accept the assimilated form, and the simplification is allowed. According to this view, there is no (or little) adaptation necessary for the listener in order to compensate for assimilations. Instead, failures of the listeners in detecting articulatory simplifications, which may arise as a consequence of a perceptual integration of assimilated and assimilating segment, influence the kind of assimilations the speaker is allowed to apply. This perceptual account for assimilated rules resonates with the perceptual-integration account proposed here, assuming that general perceptual processes lead to compensation for assimilation.

The present experiments

The three accounts just sketched differ on three counts, the level at which assimilation occurs, and thus whether there is a 'word-superiority effect' in compensation for assimilation, and the question whether specific language experience is necessary in order to compensate for assimilation. However, there is yet an even more fundamental issue to resolve. All earlier studies on compensation for assimilation made use of place assimilation rules that are quite similar, and all involved strongly related Germanic languages. Therefore, one may question the generality of

the compensation-for-assimilation phenomenon as it has only been shown for very similar assimilation rules in very similar languages. Therefore, the question is whether compensation for assimilation is a general phenomenon or restricted to the place assimilations in Germanic languages.

In order to investigate this issue, we used a manner assimilation rule, liquid assimilation, which is applied in the Hungarian language, a Finno-Ugric language (see Olsson, 1992, p.57; Siptár & Törkenczy, 2000, p. 182): If an apical lateral is followed by an apical trill, then the lateral may also be pronounced as a trill. Thus, the Hungarian word for 'from the left' /bəlro:l/ may be pronounced [bərro:l], but the Hungarian word for 'at the left' /bəlna:l/ may not be pronounced *[bərna:l]. (/bəl/² is the Hungarian word for 'left,' while /ro:l/ and /na:l/ are directional case suffixes.) If compensation for Hungarian liquid assimilation is similar to compensation for major place assimilation in Germanic languages, we can expect that listeners will make a clear distinction between a canonical form [bəlna:l] and an unviable variant *[bərna:l], but do not distinguish a canonical form [bəlro:l] from a viable alternative [bərro:l].

In order to distinguish the three accounts, which explain such a compensation for assimilation, the issue of the level at which compensation for assimilation occurs is crucial. In order to investigate this issue, we made use of a word-identification and a discrimination task. This provides a litmus test for the perceptual-integration account (see Kingston & Macmillan, 1995; Macmillan et al., 1999). The identification task measures which phonological properties listeners associate with a target. Therefore, we expect that listeners will identify the viable alternative [bərro:l] as an instance of the canonical form [bəlro:l], while the unviable alternative *[bərna:l] is not identified as an instance of the canonical form [bəlna:l]. This is predicted by all three accounts. A different prediction, however, can be derived for the discrimination task, in which performance reflects the distance in perceptual space between two stimuli. According to the perceptual-integration account, the context utterance [ro:l] influences the perception of the assimilated form [bər] at early, auditory levels. This influences the distance between [bər] and [bər] in perceptual space. Hence, discrimination between [bər] and [bər] should be more difficult in the context [...ro:l], which allows the assimilation, than in the context [...na:l], which does not allow the assimilation. Both the feature-parsing and the phonological-inference account do not assume that the context effect arises in auditory processing. Hence, the distance in a perceptual

² Note that the written form is 'bal'. However, Siptár and Törkenczy (2000) transcribe the short open, written as "a" vowel as /ɒ/, and we follow their example.

space between an assimilated form and a canonical form is not influenced by the phonological context. Therefore, these accounts predict no influence of context in a discrimination task.

This leaves two issues to be resolved. First, is there a word-superiority effect in compensation for assimilation, and, second, is specific language experience necessary for compensation for assimilation. These issues were investigated by presenting the Hungarian tokens in three conditions: Hungarian listeners hearing Hungarian words, second, Hungarian listeners hearing the stimuli with a changed onset, which yields the Hungarian pseudoword [zɔl...], and Dutch listeners hearing the original Hungarian word. In order to have an identical experimental procedure in all these three conditions, this was operationalized as a between-subject variable. Comparing the two Hungarian groups of listeners with each other allows us to investigate the impact of the lexical status of the stimulus word on the perception of assimilated word forms. Comparing the two Hungarian groups of listeners with the third, the Dutch group allows us to investigate the issue whether perceptual compensation for phonological assimilation depends on specific language experience. The phonology of Dutch has no assimilation rule similar to the liquid assimilation rule in Hungarian (cf. Booij, 1995). If compensation for assimilation depends on experience with an assimilation rule, Dutch listeners should not be able to compensate for the Hungarian liquid assimilation.

It should be noted that the Dutch language has an /l/-/r/ contrast. The /l/ is mostly realized as an apical lateral. The phonetic implementation of the /r/ phoneme varies considerably within the Netherlands, with a uvular trill as the most common exemplar. However, the Hungarian standard of an apical trill is used by a minority of Dutch speakers. Thus, Dutch listeners should clearly hear a phonological opposition between a liquid and a trill, as they are familiar with apical trills and laterals.

Experiment 1: Identification

In this experiment, we tested the identification of a [Cɔl]-[Cɔr] continuum in three conditions: in isolation, in a context that allows a change from /l/ to /r/, and in a context that does not allow this change. Identification was tested in a 2AFC task in which no feedback was supplied. Stimuli were presented to Hungarian listeners with an initial consonant that yielded a word ([bɔl], English 'left') and with an initial consonant that yielded a nonword ([zɔl]). In addition, the Hungarian word was also presented to Dutch listeners.

This experiment evaluates three issues. First, it tests whether compensation for assimilation as reported for place assimilations in Germanic languages generalizes to other assimilation rules in different language families. Second, the comparison of Hungarian words and

nonwords, presented to Hungarian listeners, allows us to evaluate whether there is an influence of lexical status of the stimulus word on compensation for assimilation. Third, probing the perception of Hungarian utterances by Dutch listeners allows us to test whether experience with liquid assimilation is critical for compensation for liquid assimilation.

Method

Participants

All participants were psychology students from the University of Maastricht or the Pázmány Péter Catholic University of Budapest and participated for course credit. All participants were free of a history of hearing problems. The Hungarian participant group hearing the original word utterances consisted of ten listeners (six female, four male) with a mean age of 22.5 years. All participants had some foreign language experience. The English (nine participants) and the German (four participants) language accounted for the majority of the foreign-language knowledge. In addition, two participants were acquainted with Italian and Latin, respectively. However, nobody in the sample had any knowledge of Dutch.

The Hungarian group hearing the Hungarian nonwords consisted of twelve participants (six female, six male) with a mean age of 21.7 years. Two participants had no foreign language skills. The other ten participants all spoke some English, and seven participants also spoke some German. In addition, one participant spoke some Esperanto. Nobody in the sample had any knowledge of Dutch.

In the Dutch group, there were 3 male and 9 female participants, with a mean age of 21.1 years. All participants spoke English and either German or French as foreign languages. However, no participant had any knowledge of Hungarian.

Materials

A female native speaker of Hungarian was recorded uttering multiple tokens of the canonical form [bɒlna:l] (Engl. 'at the left') and an unviable variant *[bɒrna:l] and the canonical [bɒlro:l] form (Engl. 'from the left') and a viable alternative [bɒrro:l]. The Hungarian context suffixes of the cases called Delative (from where?) and Adessive (where at?) were chosen, because they are phonetically quite similar. Both start and end with a sonorant and contain a long vowel. Any differences that these stimuli cause in the perception of the preceding segments can thus not be attributed to gross acoustic differences (overall amplitude, presence of voicing) between the context sounds.

The sample frequency for the recording of the natural utterances was 22050 Hz. Recordings were band-pass filtered from 130 to 8000 Hz. From one utterance of [bɒlna:l], the first syllable was spliced out and edited with the software package PRAAT 4.0 (Boersma & Weenink, 2002). This syllable was then edited in order to create a continuum of speech sounds from the original apical lateral to an apical trill, using the purposefully assimilated utterances as a template.

A linear-predictive-coding analysis with sixteen predictors yielded a stable solution and was used in order to estimate source and filter for this utterance. Given the identical place of articulation, the primary cue for the lateral/trill distinction is the presence of amplitude modulation (AM) in the trill. For the edited sound stimuli, one cycle of AM (20 Hz) was added to the estimated source using five steps from 0 dB (no AM) to 12 dB (strongest AM). This was done by editing the intensity function of the source as estimated by PRAAT. The middle panel of Figure 6.1 shows the intensity as manipulated for the stimulus without AM, the stimulus with 6 dB AM, and the stimulus with 12 dB AM. The source was multiplied by these edited intensity functions.

Formant estimations showed similar formant transitions into the consonant for apical and trills, but a slightly (200 Hz) lower third formant for the resonant period of the trill. Therefore, the third formant was also lowered in five steps of 0.1 bark. Filtering the edited sources with the edited filters yielded five stimuli, in which the first stimulus ended in an unaltered apical trill and the last stimulus emulated a final apical trill. In order to convey a casual speaking style, pre-voicing was shortened from 40 ms to 27 ms by cutting two complete cycles in all stimuli starting with a voiced labial stop. In order to create the Hungarian nonwords, the filter was manipulated for the initial 105 ms of the utterance in order to emulate the estimated filter function in natural utterances of an alveolar voiced fricative of the speaker.

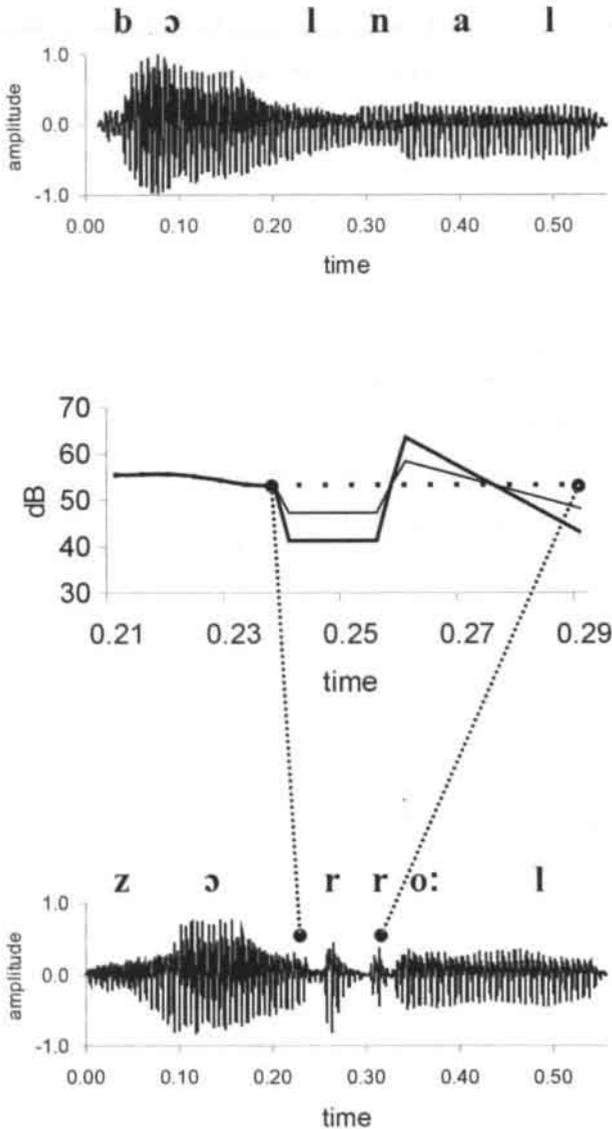


Figure 6.1 The upper panel shows the stimulus 'balnal' without AM. The short silence at the beginning is due to the shortening of the pre-voice bar. The middle panel shows the parameters used to create AM during the /l/ in 'bal' and 'zal.' The lower panel shows the result of this editing, but here with the nonword stimulus 'zarrol.' The first AM at 0.265 is due to the manipulation of an original lateral, while the second peak at 0.315 s stems from the natural suffix 'rol.'

For the first 75 ms, the filter settings were stable, and from 75 to 105 ms, the filter settings were interpolated from the /z/ settings to the original /ɔ/ settings using the cosine function in $[0, \pi]$. This yields a slow initiation of the formant movement, a maximal acceleration at the midpoint slowing down again toward the endpoint. In addition, the original voice source was attenuated to one half of the original amplitude and noise was added for the first 75 ms. From 75 ms to 105 ms the noise was faded out and the original voiced source was restored to the original level using linear interpolation. The acoustic changes, with which to achieve a continuum ranging from an apical lateral to an apical trill in the post-vocalic position, were identical to the Hungarian word continuum. These [zɔl]-[zɔr] stimuli were 13 ms longer than the [bɔl]-[bɔr] stimuli, reflecting the fact that fricatives are intrinsically longer than stops.

The context case suffixes [na:l] (Engl. 'to' in answer to the question 'where?' as in 'to your left') and [ro:l] (Engl. 'from') were spliced from other utterances, equalized in overall energy and concatenated with the Hungarian word and nonwords stimuli. The amplitude relation of words and nonwords to the case suffixes were edited to emulate the amplitude relation of the first (always stressed) syllable to the second syllable in the natural utterances. Figure 6.1 shows two stimuli that resulted from the editing procedure. The upper panel shows the stimulus [bɔlna:l] in which the first syllable has no amplitude modulation and the second syllable is the case suffix [na:l] that does not allow the assimilation of /l/ to /r/. The lowest panel shows the stimulus [zɔrro:l] in which the original /b/ onset was manipulated as described above. The first AM peak at 0.265 s is the result of the editing of the intensity function as described above. The second peak at 0.315 s stems from the natural trill of the case suffix [ro:l]. These stimuli were presented to the participants via headphones (SONY MDR-V 900) using E-prime (Psychology Software Tools, Version Beta 5.0, for the Hungarian listeners with Hungarian words), the Presentation software (Neurobehavioral Systems, for the Hungarian listeners with Hungarian nonwords), or ERTS (Behringer, 1996, for the Dutch listeners).

Procedure

Experiments were run with participants facing a computer screen and instruction was given in written form via the computer screen. All participants were instructed that they were going to hear the Hungarian word for left 'bal'—or the nonword 'zal'—spoken by a Hungarian speaker who sometimes make an error, and pronounces 'bal' as 'bar.' Participants were asked to indicate, after hearing a speech sound, whether the speech sound was 'bal' or 'bar.' In order to prevent any stimulus-response incompatibility for the Hungarian participants, participants were instructed to hit the left key when hearing 'bal' (Hungarian for 'left') and the right key when hearing 'bar' (of either the computer keyboard for Hungarian listeners, or the left button of a ERTS response box for the Dutch listeners).

500 ms before hearing a stimulus, the computer screen displayed the two answer alternatives ('bal' versus 'bar' or 'zal' versus 'zar') on the left and right of the screen, corresponding to the response key allocation. After hearing a stimulus participants had 2.5 s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, the word on the screen corresponding to the response alternative was moved up and to the margin of the computer screen by 6 pixels while the other alternative disappeared. This indicated to the participants that their answer had been registered by the computer.

Conditions of context suffixes (no suffix, unviable suffix [na:l], and viable suffix [ro:l]) were blocked. The order of presentation was randomized within blocks. All participants started by judging the stimuli without context suffixes. Then, the presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants first heard the stimuli with the viable suffix, while the other half first heard the stimuli with the unviable suffix.

Design

The design entails three independent variables, one between-subject variable with three levels and two within-subjects variables with three and five levels respectively. The between-subject variable is Listener group with the levels Native Listener-Word, Native Listener-Nonword, and Nonnative Listener. The two within-subject variables are Context (none, viable, unviable) and AM depth (five levels). There were fifteen measurements for every cell in this design. The dependent variable is the percentage of 'bal' responses calculated from the fifteen trials per cell.

Results

The mean percentages of [Cɔl] responses are shown in Figure 6.2 for all participant groups. A repeated-measurement ANOVA with participant group as between subjects variable and AM depth and Context (none, viable, unviable) as within-subject variables revealed a

significant effect of AM depth ($F(4, 124) = 290.85, p < 0.001$), a significant effect of Context ($F(2, 62) = 24.93, p < 0.001$), and a trend for the variable listener group. ($F(2, 31) = 2.94, p < 0.1$). These simple effects were, however, qualified by significant two-way interactions between Context and AM depth ($F(8, 248) = 39.10, p < 0.001$), AM depth and listener group ($F(8, 124) = 3.07, p < 0.005$) and a significant three-way interaction ($F(16, 248) = 2.65, p < 0.005$).

In order to investigate the nature of these interactions, we broke down the factorial design in order to evaluate simple effects that were not qualified by an interaction (cf. Keppel, 1991, p. 236): First, we examined the effect of the factors Context and listener group for each level of AM depth separately. If the interaction of this analysis is significant, we then examine the effect of group for each level of context. In case there was a significant effect of listener group, the source of this effect was investigated by linearly independent planned comparisons to evaluate the effect of Wordness (Hungarian listeners with words versus Hungarian listeners with nonwords) and the effect of Native Language (Hungarian listeners versus Dutch listeners).

This procedure yielded the following results: For the stimuli without AM modulation, there was a significant effect of Context ($F(2, 62) = 12.3, p < 0.001$) and Group ($F(2, 31) = 8.9, p < 0.01$). These effects were, however, qualified by a significant interaction of Group and Context ($F(4, 62) = 2.7, p < 0.05$). Separate ANOVAS for all levels of Context showed that there was an effect of Group for the stimulus in the viable context ($F(2, 31) = 6.3, p < 0.01$), but only a trend in the unviable context ($F(2, 31) = 3.1, p < 0.1$), and no significant effect for the stimulus without context ($F(2, 31) = 1.8, p > 0.1$). Planned comparisons for the viable-context stimulus yielded a significant effect for Native Language ($t(31) = 3.2, p < 0.05$), but no effect of Wordness ($t(31) = 1.6, p > 0.1$). As Figure 6.2 shows, Dutch listeners were more inclined to hear a trill in the viable context, even when there was no AM.

For the stimuli with 3 dB AM, there was a significant effect of Context ($F(2, 62) = 7.1, p < 0.01$) and Group ($F(2, 31) = 8.3, p < 0.01$), but no interaction ($F(4, 62) = 1.2, p > 0.1$). Planned comparisons for the listener group showed that Wordness did not have a significant influence ($t(31) = 1.1, p > 0.1$), but Native Language did ($t(31) = 4.0, p < 0.001$). Dutch listeners were overall less inclined to hear a lateral in this condition (73.9%) than Hungarian listeners (88.9%). Post-hoc tests (HSD, $p < 0.05$) for the different levels of the Context factor showed that the stimuli in the no-context condition were more often labeled as a lateral (91.3 %) than the viable-context (77.1 %) and the unviable-context (83.7 %) condition. The two context conditions did not differ significantly.

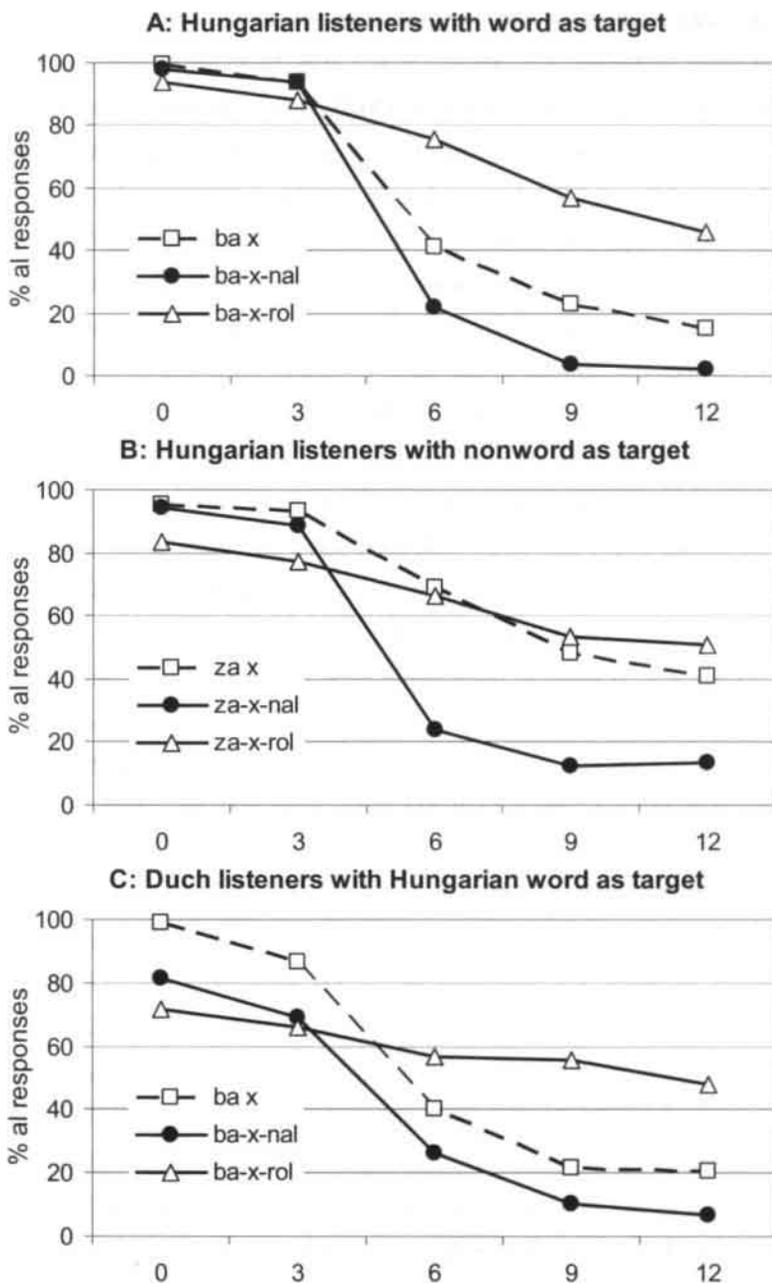


Figure 6.2 Identification performance in Experiment 1 as percentage 'Cal' identifications. Figure 6.2A shows the results for the Hungarian listeners with Hungarian words, Figure 6.2B shows the results for the Hungarian listeners with Hungarian nonwords, and Figure 6.2C shows the results for the Dutch listeners with Hungarian words.

For the stimulus with 6 dB AM, there was a significant effect of Context ($F(2, 62) = 40.8$, $p < 0.01$), no effect of Group ($F(2, 31) = 1.9$, $p > 0.1$). In addition, there was a significant interaction ($F(4, 62) = 3.7$, $p < 0.01$). Separate ANOVAs for each level of Context yielded an effect of Group for the stimulus without context ($F(2, 31) = 3.8$, $p < 0.05$) and the viable-context stimulus ($F(2, 31) = 3.7$, $p < 0.05$), but not for the unviable-context stimulus ($F < 1$). For the stimulus without context, planned comparisons revealed a significant effect of Wordness ($t(31) = 2.2$, $p < 0.05$), but no effect of Native Language ($t(31) = 1.5$, $p > 0.1$). The stimulus in the word condition was more likely to be perceived as a trill (58.8%) than the stimulus in the nonword condition (31.1%). For the stimulus in the viable condition, there was no effect of Wordness ($t(31) = 1.4$, $p > 0.1$), but a significant effect of Native Language ($t(31) = 2.5$, $p < 0.025$). The stimulus with medium AM in the viable context condition was more likely to be perceived with the 'canonical' lateral by the Hungarian listeners (70.4%) than by the Dutch listeners (56.9%).

For the stimuli with 9 and 12 dB AM, there was a significant effect of Context (9 dB: $F(2, 62) = 39.0$, $p < 0.001$, 12 dB: $F(2, 62) = 39.0$, $p < 0.001$), while both the Group variable (9 dB: $F(1, 31) = 1.5$, $p > 0.1$, 12 dB: $F(1, 31) = 3.0$, $p > 0.1$), and the interaction (9 dB: $F(4, 62) = 1.9$, $p > 0.1$, 12 dB: $F < 1$) failed to reach significance. For both levels, post-hoc tests (HSD) revealed that all context conditions differed significantly from each other in how often the stimuli were labeled as a lateral (no context: 31.4% at 9dB AM and 26.2% at 12dB AM, unviable context: 8.9% at 9dB AM and 7.7% at 12dB AM, viable context: 55.4% at 9dB AM and 48.2% at 12dB AM).

In summary, the results show that the AM was successful in leading to a perceptual change from /l/ to /r/ in the unviable and, to a lesser degree, in the no-context condition. In the viable-context condition, however, even strong AM modulation does not lead to a clear percept of an apical trill. This general pattern is moderated by the fact that Dutch listeners were less likely to perceive the 'canonical' lateral in the viable context condition than the Hungarian listeners. In addition, there was a difference between the word and nonword condition in how the medium AM was perceived without context. Moderate AM was more likely to be perceived as a trill in a word than in a nonword. Most importantly, however, all listener groups perceived the maximal AM modulation as a trill in the unviable context, but as ambiguous in the viable context condition.

Discussion

The most basic question in this experiment was whether perceptual compensation for phonological assimilation, as reported for major place assimilation of word-final nasals and stops in Germanic languages, generalizes to an example of a manner assimilation rule in a Finno-Ugric

language. Earlier studies (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2002a; Coenen et al., 2001, see also chapter three) have shown that assimilated word forms are only accepted in contexts in which the assimilation is allowed. This is replicated with Hungarian listeners: A deviation from a canonical form is more likely to be perceived as an instance of the canonical form if the context allows the deviation. Thus, [bɔr] is perceived as similar to the canonical form of the Hungarian word for 'left,' /bɔl/, only when the context allows the assimilation of /l/ to /r/.

A second question was whether compensation for assimilation is influenced by the lexical status of the assimilated word. This does not seem to be the case. In the viable and unviable context, the speech-sound continuum from [Cɔ] to [Cɔr] was perceived similarly in both the word and the nonword condition. This replicates the results reported in chapter three. They showed that German and Dutch listeners compensated for assimilation equivalently when presented with a Dutch word, a German nonword, to which an assimilation has been applied that was viable in the German and the Dutch language.

A third question was whether compensation for liquid assimilation is dependent on experience with this assimilation. It seems that compensation for assimilation is influenced by specific language experience. Hungarian listeners were more inclined to respond with the canonical form [Cɔ] than Dutch listeners. However, the results also show that context sensitivity in the perception of assimilated utterances is not wholly dependent on specific language experience. The Dutch listeners showed an overall pattern similar to that of the Hungarian listeners. They clearly perceived an opposition between a lateral and a trill in the unviable context but there was no clear category shift in the viable context. While both groups had a steeper identification function in the unviable context than in the viable context, only the native listeners showed a bias towards the 'canonical' form. This replicates the results from chapter four. There, we presented Dutch stimuli with nasal place assimilation to Portuguese listeners, who have no experience with nasal place assimilation. Both native and nonnative listeners found it more difficult to perceive the difference between an alveolar and a labial nasal (/n/ versus /m/) in a context that allowed an alveolar to assimilate to a labial ([tɔɛnbank], Engl. 'garden bench') than in a context that did allow such an assimilation ([tɔɛnstu'l], Engl. 'garden chair'). However, not all effects were independent of language experience. In agreement with the present results, only listeners with experience with nasal place assimilation tended to resolve this ambiguity with a bias toward the 'canonical' /n/.

An unexpected result was the finding that perception of the speech-sound continuum was more categorical in the unviable context than in the no-context condition. A likely interpretation of this result is that, in phrase-final position, a trill with one AM period is not an acceptable phonetic implementation of a trill. However, within a phrase, this is a valid implementation of a trill. Therefore, we find a strong perceptual switch in the unviable context condition (one AM period within a phrase), but less so in the no-context condition (one AM period in phrase-final position).

The current experiment succeeded in showing that perceptual compensation for a phonological assimilation is not restricted to major place assimilation in Germanic languages. In addition, it seems that the lexical status of the assimilated form does not have an influence on perceptual compensation for assimilation, while experience with this assimilation only sharpens, but not completely governs, the perception of assimilated forms.

One issue remains unresolved. It is still unclear whether compensation for phonological assimilation is based on higher-level language processes or on a perceptual integration of target and context at auditory levels of processing. In order to investigate this, we also performed a discrimination task, in which participants were not asked to label the stimuli, but to discriminate them. If perceptual compensation for phonological assimilation is based on phonological analyses, earlier processing levels should be context-insensitive. Using a discrimination task with direct feedback should incline listeners to probe an auditory level of analysis (see Beddor & Krakow 1999, for a similar view). If the context sensitivity in the perception of assimilated utterances is based on a late, language-specific processing stage, listeners should be less influenced by context in a discrimination task. If however, perceptual compensation for phonological assimilation is based on early processing levels, listeners should be influenced by context in the discrimination task just as in the identification task.

Experiment 2: Discrimination

In this experiment, we used the same stimuli as in Experiment 1 but employed a discrimination task. The phonological-inference account and the feature-parsing account should predict that performance in a discrimination task is not influenced by the phonological categorization of the stimuli in question. In order to achieve this, we used the four-interval oddity (4I-oddity) task. Gerrits (2001, chapter 3; see also Schouten, Gerrits, & van Hessen, 2003) showed that, out of the commonly used discrimination tasks, the 4I-oddity task shows the least influence of the phonological categorization on the to-be-discriminated stimuli. In this task, the likelihood that participants are able to discriminate between two stimuli is the same for within-

category and between-category pairs, given the same acoustic difference. Therefore, we may expect a dissociation between identification and discrimination if compensation for assimilation is based on phonological processing, be it in the form of regressive inference or feature parsing. However, a perceptual-integration account predicts that discrimination performance is influenced by context.

In the 4I-odddity task, four stimuli presented at a constant inter-stimulus interval (ISI), of which three are identical (the standard) and one is different (the 'odd'), are presented to the listener. The listener is (correctly) informed that either the second or the third is the 'odd,' and the task of the listeners is to indicate whether the second or third stimulus was the odd.

In order to prevent a combinatorial explosion of conditions, we did not probe all possible contrasts. As we are interested in how far listeners are able to detect a mismatch with a canonical pronunciation, only the original stimulus with a lateral /l/ was used as standard in the 4I oddity task, while the stimuli with differing degrees of AM served as odds.

Method

Participants

The same three participant groups as in Experiment 1 participated in Experiment 2. That is, ten Hungarian participants performed the discrimination task with the Hungarian word stimuli; twelve Hungarian participants performed the task with the Hungarian nonword stimuli; and twelve Dutch participants performed the task with the Hungarian word stimuli. The discrimination task followed the identification task in all groups.

Materials

The speech sound stimuli were the same as in Experiment 1. Sound files for the discrimination task were created before the experiment to prevent on-line timing problems. For each trial, four sounds were concatenated with an ISI of 425 ms. The standard stimulus was always the sound without AM and the deviant was one of the stimuli with AM. This odd stimulus was either in the second or the third position in the train of four stimuli.

Procedure

Experiments were run with participants facing a computer screen. Instructions were also presented via the computer screen. All participants were instructed that they would hear a series of four stimuli, in which either the second or the third stimulus differed from the other stimuli. They were explicitly instructed that two sounds, which might be written in the same way, may still differ in how they sound. Participants were asked to indicate which one was the odd after hearing a series of four speech sounds.

500 ms before hearing a stimulus consisting of four speech sounds, the computer screen displayed the two answer alternatives (the Hungarian or the Dutch words for 'two' and 'three') on the left and right of the screen, corresponding to the response key allocation. After hearing a train of four speech sounds, participants had 3 s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, feedback indicated whether the choice was correct or not.

Conditions of context suffixes (unviable suffix [na:l] and viable suffix [ro:l]) were blocked in four blocks of forty trials each. Within blocks, order of presentation was randomized. The presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants heard the stimuli in the viable context in the first and third block and the stimuli in the unviable context in the second and fourth block. For the other half, the unviable context stimuli were presented in the first and third block and the viable-context stimuli in the second and third block.

Design

The design entails three independent variables, one between-subject variable with three levels and two within-subject variables with two and four levels respectively. The between subject variable is Listener group with the levels Native Listener-Word, Native Listener-Nonword, and Nonnative Listener. The two within-subject variables are Context (viable or unviable) and ΔAM between the standard and the odd stimulus with four levels from 3 dB ΔAM to 12 dB ΔAM .

Ten trials each were presented with the odd stimulus in the second and third position for all eight condition of conditions arising from the crossing of the ΔAM and the Context factor. The dependent variable d' was calculated from the 20 trials per cell. We arbitrarily defined the second position as the target. Hence, a correct response in a trial with the odd stimulus at the second position is counted as a hit, while an error in a trial with the odd stimulus at the third position is counted as a false alarm. In order to calculate d' we forcibly reduced the range of the percentages of correct responses from 0.1% to 99.9%.

Results

Figure 6.3 shows the mean d' data for all conditions. An ANOVA with ΔAM , Context (viable versus unviable), and Listener Group revealed a significant effect of Context ($F(1, 31) = 115.8, p < 0.001$) and a significant effect of ΔAM ($F(3, 93) = 49.0, p < 0.001$). These two within-subject variables did not interact significantly ($F(3, 93) = 1.7, p > 0.1$). Neither the main effect of Listener Group ($F < 1$) nor any of its interactions (by Context: $F < 1$, by ΔAM : $F(5, 93) = 1.9, p = 0.08$, by Context-by- ΔAM : $F < 1$) was significant.

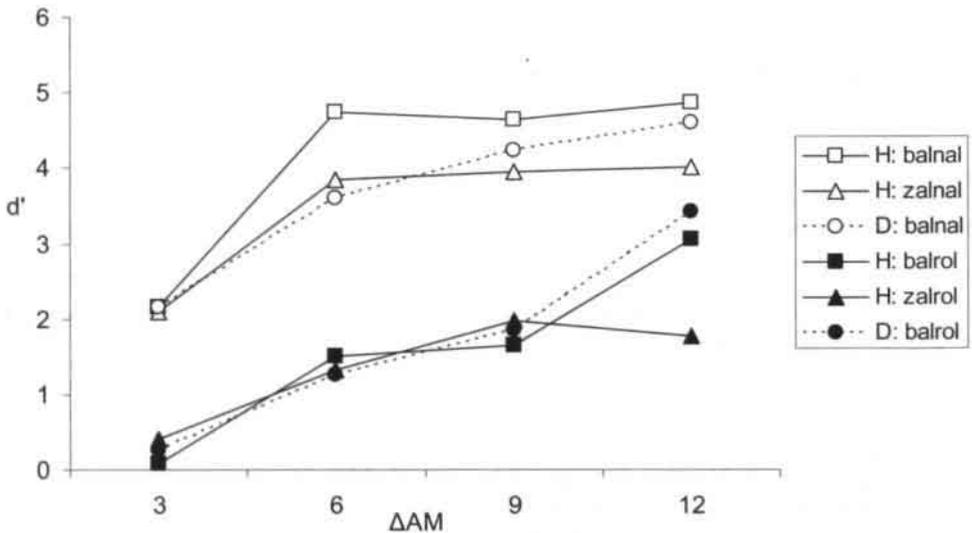


Figure 6.3 Discrimination performance in Experiment 2. The open symbols represent the data for the unviable-context condition and the closed symbols represent the unviable conditions. The squares represent the data of the Hungarian listeners with the Hungarian word, the triangles represent the data of the Hungarian listeners with the Hungarian nonword, and the circles represent the data of the Dutch listeners with the Hungarian word.

The main effect of Context shows that discrimination was significantly better in the unviable context ($d' = 3.76$) than in the viable context ($d' = 1.54$). In order to evaluate the main effect of ΔAM , post-hoc tests (HSD, $p < 0.05$) were performed. This showed that the level 3 dB ΔAM ($d' = 1.19$) was significantly different from all other levels. The levels 6 dB ΔAM ($d' = 2.71$) and 9 dB ΔAM ($d' = 3.05$) did not differ significantly. In addition, discrimination performance was better at the maximal ΔAM ($d' = 3.62$) than at the level 6 dB ΔAM .

Discussion

In this experiment, we examined whether the context effects in the perception of phonological changes extend to a discrimination task. This was clearly the case. Even in a discrimination task, performance is context-dependent. Deviations from a canonical form [Ca] were less salient in the viable context than in the unviable context. This was independent of the lexical status of the assimilated form and independent of experience with the assimilation rule in question.

These results support the assumption that the context sensitivity in the perception of assimilated word forms arises at early, auditory levels of processing. The phonological-inference and the feature-parsing account cannot explain the present result unless they assume that phonological categorization influences discrimination performance. Two arguments make this seem an unlikely assertion. First, the discrimination task used does not give rise to a between-category advantage in discrimination (Gerrits, 2001). Second, the present results indicate a dissociation between identification and discrimination. Experience with the assimilation rule influenced performance in the identification task in Experiment 1, but not in the discrimination task. This indicates that the identification task and the discrimination task probe different levels of processing.

General Discussion

In the current study, we set out, first, to investigate the generality of the compensation-for-assimilation process that was first stipulated by Gaskell and Marslen-Wilson (1996). This was done by probing the perception of morphologically complex Hungarian utterances in an identification task and a discrimination task. The utterances were presented in the canonical form and in a changed form in which liquid assimilation had been applied. Assimilated and canonical utterances were presented in a context in which the assimilation might occur and in a context in which it was not allowed to occur. In addition, we estimated the influence of the lexical status of the assimilated form as well as the role of experience with the Hungarian assimilation rule for the

perception of phonological assimilations. This was done by presenting, first, Hungarian words to Hungarian listeners, second, Hungarian nonwords to Hungarian listeners, and, third, Hungarian words to Dutch listeners.

The results showed that perception of assimilated utterances is context-sensitive, independent of lexical status and independent of experience with the Hungarian assimilation rule. In the unviable context, Hungarian and Dutch participants clearly distinguished changed and unchanged utterances, while, in the viable context, all participants were less able to distinguish between a canonical and a changed form. This was observed in the identification task as well as in the discrimination task. However, in the identification task, but not in the discrimination task, this general context effect was also modified by experience with Hungarian liquid assimilation. While Dutch listeners perceive all speech sounds in the viable condition as more or less ambiguous, Hungarian listeners showed a preference for the canonical interpretation of the utterance.

First of all, the results indicate that compensation for assimilation is a general process, which is not restricted to major place assimilation in Germanic languages. Using liquid assimilation in Hungarian, a Finno-Ugric language, we found results that resemble earlier results obtained with place assimilation in English (Gaskell & Marslen-Wilson, 1996; Gow, 2002a), Dutch (chapter three), and German (Coenen, Zwitserlood, & Boelte, 2001). In the viable context, the changed form is perceived as similar to the canonical form, while this is not the case in the unviable context.

The obtained results allow us to evaluate the three accounts for compensation for assimilation. The phonological-inference account, the feature-parsing, and the perceptual-integration account differed on three counts. First, there is the issue at which level compensation occurs. This was investigated with the use of both an identification task and a discrimination task. Both the feature-parsing and the phonological-inference account assume that compensation for assimilation occurs at a level at which the input is represented as phonological features, be it in a discrete or continuous fashion. As a consequence of this assumption, only identification but not discrimination performance should be influenced by the phonological viability of the change. However, this was not the case; there was a strong context effect in both tasks. This favors the perceptual-integration account, which predicts an effect of context in both tasks.

It may be possible to reconcile the feature-parsing account with the discrimination data. If one assumes that phonological features are the objects of perception, preceding levels may not be accessible for conscious report as necessary in a discrimination task. Note, however, that this is quite unlikely, given findings that within-category discrimination of speech is mostly above

chance level (Lieberman, Harris, Hoffman, & Griffith, 1957; Massaro & Cohen, 1983). This contrasts with the assumption that phonological features are the objects of perception used in a discrimination task. Nevertheless, there is another notable difference between the feature-parsing account and a perceptual-integration account. The latter does not assume speech-specificity of the effects; therefore, it should be possible to generate similar context effects with non-speech context sounds. Currently, we are running experiments with non-speech analogues of the stimuli used here in order to settle this issue.

In relation to the question of the level at which compensation occurs, we evaluated whether the lexical status of the assimilated word influences compensation for assimilation. This was not the case in the present results. The word and nonword stimuli gave rise to equivalent context effects. This replicates the results reported in chapter three, but contradicts the earlier results by Gaskell and Marslen-Wilson (1998). In their experiment, participants were performing a phoneme monitoring task and were more likely to give a 'compensatory response' (hearing a /t/ in 'freigh~~p~~ bearer') if the compensatory response yielded a word (here: 'freight'). Given these repeated negative results to replicate a 'word-superiority effect' on perceptual compensation for phonological assimilation in a word-identification task, it may be asked whether the reported lexical influence was introduced by the use of a phoneme-monitoring task. Recent electrophysiological results revealed that the phoneme-monitoring task cannot be performed by relying on prelexical representations. Thierry, Doyon, and Démonet (1998) showed that lexical access occurs before information about individual phonemes of a word is available. This might account for the fact that more compensatory responses are found for words than for nonwords in a phoneme-monitoring task in the Gaskell and Marslen-Wilson (1998) study. Our results indicate that lexical influences do not appear in an identification task, which does not necessitate access to individual phonemes.

In consequence, our results indicate a pre-lexical locus of the compensation-for-assimilation process. This can, in principle, be accounted for by all three accounts. However, in their current form, the phonological-inference account (Gaskell & Marslen-Wilson, 1998) assumes that there is a 'word-superiority effect,' and the feature-parsing account (Gow, 2003) does not preclude a lexical effect. Although our findings regarding the role of the lexicon in compensation for assimilation may thus not be critical in deciding between the different accounts, they certainly constrain the precise implementations of the possible accounts by showing that there is no 'word-superiority effect' on compensation for assimilation.

A final issue that distinguishes the different accounts is the question whether listeners need experience with a given assimilation rule in order to develop adequate compensation for

this assimilation. In contrast to the phonological-inference account, both the feature-parsing account and the perceptual-integration account assume that specific language experience is not necessary. The answer to the question whether specific language experience is crucial for compensation for assimilation turns out to be more complicated. If the question is whether experience influences the perception of assimilated utterances, the answer is clearly yes. Hungarian listeners were more likely to respond with the canonical form in the viable context than Dutch listeners. This result indicates that native-language experience makes a difference. However, if the question is whether experience is pivotal for context sensitivity in the perception of assimilated utterances, the answer is clearly no. In the discrimination task, there was a strong effect of context independent of specific language experience. Similarly, the overall pattern of performance in the identification task—steeper identification functions in the unviable than in the viable context—was the same for Dutch and Hungarian listeners. Therefore, it seems that compensation for assimilation is mainly driven by a perceptual integration of target and context, independent of specific language experience. Language experience only moderates these language independent biases.

The fact that the perception of assimilated utterances proceeds mainly unaffected by specific language experience also supports the related assumption that perceptual constraints influence the shape of phonological assimilation rules, in line with the ‘perception influences phonology’ view (Hume & Johnson, 2001; Ohala, 1990). It seems that the assimilations observed in languages today are restricted to cases in which the assimilated form is not saliently different from the canonical form, independent of any experience with assimilation rules. Thus far, such perceptual explanations for the shape of attested assimilation rules have mostly considered the strength of the cues for a to-be-assimilated segment in its position within the syllable (Steriade, 2001). Here, we want to add that the perceptual saliency of a to-be-assimilated target x may not just be a property of this target. In addition, a *perceptual integration* of the acoustic properties of segment x with the assimilating segment $x\pm 1$ seems to influence the saliency of the segment x , which in turn allows this segment to be assimilated. That is, certain assimilations may be ‘perceptually licensed,’ independent of language experience, because the assimilatory change of the target *in its specific context* is not salient.

In this context, it is important to note that the model of phonological inference is particularly questioned by the fact that assimilations tend to be constrained by perceptual saliency. Phonological inference is supposed to rest on the learning of the co-variance of phonological features (Gaskell et al., 1995). Therefore, the model can in principle account for the compensation for any assimilation, as long as the co-variance of phonological features makes the

assimilation predictable. Hence, the model seems able to learn to compensate for unlikely, but very salient assimilation rules as, for instance, the imaginary rule that fricatives become vowels if followed by nasals. Most assimilation rules, however, involve feature changes that are perceptually difficult to detect (see Kohler, 1990). This, in consequence, questions the assumption that assimilations are compensated for by a model that learns to predict assimilations on the basis of feature co-occurrence.

However, it should also be noted that the current data partly confirm the predictions of a phonological-inference account. There is a selective effect of specific language experience on the *identification* of assimilated word forms, as predicted by the phonological-inference account. Native listeners were more likely than non-native listeners to perceive the stimuli in the viable context as canonical forms. In addition, this effect seems to be located at a phonological processing stage, as it was only evident in the identification task, but not in the discrimination task. This may indicate that listeners learn which phonemes of their native language are prone to assimilation. When in doubt about the identity of a given phoneme, listeners may be biased by language experience to perceive those phonemes which may undergo assimilation. However, the phonological-inference account is difficult to reconcile with the finding that the context effects on the perception of assimilated forms are to a large part independent of language experience, and that the context effects extend to a discrimination task.

The feature-parsing account, in contrast, is supported by the finding that context effects are largely independent of specific language experience. However, this account is difficult to reconcile with the results of the discrimination task. In contrast with the prediction of this account, there was a strong context effect on discrimination performance. In addition, the effect of language experience on identification performance is also difficult to capture in the framework of feature parsing.

The perceptual-integration account, on the other hand, can easily account for the effects of context on discrimination performance and for the independence of the context effects from specific language experience. However, it also leaves unexplained the effect of language experience on identification performance. Therefore, it seems that at present the perception of assimilated word forms can best be accounted for by assuming that a process of perceptual integration of target and context drives the perception of assimilated utterances on auditory levels of processing. This basic effect is moderated by language experience, probably due to the acquisition of (implicit) phonological knowledge about assimilation rules of one's native language.

The Automatic Processing of Assimilated Word Forms Indexed with the MMN¹

In a series of three experiments, the automatic, pre-attentive processing of utterances that have been changed in a viable and unviable way according to rules of phonology of Hungarian is investigated by measuring the event-related brain potentials during passive listening to oddball series. The main hypothesis is that the mismatch negativity (MMN) evoked by the stimulus deviation in the oddball series is larger if the deviant is an unviable alternative pronunciation of the standard than in cases where the deviant is a viable alternative pronunciation of the standard. In addition, the role of lexical status of the assimilated word and the experience of the listeners with the assimilation rule is varied by presenting the utterances to Hungarian and Dutch listeners. Results indicate that neither lexical status nor language experience contribute to the pre-attentive processing of assimilated utterances. Instead, acoustic details of the assimilated utterance play a pivotal role in how the utterance is processed on pre-attentive processing levels.

¹ This is an adapted version of Mitterer, H., Csépe, V., Honbolygo, F., Vlamings, P., and Blomert, L. (submitted). The automatic processing of assimilated word forms indexed with the MMN.

How is a spoken word recognized? According to the predominant view, the acoustic input is first transformed into a more abstract code. This abstract code is then used to achieve lexical access (McQueen & Cutler, 2001). Although the grain size of the basic unit of this code is still under debate—with proposals ranging from phonological features (Marslen-Wilson & Warren, 1994) to demi-syllables (Massaro, 1998)—most psycholinguistic approaches converge on the assumption of intermediate representation of a certain amount of abstraction in spoken-word recognition. However, all of these proposals put a rather heavy load on the assumed pre-processing mechanisms that have to transform the acoustic input into some form of such an intermediate representation. Achieving an intermediate, phoneme-like representation of an auditory input is hampered by a number of non-linguistic influences on the speech signal. The acoustic implementation of, for instance, a phoneme depends, amongst other influences, on the speech rate, the surrounding phonemes (Gussenhoven & Jacobs, 1998; Farnetani, 1997), the likelihood of the word in the sentence context (Jufursky, Bell, & Girand, 2002), as well as the height, sex, age, emotional state, and sociological status of the speaker (van Bezooijen, 1995; Ladefoged, 1996; Lerner & Loewe, 1956; Traunmüller & Eriksson 2000). Given this list of variation-adding influences, it becomes evident that the pre-processing mechanisms, which transform the acoustic input into an intermediate representation before lexical access, face a rather complicated task.

However, even if the acoustic realization of a phoneme varies strongly due to the variation-adding influences, an intermediate representation may still be achieved by a context-sensitive 'phoneme-detection mechanism.' Indeed, for all influences listed above, which add variance to the speech signal, evidence has been found for a context-sensitive interpretation of the speech signal (Beddor & Krakow, 1999; Evans & Iverson, 2002, Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2001, 2002a, 2003; Fowler & Brown, 2000; Fowler, Brown, & Mann, 2000; Ladefoged & Broadbent, 1957; Lotto & Kluender, 1998; Lotto & Kluender, & Holt, 1997; Mann, 1980; Mann & Repp, 1981; Nearey, 1989; Sussman, Fruchter, Hilbert, Sirosh, 1998). These studies have probed context sensitivity of the (speech)-perception system by presenting a given stimulus in different acoustic and phonetic contexts. In all these studies, it has been shown that the speech-perception system is context sensitive, so that the percept derived from a speech sound depends on its context. In addition, the context sensitivity is usually 'compensatory.' The context effects in speech perception are the *inverse* of the variation-adding influences in speech production. For instance, as a vowel may become nasalized in production due to coarticulation with an adjacent nasal consonant, some nasalization is parsed from a vowel if it is accompanied by a nasal consonant in perception (see, e.g., Beddor & Krakow, 1999). In effect, this may yield a

more or less stable phoneme-like code that resists context-dependent perturbations occurring during production.

Another instance of such an inverse relation between speech production and speech perception is observed with phonological assimilations. Phonological assimilations are driven by context-dependent rules that regulate how a phoneme may be influenced by a following or preceding phoneme. Such assimilation rules are language specific (Gussenhoven & Jacobs, 1997), although some general tendencies were also reported (Kohler, 1990; Steriade, 2001). A typical exemplar of an assimilation rule is the rule for coronal place assimilations in English. According to this rule, word-final coronal stops and nasals may take over the place of articulation of the subsequent phoneme. Hence, the phrase 'freight bearer' may be pronounced as 'freighp bearer,' but the phrase 'freight carrier' may not be pronounced as 'freighp carrier.' In perception, the opposite effect has been observed. A word-final /p/ is sometimes perceived as a /t/ in 'freighp bearer' but less so in 'freighp carrier' (Gaskell & Marslen-Wilson, 1998). Hence, the context-sensitive change in production—/t/ becomes /p/ before /b/, but not before /k/—is countered by an inverse context-sensitive change in perception—/p/ becomes /t/ before /b/ but not before /k/. The existence of such a perceptual compensation for phonological assimilations, based on a context-sensitive interpretation of phonological changes, is now well established (Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2002a, 2003; this dissertation, chapter three). However, the mechanisms that drive this context sensitivity are still under debate (Gaskell, 2002; Gow, 2001, 2002a, 2003).

Especially two issues have evoked controversy. First, there is the question of a possible influence of the lexicon on this 'compensation for assimilation.' This issue bears considerable theoretical weight, as it might indicate in whether there is feedback from the lexicon on pre-lexical processing (McQueen & Cutler, 2001; Pitt & McQueen, 1998). Gaskell and Marslen-Wilson (1998) argued that compensation is more likely if a compensatory response yields a word ('freighp bearer' plus compensation yields 'freight') than in cases it yields a nonword ('preighp bearer' plus compensation yields 'preight'). However, in chapter three, we failed to find evidence for the supposed influence of lexical processing on compensation for assimilation.

A second issue focuses on the question whether experience with an assimilation rule is necessary in order to be able to compensate for phonological assimilations. There is some evidence that suggests that the listener must be exposed to a certain amount of co-variation—as, for instance, the fact that English vowels are commonly nasalized in the neighborhood of a nasal consonant—before the inverse context effect can arise in perception (Beddor & Krakow, 1999; Holt, Lotto, Kluender, 2001; Sinnot & Sapirota, 2000). However, there is also evidence

indicating that some context sensitivities—for instance, compensation for coarticulation—can arise without learning (Fowler, Best, & McRoberts, 1990; Lotto, Kluender, & Holt, 1997). Therefore, the question arises whether compensation for assimilation depends on experience with the specific assimilation rules of a given language. In their model of phonological inference, Gaskell and Marslen-Wilson (1998) pleaded for a co-variant learning mechanism leading to compensation for assimilations.

However, there is little evidence—whether supporting or contradictory—for the supposed lexical influence on compensation for assimilation as well as for impact of native-language experience with an assimilation rule on the processing of assimilated word forms. Therefore, we (see chapter six) investigated the role of these variables in a cross-linguistic study. The perception of assimilated word forms was probed by an identification and a discrimination task. The assimilation rule used in this study was the rule of liquid assimilation that is applied in Hungarian (see Olsson, 1992, p.57; Siptár & Törkenczy, 2000, p. 182): the Hungarian word for ‘left’ [bɔl] may be pronounced [bɔr] if the next phoneme is also an /r/ (as in [bɔlro:l], Engl. ‘from the left’). However, this change from /l/ to /r/ is not allowed to happen in [bɔlna:l] (Engl. ‘at the left’). Based on previous research (Gaskell & Marslen-Wilson, 1998; Gow, 2002a; chapter 3), we expected that the /l/-to-/r/ change would be less noticeable in a context in which the /l/-to-/r/ change is a viable assimilation according to the Hungarian liquid assimilation rule. This was indeed the case. In an identification and a discrimination task, the /l/-to-/r/ change was more noticeable in case of an unviable change ([bɔlna:l] → *[bɔrna:l]) than in case of a viable change ([bɔlro:l] → [bɔrro:l]). This pattern of results remained unchanged qualitatively and quantitatively when the change was applied to the Hungarian nonword [zɔl]. This indicates that the lexical status of the assimilated form does not influence the compensation-for-assimilation process.

In order to test the impact of experience with the liquid assimilation rule, we (chapter six) tested the perception of the Hungarian words by Dutch listeners. The Dutch language does not know an /l/-to-/r/ assimilation rule (cf. Booij, 1995). This absence of experience with the assimilation rule, however, only had an impact on the perception of the assimilated utterances in the identification task, and no differences at all were evident in the results of the discrimination task. Overall, the commonalities between Dutch and Hungarian listeners outweighed the differences. Participants, irrespective of language experience, experienced the two forms [bɔl] and [bɔr] as more similar, if these sounds were followed by [ro:l] than if they were followed by [na:l]. This indicates that experience with an assimilation rule is not pivotal in order to compensate perceptually for assimilation.

In this study, we attempt to broaden the empirical scope of these assertions made previously (chapter six) by using a different method to study the perception of assimilated word forms. This is achieved by probing the perception of assimilated word forms by using event-related potentials (ERPs). In the previous behavioral study, we examined the perception of assimilated utterances during attentive listening. We now aim to contrast this by examining the pre-attentive processing of these stimuli by presenting the stimuli in an oddball series while measuring the electro-encephalogram. This allows us to probe the perception of assimilated word forms without asking participants to make decisions on what they hear; a process which may be vulnerable to the familiarity with the stimulus material and may contaminate the comparison of Hungarian and Dutch listeners. In this paradigm, we can make use of a “powerful tool of cognitive neuroscience” (Näätänen, 1995, p. 6) called the mismatch negativity (MMN). The MMN is a component of the event-related potential (ERP) that is elicited when participants are exposed to a train of auditory stimuli consisting of—at least—two types of stimuli. If one stimulus accounts for the majority of the stimuli in the series and an ‘oddball’ accounts for the minority of the stimuli, then the ERP to the deviant stimulus is usually more negative than the ERP to the standard stimulus. This difference occurs at 100-300 ms after the onset of the mismatch in the acoustic stimulus with a maximum at the frontal and central leads. This negativity, called MMN, seems to be sensitive to the perceived distance rather than to the acoustic difference between the standard and the deviant (see, e.g., Näätänen & Winkler, 1999). This means that the size of the MMN does not solely depend on the acoustic difference between the standard and the deviant stimulus, but is also influenced by the weight listeners assign to this acoustic difference (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen 1993). In addition, the MMN reflects whether the difference between the standard and deviant is phonemic in the native language of the listener (Näätänen et al., 1997; Phillips et al., 2000). Sharma and Dorman (1999), for instance, showed that the MMN is larger for a standard-deviant pair from different phonological categories than for a MMN to a within-category contrast.

Therefore, if listeners compensate automatically for Hungarian liquid assimilation, we may expect the MMN to a deviant [bɔ̃rro:l] contrasting with a standard [bɔlro:l] to be smaller than the MMN to a deviant *[bɔ̃rna:l] contrasting with a standard [bɔlna:l]. This leads to the basic two-factorial design used in the present study: An oddball series is presented in which the standard is a word in its canonical form ([bɔl]) and the deviant is a changed form ([bɔ̃r]). As a second factor, the context in which these forms are presented either allows the assimilation ([...ro:l]) or does not allow the assimilation ([...na:l]). An automatic compensation is indexed by an interaction of the MMN evoked by the oddball series with the context in which the deviant is

presented. The MMN in the viable-context condition should be smaller than the MMN in the unviable-context condition.

In order to evaluate a possible impact of first, the lexical status of the assimilated forms, and second, the specific language experience on compensation for assimilation, we used this basic design with words and nonwords; and with listeners with and without experience with the Hungarian rule of liquid assimilation. By crossing these two independent variables, we get for conditions: First, Hungarian listeners listened to Hungarian words; second, Hungarian listeners listened to Hungarian nonwords, third, Dutch listeners listened to Hungarian words, hence Dutch nonwords, and finally, Dutch listeners listened to Dutch words. If there is a lexical influence on compensation for assimilation, the MMN to words should be more context-dependent than the MMN for nonwords. If language experience is necessary for compensation for assimilation, the MMN should only be context-dependent for Hungarian listeners. Finally, if both language experience and the lexical status of the assimilated form play a role, we should find an interaction of the MMN with the context in Dutch listeners, a minor effect in Hungarian listeners listening to Hungarian nonwords, and a major effect in Hungarian listeners listening to Hungarian words.

Experiment 1: Hungarian listeners

In this experiment, we evaluated whether the MMN to phonological changes is context-dependent when Hungarian words and nonwords are presented to Hungarian listeners. To this end, we measured the ERPs to auditory stimuli in an oddball series in which the standard stimulus ($p = 0.85$) always ended in an apical lateral /l/ and the deviant ended in an apical trill /r/. This change was applied to the Hungarian word [bɔl] (Engl. 'left') and the Hungarian nonword [zɔl]. In addition, the stimulus words were followed by the case suffixes [ro:l] (Delative, similar to 'from') and [na:l] (Adessive, similar to 'at'). According to the rule of Hungarian liquid assimilation (see Olsson, 1992, p.57; Siptár & Törkenczy, 2000, p. 182), the change from /l/ to /r/ is a possible assimilation before the Delative suffix [ro:l] but not before the Adessive suffix [na:l]. Thus, based on the assumption that there is a context-sensitive compensation mechanism we expect the MMN to the pair [bɔlro:l]-[bɔrrɔ:l] to be smaller than the MMN to the pair [bɔlna:l]-*[bɔrna:l].

Method

Participants

Twelve students (six male, six female) from different Budapest universities participated in the study. They were paid for participation. Participants had an age of 16 to 28 years (mean: 20.5 years); ten participants were right-handed, two were left-handed. All participants gave their informed consent, completed a handedness questionnaire, and were screened for hearing problems. They were all right-handed native speakers of Hungarian.

Materials

A female native speaker of Hungarian was recorded uttering multiple tokens of the canonical form [bɒlna:l] (Engl. 'at the left') and an unviable variant *[bɒrna:l] and the canonical [bɒlro:l] form (Engl. 'from the left') and a viable alternative [bɒrro:l]. The Hungarian context suffixes of the cases called Delative (from where?) and Adessive (where at?) were chosen, because they are phonetically quite similar. Both start and end with a sonorant and contain a long vowel. Any differences that these stimuli may cause in the perception of the preceding segments can thus not be attributed to gross acoustic differences (overall amplitude, presence of voicing) between the context sounds.

Sample frequency for recording the natural utterances was 22050 Hz. Recordings were filtered from 130 to 8000 Hz. (f_0 was 150 Hz.) The first syllable from one [bɒlna:l] utterance was spliced off and edited with the software package PRAAT 4.0 (Boersma & Weenink, 2002). This syllable [bɒl] was then edited in order to create a speech sound in which the only difference is the manner of articulation of the final phoneme of the first syllable.

Given the similar place of articulation for the apical trill and the apical lateral, the primary cue for the lateral/trill distinction is the presence of amplitude modulation (AM) in the trill (Ladefoged & Maddieson, 1996). In addition, the resonant characteristics of the lateral and the open period of the trill are different. In order to manipulate both cues, the original stimulus was divided in a source and a filter by a LPC (linear predictor coefficients) analysis (16th order). For the edited sound stimuli, one cycle of AM (20 Hz) with 12 dB AM depth was added to the estimated source. This was done by editing the intensity function of the source as estimated by PRAAT. Formant estimations showed a slightly (200 Hz) lower third formant for the resonant period of the trill. Therefore, the third formant was also lowered by 0.5 barks. Filtering the edited source with the edited filter yielded a stimulus that was perceived as an apical trill in the behavioral study (chapter six). In order to convey a casual speaking style, pre-voicing was shortened from 40 ms to 27 ms by cutting two complete cycles of pre-voicing in all stimuli starting with a voiced labial stop. These stimuli had a length of 544 ms.

In order to create the Hungarian nonwords, the filter was manipulated for the initial 105 ms of the utterance in order to emulate the estimated filter function in natural utterances of an alveolar voiced fricative of the speaker. The filter settings were stable for the first 75 ms. From 75 to 105 ms, the filter settings were interpolated from the /z/ settings to the original /ɔ/ settings using the cosine function in $[0, \pi]$. This yields a slow initiation of the formant movement, a maximal acceleration at the midpoint, which then slows down again toward the endpoint. In addition, the original voice source was attenuated to one half of the original amplitude and noise was added for the first 75 ms. From 75 ms to 105 ms the noise was faded out linearly, and the original voiced source was restored to the original level using linear interpolation. The apical lateral-apical trill continuum was identical to the Hungarian word continuum. These [zɔl]-[zɔr] stimuli were twelve ms longer than the [bɒl]-[bɒr] stimuli, reflecting the fact that fricatives are intrinsically longer than stops.

The context case suffixes [na:l] (Engl. 'at' in answer to the question 'where?' as in 'at your left') and [ro:l] (Engl. 'from') were spliced from other utterances, equalized in overall energy and concatenated with the Hungarian word and nonwords stimuli. The amplitude relation of word and nonword to the case suffixes were edited so that they emulated the amplitude relation of the first (always stressed) syllable to the second syllable in the natural utterances. Figure 7.1A and 7.1B show the four stimuli that resulted from this editing procedure. Figure 7.1A shows the stimulus [bɒlna:l] in which the first syllable has no amplitude modulation and the second syllable is the case suffix [bɒrna:l] that does not allow the assimilation /l/ to /r/. Figure 7.1B shows the stimulus [zɔrro:l] in which the original /b/ onset was manipulated as described above. These stimuli were presented to the participants over headphones using the Presentation software (Neurobehavioral Systems).

Procedure

After mounting of the electrodes, participants were seated in a comfortable chair and watched a silent video. At the same time, participants heard an oddball series over headphones with one stimulus every 1.2s. The recordings were spread out over two recording sessions on different days of about one hour each. Within one recording session, there were four blocks with 805 stimuli each. Each block started with five standard stimuli that were not used for the ERP averaging. In addition, every deviant was followed by at least one standard stimulus.

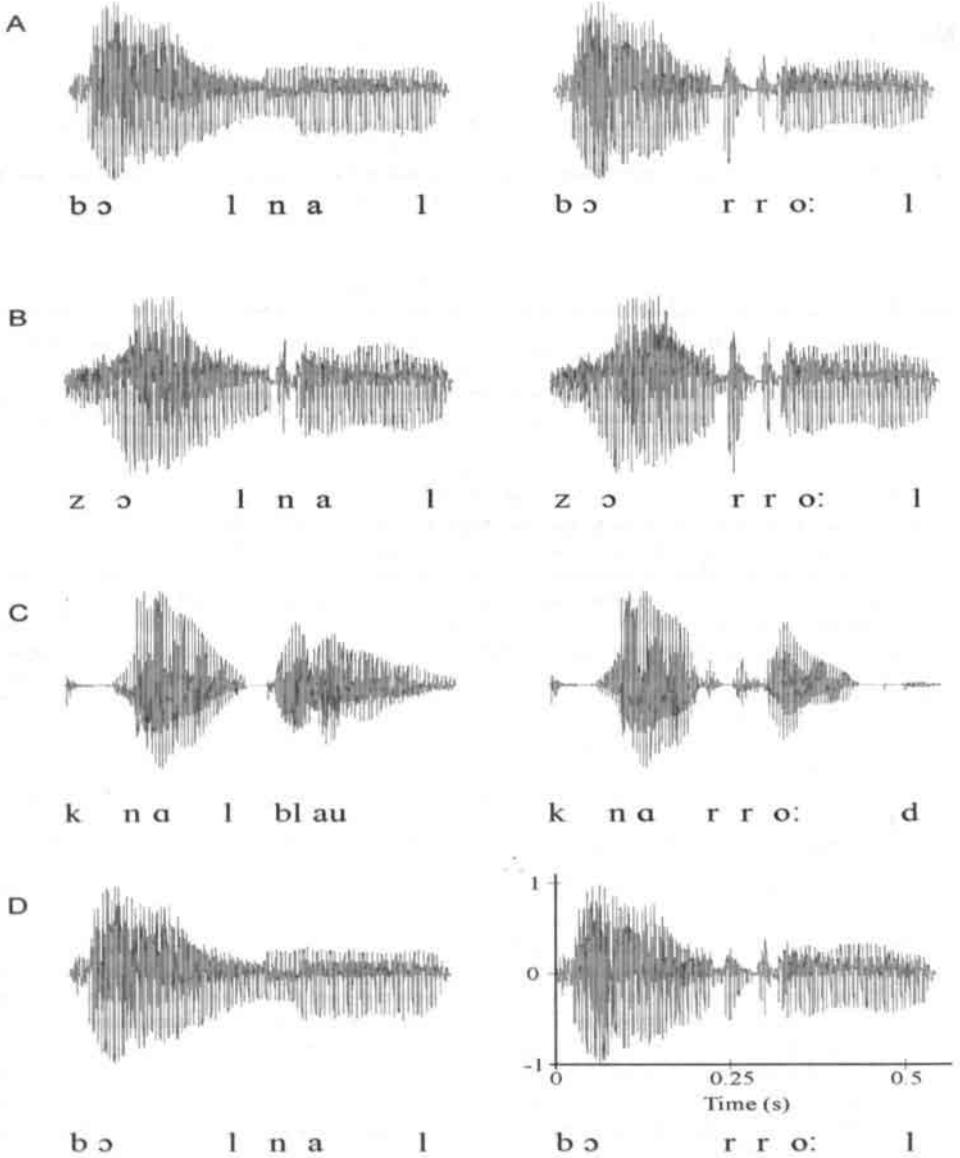


Figure 7.1 Stimuli used in Experiments 1 to 3. The stimuli on the left side are stimuli in their canonical form in a context that does not allow the phonological change. The stimuli on the right side are in the non-canonical form with a morpheme-final /r/ in context that allow the assimilation according to Hungarian phonology. Panel A shows the Hungarian word (and Dutch nonword) stimuli, Panel B shows the stimuli for the Hungarian nonword stimuli, and Panel C shows the Dutch word stimuli. Panel D shows another version of the Hungarian word stimuli, which were used in Experiment 3.

Within one recording session, two different oddball series were presented in two blocks each. A block comprised 805 stimuli with five standards at the start of the block. In one series, the standard ($p = 0.85$) was [(b/z)ɔlro:] and the deviant was the viable alternative pronunciation [(b/z)ɔrro:]. In the second series, the standard was [(b/z)ɔlna:] ($p = 0.85$) and the deviant was the unviable alternative pronunciation *[(b/z)ɔrna:]. (Following the linguistic formalism, we used the *star to indicate forbidden forms.) In half of the recording sessions, the blocks with the viable alternative as deviant were the first and third block and the blocks with the unviable alternative were the second and fourth block. In the other half of the recording sessions, this order was reversed. In addition, half of the participants heard the words ([bɔ...]) in the first recording session and the nonwords ([zɔ...]) in the second session; for the other half of participants this order was reversed.

Electrophysiological recording and data reduction

The ERPs were recorded over 20 (Fz, Cz, Pz, Oz, F3, F4, F7, F8, C3, C4, T3, T4, T5, T6, P3, P4, O1, O2, M1, M2) electrodes linked to the nose as reference. Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supra-orbital ridge of the right eye. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. Using a 32 channel SynAmp amplifier and the SCAN program of the Neuroscan software package (Neurosoft Inc.) the brain electric activity was acquired in continuous mode. The sampling rate was set to 250 Hz, with a band pass of 0.1–200 Hz. The acquired samples were sliced into epochs from 52 ms before stimulus onset to 1000 ms after stimulus onset, band-pass filtered from 1 to 20 Hz, baseline-corrected from -50 to the point at which standard and deviant started to differ (220 ms and 233 in the word and nonword, respectively). Artefacts from vertical eye-movements were reduced using linear regression. After de-correlation, samples were rejected if the voltage on any channel excluding the vertical eye channel exceeded a value of $75 \mu\text{V}$. Epochs were then averaged for each stimulus type. The individual ERP averages were obtained from 240 epochs in each condition by using all deviant stimuli and the same number of standard stimuli randomly drawn from all standard stimuli, excluding standards directly following deviants.

For data analysis, a mean-amplitude measure from Fz was used. Mean amplitudes were determined from a 100 ms window with the peak of the grand-average MMN as anchor. This grand-average peak was determined by visual inspection of Fz and mastoid electrodes.

Design

The design entails two independent variables. The first variable will be coined Change, with the levels 'unchanged' and 'changed.' The second variable will be coined Context, with the levels 'viable' and 'unviable.' This design is applied to both word- and nonword-responses. The independent variable is the mean amplitude on Fz in a 100 ms window around the grand-average peak of the MMN.

Results

Nonword condition

The upper panel of Figure 7.2 shows the grand average waveforms from Fz, M1, and M2 for all four conditions of the nonword experiment. There is a well-defined MMN that reverses polarity at the mastoid electrodes (see also Figure 7.5 for the difference waves). The MMN peak is at 420 ms post stimulus onset, which is approximately 180 ms after onset of the acoustic mismatch between standard and deviant. (Note that the acoustic mismatch is introduced gradually.)

The mean amplitudes in a window 370–470 ms were $-1.53 \mu\text{V}$ for the [zɔlna:] stimulus, $-2.35 \mu\text{V}$ for the *[zɔrna:] stimulus, $-1.59 \mu\text{V}$ for the [zɔlro:] stimulus and $-2.28 \mu\text{V}$ for the [zɔrro:] stimulus. A two factorial-repeated measurement Analysis of Variance (ANOVA) revealed a significant effect of Change ($F(1, 11) = 17.85, p < 0.005$). However, neither the effect of Context nor the interaction between Context and Change ($F_s < 1$) were significant.

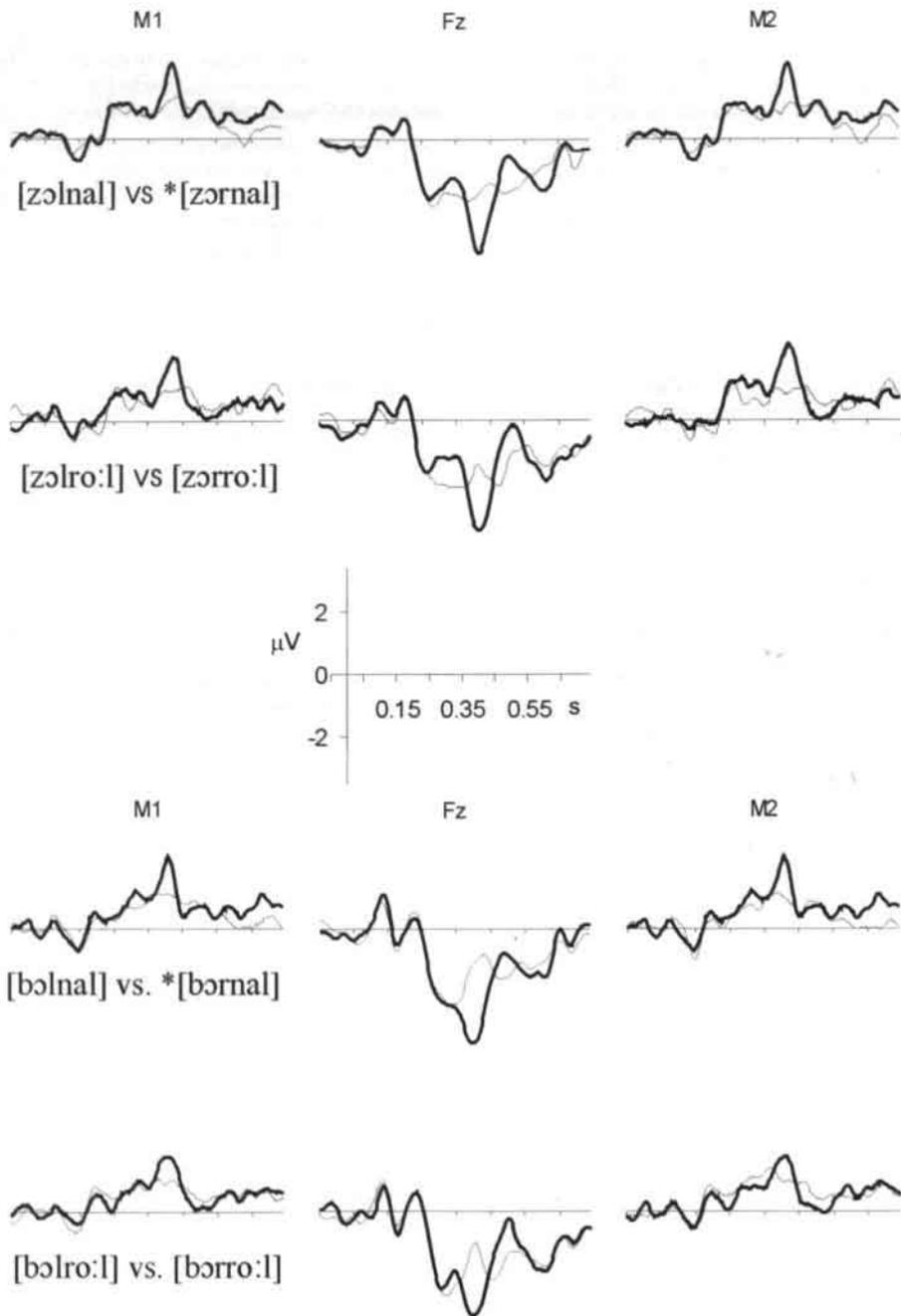


Figure 7.2 Grand averages for the electrodes Fz, M1, M2 from Hungarian listeners listening to Hungarian words (lower half) and nonwords (upper half). The standard ERP is depicted by the thin line, the deviant ERP by the thicker line. The standard-deviant pair is indicated below the panel for the ERPs at M1. Phonologically unviable deviants are indicated by a *star.

Word condition

The lower panel of Figure 7.2 shows the grand average waveforms from Fz, M1, and M2 for all four word conditions. There is a well-defined MMN that reverses polarity at the mastoid electrodes. The MMN peak is at 404 ms post stimulus onset, which is again approximately 180 ms after onset of the acoustic mismatch between standard and deviant (see also Figure 7.5 for the difference waves). Due to the slightly longer onset of the nonword stimuli, the acoustic mismatch occurs earlier in these word stimuli (see Figure 7.1 B). The mean amplitudes in a window of 354-454 ms are $-1.16 \mu\text{V}$ for the [bɔlna:l] stimulus, $-2.58 \mu\text{V}$ for the *[bɔrna:l] stimulus, $-1.46 \mu\text{V}$ for the [bɔlrɔ:l] stimulus and $-2.33 \mu\text{V}$ for the [bɔrrɔ:l] stimulus. A two factorial-repeated measurement ANOVA revealed a significant effect of Change ($F(1, 11) = 88.97, p < 0.001$). However, neither the effect of Context ($F < 1$) nor the interaction between Context and Change ($F(1, 11) = 2.97, p > 0.1$) were significant.

Discussion

The data show a consistent MMN that has a peak around 180 ms after the gradual onset of the acoustic mismatch between standard and deviant. This negative component reverses polarity at the mastoid electrodes, which fosters its interpretation as a MMN. However, there is no context sensitivity of the MMN. This contradicts our initial hypothesis. We expected the MMN to the [bɔlrɔ:l]-[bɔrrɔ:l] pair to be smaller than the MMN to the [bɔlna:l]-*[bɔrna:l] pair. However, the two MMNs are of similar size. This result is quite surprising. Note that we (chapter six) found that the same [bɔrrɔ:l] stimulus was quite often identified as [bɔlrɔ:l], while *[bɔrna:l] was seldom identified as [bɔlna:l]. In addition, discrimination was more difficult in the viable-context condition than in the unviable-context condition. The current failure to find a context effect on the MMN may indicate that these effects may arise at a decision-making level rather than on a perceptual level.

Before we elaborate on the interpretation of this result, we first investigate whether these results are moderated by native-language experience with the assimilation rule. Therefore, we turn to Dutch listeners. In our behavioral measurements (chapter six), Dutch listeners responded in a quite similar way as Hungarian listeners when identifying and discriminating phrases with Hungarian liquid assimilation. We now examine whether this similarity also holds for the automatically evoked brain potentials.

Experiment 2: Dutch listeners.

In this experiment, we probe the pre-attentive perception of the phonologically changed Hungarian utterances by Dutch listeners. The same utterances as in Experiment 1 were used. None of the utterances is meaningful in Dutch. In addition, a Dutch word pair was used to which Hungarian liquid assimilation was applied. In one word, [knaɫroːd] (Engl. 'vivid red'), the assimilation to [knaɫroːd] is viable in Hungarian. Note, however, that neither this rule nor a similar rule exists in Dutch (cf. Booij, 1995). Therefore, ^(*)[knaɫroːd] should actually be indicated with a star, as this form is not allowed in Dutch. In another word, [knaɫblau] (Engl. 'vivid blue') the assimilation to ^{*}[knaɫblau] is not only forbidden in Dutch, but also in Hungarian. These stimuli were used in oddball series in which the canonical word is the standard and the (in Hungarian) viable or unviable alternative is the deviant.

Method

Participants

Twelve students (eleven female, one male) from the University of Maastricht participated in the nonword part of the experiment. They were paid for participation. Participants had an age of 18 to 25 years (mean: 20.3 years); ten participants were right-handed, two were left-handed.

In addition, 15 participants participated in the word-part of this study for a course credit. Participants had an age of 18 to 25 years (mean: 20.3 years); participants were right-handed, two were left-handed.

All participants were native-speakers of Dutch and free of hearing problems. All of the participants spoke English reasonably well, and all but one participant spoke at least a second foreign language (either German or French). One participant spoke some Polish; however, none of the Dutch listeners knew any Hungarian. All participants gave their informed consent.

Materials

The same tokens as in the Hungarian study were used for the Dutch nonword experiment (Hungarian words and nonwords). For the word-experiment, we recorded a speaker of Dutch, who uses an apical trill for the phoneme /r/. (The majority implementation for a trill in Dutch is the uvular trill, however apical trills are also observed, as are uvular approximants). The speaker uttered the words [knaɫroːd] (Engl. 'vivid red') and [knaɫblau] (Engl. 'vivid blue') several times. One of the tokens of [knaɫ] spliced from [knaɫblau] was edited with the same methods as the Hungarian utterance in order to produce an utterance of [knaɫ], which was identical to the original [knaɫ] utterance in every respect but the last segment. The edited word started to differ from the original at 200 ms. The context words [blau] and [roːd] were spliced from two other utterances and concatenated with the original and the edited word. This procedure yielded an unchanged stimulus with a viable context [knaɫroːd], an unchanged stimulus with an unviable context [knaɫblau], a changed stimulus with an—according to the Hungarian rules—viable context [knaɫroːd], and a changed stimulus with an unviable context ^{*}[knaɫblau]. These stimuli were presented to participants over four loudspeakers (JBL Control 25) placed in the corners of the sound-attenuated EEG cabin.

Procedure

For the participants in the nonword condition of this experiment, the procedure was the same as for the Hungarian participants in Experiment 1 of this study. These participants were exposed to both the Hungarian word and nonword. The Hungarian words and the Hungarian nonwords do not resemble sensible Dutch utterances. The participants in the word condition of this experiment completed one recording sessions with 4 blocks consisting of 805 stimuli each. Eight participants heard the oddball series with the viable deviant in the first and third block and

the series with the unviable deviant in the second and fourth block. For the remaining seven participants, the reverse order was used.

Electrophysiological recording and data reduction

The nose-referenced electroencephalogram (EEG) (0.1-125 Hz, sampling rate 250 Hz) was recorded with a 32-channel electrode cap covering frontal, central, temporal, and parietal scalp areas. An electrode on the nose served as a reference. Blinks and vertical eye movements were monitored with electrodes placed at the sub- and supra-orbital ridge of the right eye. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. All electrode impedances (EEG and EOG) were kept below 5 kOhm. ERPs were obtained by averaging the EEG time-locked to the onset of the sound. The ERPs were band-pass filtered digitally from 1 to 30 Hz. The signal was normalized using a baseline that was calculated using the mean amplitude from 50 ms before stimulus onset to the onset of the mismatch (200 ms in the Dutch word, 220 ms and 233 in the Hungarian word and nonword respectively). Artefacts from vertical eye movements and blinks were reduced by linear regression. If after de-correlation, epochs still exceeded $|75| \mu\text{V}$ on any of the channels (including the vertical eye channels), they were rejected. The individual ERP averages were obtained from 240 raw epochs in each condition by using all deviant stimuli and the same number of standard stimuli randomly drawn from all standard stimuli, excluding standards directly following deviants. Mean amplitudes were determined in the same way as in Experiment 1.

Design

The design entails two independent variables. The first variable will be coined Change, with the levels 'unchanged' and 'changed.' The second variable will be coined Context, with the levels 'viable' and 'unviable.' This design is applied to both word- and nonword-responses. The independent variable is the mean amplitude on Fz in a 100 ms window around the grand-average peak of the MMN.

Results

Nonword condition

The upper panel of Figure 7.3 shows the grand average waveforms from Fz, M1, and M2 for all four conditions of the nonword experiment. As stated above, we measured ERPs to Hungarian word and nonwords in Dutch listeners. Here, we only present the data from the Hungarian words.² The ERPs reveal a well-defined MMN that reverses polarity at the mastoid electrodes (see also Figure 7.5 for the difference waves). The MMN peak is at 400 ms post-stimulus onset. This is approximately 175 ms after onset of the deviance, and is quite similar to the 180 ms observed in the Hungarian sample.

The mean amplitudes in the window 350-450 ms were $-1.78 \mu\text{V}$ for the unchanged stimuli and $-2.99 \mu\text{V}$ for the changed stimulus in the viable context. In the unviable context, the mean amplitudes were $-2.02 \mu\text{V}$ for the unchanged stimulus and $-2.78 \mu\text{V}$ for the changed stimulus. A two factorial-repeated measurement Analysis of Variance (ANOVA) revealed a significant effect of Change ($F(1, 11) = 18.80, p < 0.005$). However, neither the effect of Context ($F < 1$) nor the interaction between Context and Change ($F(1, 11) = 1.96, p > 0.15$) were significant.

² Though both Hungarian words and nonwords were nonsensical to Dutch listeners, we did not pool the data. The onset of the acoustic mismatch occurred at different time points in the Hungarian word and nonword stimuli. Consequently, pooling the data would have introduced temporal smearing of the ERPs.

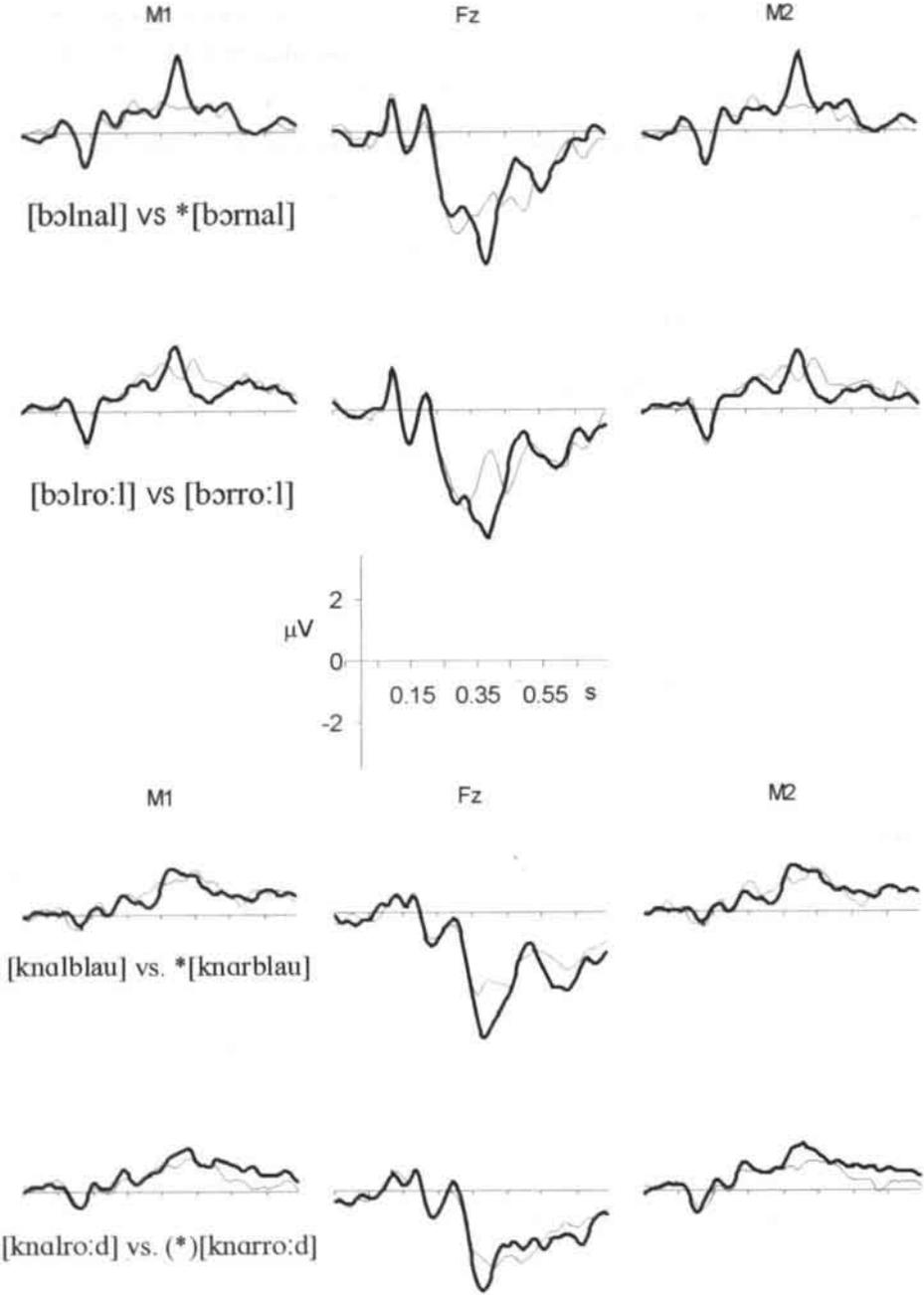


Figure 7.3 Grand averages for the electrodes Fz, M1, M2 from Dutch listeners listening to Hungarian words, being Dutch nonwords (upper half) and Dutch words (lower half). See also caption of Figure 7.2.

Word condition

The lower panel of Figure 7.3 shows the grand average waveforms from Fz, M1, and M2 for all four Dutch-word conditions. The MMN was more pronounced in the unviable-context condition and peaked at 388 ms, which is about 180ms after the onset of the change (see also Figure 7.5 for the difference waves). The mean amplitudes in the window from 338 to 438 ms were, in the viable context, $-1.80 \mu\text{V}$ for the canonical form and $-2.12 \mu\text{V}$ for the changed form. In the unviable context, the mean amplitudes were $-1.95 \mu\text{V}$ for the canonical stimulus and $-3.02 \mu\text{V}$ for the changed stimulus.

A two factorial-repeated measurement Analysis of Variance (ANOVA) revealed a significant effect of Change ($F(1, 14) = 20.27, p < 0.001$) and a significant effect of Context ($F(1, 14) = 8.71, p < 0.025$). However, both main effects are compromised by a significant interaction between Context and Change ($F(1, 14) = 7.86, p < 0.025$). This interaction is due to the fact that the effect of Change (hence, the MMN) is smaller in the viable context ($-0.31 \mu\text{V}$) than in the unviable context ($-1.07 \mu\text{V}$). Subsequent *t*-tests showed that the effect of Change was significant in the unviable context ($t(14) = 4.40, p < 0.005$), but constituted only a trend in the viable context ($t(14) = 2.02, p < 0.1$).

Discussion

In this experiment, the MMN was context sensitive. This expected pattern of this context sensitivity was, however, only observed in the word-condition. This is difficult to interpret in terms of the hypothesis reviewed in the introduction. We expected that there might be a lexical influence for both Hungarian and Dutch listeners, or that the lexical influence might be stronger for the Hungarian listeners. However, it is difficult to accept that the lexical status of the assimilated form should only matter for the listener group that is not familiar with the assimilation rule. It is possible, however, that the difference between the word and nonword condition in Dutch listeners does not arise as a consequence of the lexical status of the assimilated form, but rather represents an effect of particular acoustic differences. Figure 7.1 shows the stimuli used in this study. A detailed comparison of the deviant stimuli, the right panels of Figure 7.1A, 1B, and 1C, reveals that the peak of the AM in the deviants—the first peak—was larger than the AM peak in the viable-context stimulus—the second AM peak—in all cases but the Dutch word. Therefore, the alleged lexical influence on compensation for assimilation is obscured by the confounding of lexical status with the relation of AM peaks in deviant and context sounds. If this acoustic confound is the crucial factor, a similar effect should arise when an altered Hungarian word is presented to Dutch listeners. This was tested in the next experiment.

Experiment 3a

In the previous experiment, a context-sensitive MMN was only observed when Dutch listeners listened to Dutch words. In this experiment, we tested whether such results can also be obtained when altered versions of the Hungarian words used in Experiments 1 and 2 were presented to Dutch listeners. To this end, we used the Hungarian word [bɔl] in its original form as standard. A less drastically different form with a lower AM peak (compare Figure 7.1A with Figure 7.1D) was used as deviant. This stimulus is similar to the stimulus with an AM depth of 9 dB, which was perceived predominantly as [bɔr] in our behavioral study (see chapter six). As in the previous experiments, these stimuli were presented in a context that allows the assimilation [...ro:l] and one that does not allow the assimilation [...na:l].

Method

Participants

Ten students (eight female, two male) from the University of Maastricht participated in the experiment. They were paid for participation. Participants had an age of 19 to 26 years (mean: 21.0 years); nine participants were right-handed, one was left-handed. All participants were native-speakers of Dutch and free of hearing problems. All of the participants spoke English reasonably well, and all but two participants spoke at least a second foreign language (either German or French). No other languages were spoken by any participant. All participants gave their informed consent.

Materials

The same standard stimuli as in Experiment 2 in the nonword condition were used. New deviant stimuli were created by choosing an AM depth that lead to similar maximal amplitudes of the AM in the target and the AM in the context stimulus (see Figure 7.1D).

Procedure, electrophysiological recording, and data reduction

Procedure, electrophysiological recording, and data reduction were the same as in Experiment 2.

Design

The design entails two independent variables. The first variable will be coined Change, with the levels 'unchanged' and 'changed.' The second variable will be coined Context, with the levels 'viable' and 'unviable.' The independent variable is the mean amplitude on Fz in a 100 ms window around the grand-average peak of the MMN.

Results and Discussion

Figure 7.4 shows grand averages for all four conditions. There is a clear peak of the MMN at Fz at 404 ms accompanied by a positive peak at the mastoid electrodes (see also Figure 7.5 for the difference waves). This replicated the latency value of the MMN found in the earlier experiments. The mean amplitudes in the 354-454 ms window at Fz were $-1.67 \mu\text{V}$ for the unchanged stimulus and $-2.41 \mu\text{V}$ for the changed stimulus in the unviable-context condition.

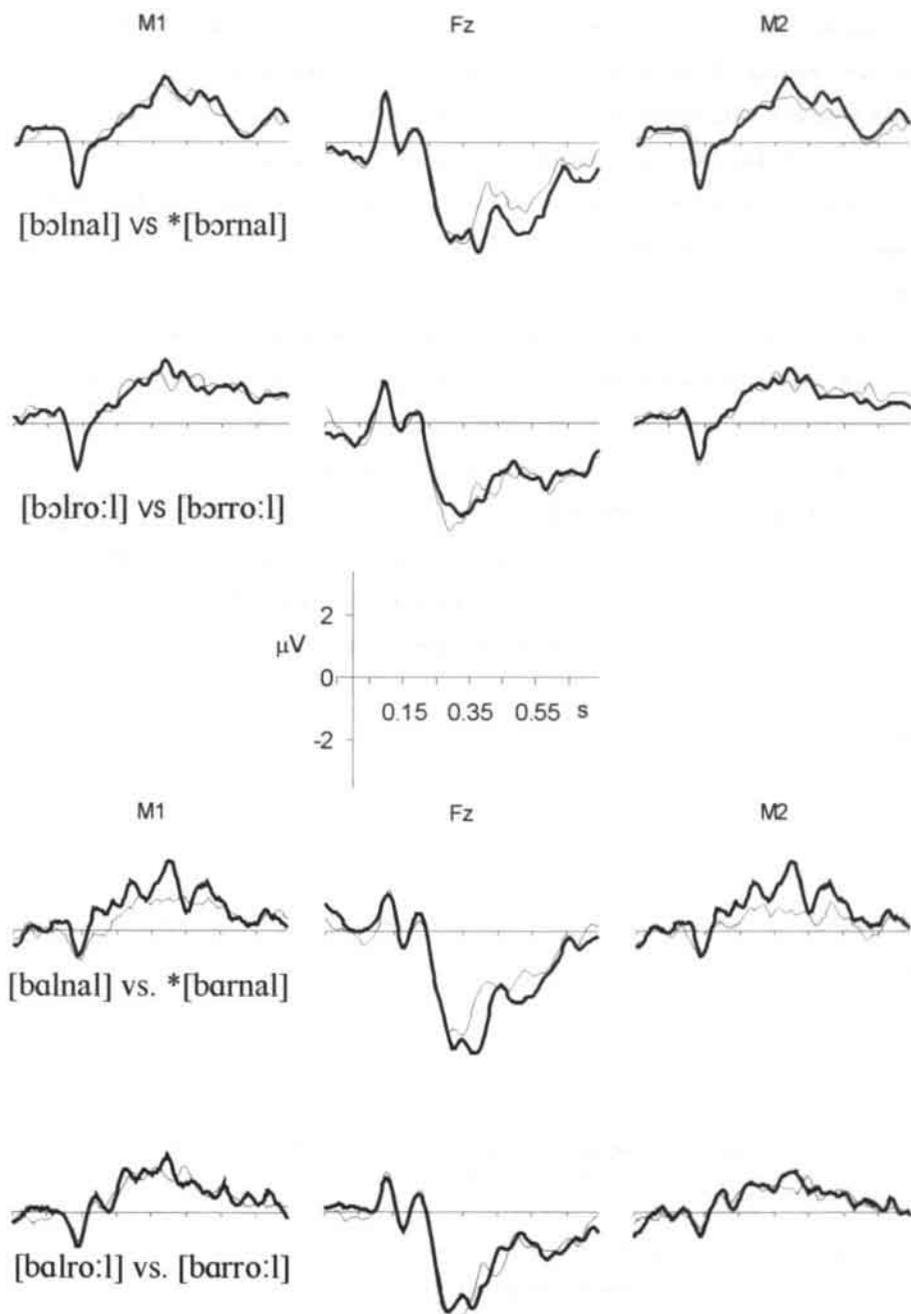


Figure 7.4 Grand averages for the electrodes Fz, M1, M2 from Dutch listeners listening to Hungarian words, being Dutch nonwords (upper half) and Hungarian listeners listening to the same Hungarian words (lower half). See also caption of Figure 7.2.

In the viable context, the mean amplitudes were $-1.87 \mu\text{V}$ for both the changed and the unchanged stimulus. A repeated-measure ANOVA with Change and Context as predictors revealed a trend for the effect of Change ($F(1, 9) = 4.06, p < 0.1$) and no significant effect of Context ($F < 1$). However, the interaction between both predictors was significant ($F(1, 9) = 10.96, p < 0.01$). Consecutive t-tests showed that the effect of change—and thus the MMN—was significant for the unviable-context condition ($t(9) = 2.72, p < 0.025$) but not for the viable-context condition ($t < 1$).

This result clarifies the reason for the difference in results between the word and the nonword condition of Experiment 2. This difference was not due to the lexical status of the assimilated form, but rather to stimulus differences. That is, a context-sensitive MMN can be found when Hungarian words, Dutch nonwords, are presented to Dutch listeners. However this only occurs, if the peak of the AM of the deviant stimulus is not larger than the peak of the AM in the context (See Figure 7.1). The question now arises whether this acoustic detail would also explain the absence of context sensitivity in the Hungarian sample in Experiment 1. To this end, we re-ran Experiment 3, but now with Hungarian participants.

Experiment 3b

In the first two experiments, the results of Hungarian and Dutch listeners were similar if *identical* stimuli were used in both samples (cf. the word condition of Experiment 1 with the nonword condition of Experiment 2). In this experiment, we tested whether this also holds for the Hungarian stimuli which induced a context-sensitive MMN when presented to Dutch listeners.

Method

Participants

10 students (all female) from different Budapest Universities participated in the experiment. They were paid for participation. Participants had an age of 18 to 29 years (mean: 22.4 years); All participants gave their informed consent, completed a handedness questionnaire, and were screened for hearing problems. They were all native speakers of Hungarian.

Materials and Procedure

The same stimuli and Procedure as in Experiment 3a were used.

Electrophysiological recording and data reduction

Electrophysiological recording and data reduction were the same as in Experiment 1.

Design

The design entails two independent variables. The first variable will be coined Change, with the levels 'unchanged' and 'changed.' The second variable will be coined Context, with the levels 'viable' and 'unviable.' The independent variable is the mean amplitude on Fz in a 100 ms window around the grand-average peak of the MMN.

Results and Discussion

Figure 7.4 shows the grand averages for all four conditions. There is a clear peak of the MMN at Fz with a peak at 396 ms accompanied by a positive peak at the mastoid electrodes (see also Figure 7.5 for the difference waves). This replicated the latency value of the MMN found in the earlier experiments. The mean amplitudes in the 346–446 ms window at Fz were, in the unviable context, $-1.96 \mu\text{V}$ for the unchanged stimulus and $-2.86 \mu\text{V}$ for the changed stimulus. In the viable context, the mean amplitudes were $-2.13 \mu\text{V}$ for the changed and $-2.04 \mu\text{V}$ for the unchanged stimulus. A repeated-measure ANOVA with Change and Context as predictors revealed a significant effect of Change ($F(1, 9) = 10.52, p < 0.05$) and a trend for the effect of Context ($F(1, 9) = 4.30, p < 0.1$). However, the interaction between both predictors was significant ($F(1, 9) = 12.38, p < 0.01$). Consecutive t-tests showed that the effect of change—and thus the MMN—was significant for the unviable-context condition ($t(9) = 3.80, p < 0.01$) but not for the viable-context condition ($t < 1$).

This replicated the results of Experiment 3a. It also indicated that the MMN to phonological changes seems to be independent of native-language experience with the given changes. It seems that, if identical stimuli were presented to both Hungarian and Dutch listeners, the results were equivalent.

General Discussion

In the present study, we investigated whether the process of compensation for assimilation, first, is subject to a lexical influence, and, second, depends on language experience with a given assimilation rule. Utterances in which Hungarian liquid assimilation was applied either viably or unviably to words and nonwords were presented to Dutch and Hungarian listeners. Using a passive-listening task with an oddball series, we expected the MMN to be smaller for a standard-deviant pair consisting of a canonical form (e.g. [bɔlro:l]) and a viable alternative pronunciation (e.g. [bɔrro:l]) than the MMN for a standard-deviant pair consisting of a canonical form (e.g. [bɔlna:l]) and an unviable alternative pronunciation (e.g. *[bɔrna:l]).

However, in the initial two experiments, this was only the case for Dutch listeners listening to Dutch words. Detailed examination of the stimulus material showed that this standard-deviant pair differed acoustically from the other stimuli. In particular, the AM that introduced the acoustic mismatch between standard and deviant was larger than the AM in the context sound in all cases but the Dutch words presented to Dutch listeners.

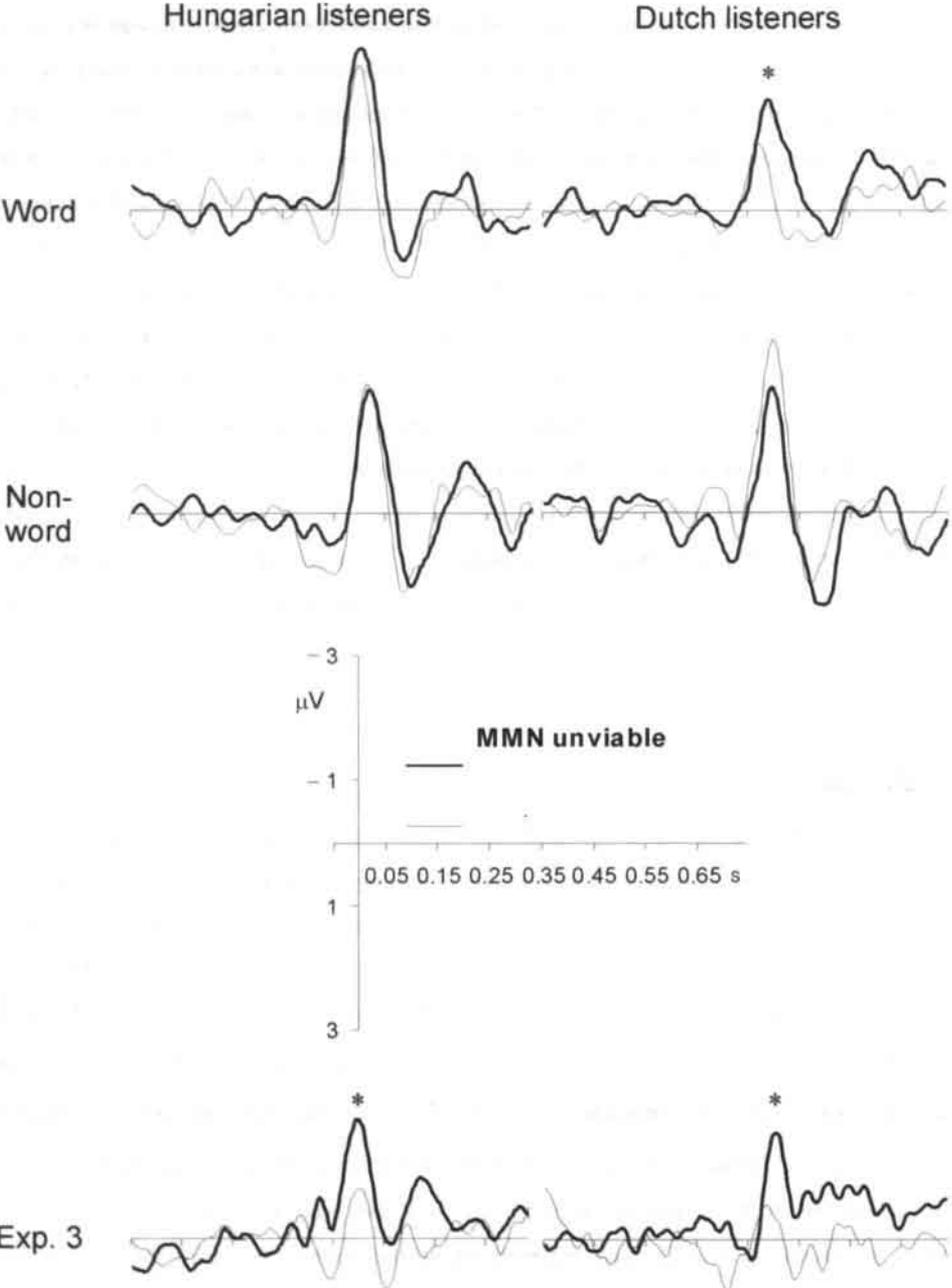


Figure 7.5 Difference waves (deviant – standard) at Fz for all experiments. The two upper panels show the difference wave for Experiment 1 and 2. The lower panel shows the data from Experiment 3. A star over the peak of a MMN indicates a significant difference between the difference waves displayed.

A final experiment showed that this acoustic detail was critical. Both Dutch and Hungarian listeners showed a context-sensitive MMN when the AM peak in the deviant stimulus was of a size similar to the AM peak of the context. Figure 7.5 provides an overview of the results with comparisons of the difference waves over all experiments.

The main impetus of this study was to investigate the impact of the lexical status of the assimilated form and native-language experience with an assimilation rule on the perception of assimilated word forms. In the course of this study, neither of these variables contributed substantially to the results. If *acoustically identical* stimuli were used, Dutch and Hungarian listeners showed comparable results. This was the case in the comparison of the word condition with Hungarian listeners (Experiment 1) with the nonword condition with Dutch listeners (Experiment 2), in which the Hungarian word [bɔl] was presented, with the strong AM in the deviant [bɔr]. This was also the case for the comparison of Experiment 3a with Experiment 3b in which the Hungarian word [bɔl] was presented, with the less strong AM in the deviant [bɔr]. It seems that experience with a given assimilation rule hardly influences the automatic processing of assimilated word forms. This replicated our behavioral results (chapter six). In this study, language experience of the assimilated form did not influence subjects' discrimination performance, and only moderated the context sensitivity when identification was required.

In addition, the lexical status of the assimilated word also did not influence compensation for assimilation (see Experiment 1 and the comparison of the word condition of Experiment 2 with Experiment 3a). Again, this replicated our behavioral results (chapter six), which also failed to show an effect of lexical status of the assimilated form on compensation for assimilation.

However, an acoustic variable—the relation of the amplitude peak of the acoustic mismatch to the amplitude peak of the context AM—was pivotal in controlling context sensitivity. This indicates that acoustic detail plays a critical role in the perceptual compensation for phonological assimilation. Similar results have been reported by Gow (2002a, 2003) and ourselves (chapter five) for other assimilation rules. In these studies, instances of non-alveolar segments are presented in contexts in which they might arise as a consequence of an intended non-alveolar segment or as a consequence of an assimilated alveolar segment. An exemplar of such a segment is the /p/ in 'ripe berries.' The /p/ in 'ripe' might be due to either an intended /p/ or could arise if the speaker intended to say 'right berries,' but assimilated the /t/ in 'right' to a /p/. Gow showed that good instances of non-alveolar segments are not subject to context effects and are perceived as intended non-alveolars, even if the context would allow an alveolar segment to assimilate. Similarly, we (chapter five) showed that a labial nasal is only subject to

context effects—being interpreted as an alveolar nasal— if the nasal is relatively short. A short labial nasal followed by a labial obstruent was often identified as an alveolar nasal, which is a compensation for a possible nasal place assimilation in Dutch. However, a long labial nasal with the same context was nearly never identified as an alveolar. This mirrors the present finding that a stimulus with a strong AM—corresponding to a trill being spoken with some emphasize—is not subject to a context effect while a less strong trill is. Therefore, the data of Gow (2002a), our earlier results (chapter five), and the present results converge on the finding that a good exemplar of a phonological category is not subject to context effects, while a less good exemplar, still bearing some cues for the original, unassimilated segment, is subject to a context effect. This indicates that acoustic/phonetic detail has an important role to play in the perception of assimilated utterances. The current results additionally highlight that the role of phonetic detail is independent of native-language experience with a given assimilation rule. Therefore, it is unlikely that the listener has to be confronted with assimilated forms in order to be able to compensate for the given assimilation.

A surprising aspect of the current data is that the initial results did not converge with our behavioral results (chapter six). Although Experiment 3 showed that the behavioral context effects observed previously (chapter six) were to at least to a substantial degree perceptual in nature, and not solely arising on a decision-making level, it remains puzzling that the effect of acoustic detail was not prominent in the discrimination and identification data reported in chapter six. According to the standard interpretation, the MMN is a “*measure of auditory discrimination accuracy*” (Näätänen, 2001, p. 1). However, in our behavioral data (chapter six), discrimination performance between the same stimuli as used in Experiment 1 was influenced by context and this context effect was not moderated by acoustic details. That is, d' was worse in case of the viable context, and this context effect was the same for less deviant stimuli as used in Experiment 3 as well as for strongly deviant stimuli as used in Experiment 1. This might indicate that d' and the MMN are not two measures of the same construct, auditory discrimination accuracy, or that they measure different aspects of the perceptual accuracy, that is detection and discrimination. However, it is important to note an important methodological difference between the discrimination paradigm used in chapter six and the present experiments. In chapter six, we presented stimuli with differing acoustic detail in the same block. In the present experiment, differences in phonetic detail were dispersed over different experimental blocks and even different experiments. It has long been established that the participant's reaction to a given stimulus does not only depend on the stimulus situation and instructions on a given trial, but also on the range of stimuli presented in the same experimental block (Van der Heijden, Hagenaar, &

Bloem, 1984). It remains to be seen whether such 'mixing costs' (see Los, 1999) may also influence the MMN if different levels of acoustic mismatch are introduced in an oddball series.

Most importantly, however, the current data converge with our behavioral data (chapter six) with regard to the influence of specific-language experience and the lexical influence on compensation for assimilation. Although lexical processing and language experience are pivotal in order to achieve stable speech recognition in one's native language (see, e.g., McQueen, 1998; Näätänen et al., 1997; Norris, McQueen, & Cutler, 1995; Suomi, McQueen, & Cutler, 1997; Winkler et al., 1999), it seems that compensation for assimilation may be achieved by basic perceptual processes, which are independent of higher-level influences and can arise without experience with a given assimilation rule. Any theory of compensation for assimilation will have to account for the fact that compensation happens at least partially at a pre-attentive level of processing and is, at this level, independent of language experience, not subject to a top-down influence of the lexical information, and that acoustic details are taken into account.

Compensation can be Triggered by Non-Speech Sounds

In a series of three experiments, it was investigated whether perceptual compensation for phonological assimilations can be triggered by non-speech sounds. To this end, Hungarian utterances, to which a possible assimilation (/l/ → /r/) was applied, were presented to listeners. Words were concatenated with different non-speech context. Non-speech context sounds were designed to resemble speech contexts that can trigger the assimilation to some extent. Speech contexts that can trigger the assimilation have an amplitude modulation in the beginning.

Therefore, Experiment 1 tested whether a noise segment to which amplitude modulation was added, can cause similar context effects as a speech-sound context. Experiment 2 tested effects of amplitude modulation applied to a pure tone and Experiment 3 tested the effects of a frequency-modulated context. All three experiments revealed context effects similar to the effects caused by speech sounds. It is argued that these results show that compensation for phonological assimilation builds on general perceptual and not on speech-specific processes.

Perception, as a rule, is context-sensitive. In the visual domain, this is evident in, for instance, the demonstrations of the Gestalt school, *mach bands*, the phenomenon of color constancy, and the Müller-Lyer illusion, to name but a few. In the auditory domain, similar adaptations to context have been observed (see Warren, 1999, for an overview). To give an example: a monaurally presented sound is perceived to have a more central location if a sound with a similar spectral make-up is presented simultaneously to the other ear (e.g., Warren & Bashford, 1976). It is also generally accepted that speech perception is, for various reasons (Allen & Miller, 2001), context-sensitive. Nevertheless, the mechanisms that give rise to the context sensitivity in speech perception are sometimes fiercely debated (see, e.g., Fowler, 1992; Fowler, Brown, & Mann, 2000; Lotto & Kluender, 1998).

One aspect of context sensitivity in speech perception is that the interpretation of a speech segment depends on the phonetic environment in which a speech segment is embedded (Beddor & Krakow, 1999; Fowler & Brown, 2000; Fowler et al., 2000; Lotto & Kluender, 1998; Lotto & Kluender, & Holt, 1997; Mann, 1980; Mann & Repp, 1981; Nearey, 1989; Sussman, Fruchter, Hilbert, & Sirosh, 1998). These influences are the inverse to signal-disturbing context effects in speech production (see, e.g., Farnetani, 1997) and are therefore functional in achieving a reliable, abstract code with which lexical access can be achieved (McQueen & Cutler, 2001). That is, in production, a vowel may become more nasalized by an adjacent nasal consonant, while, in perception, the same vowel is perceived as *less* nasalized if accompanied by a nasal consonant. Although the first reports of such context effects date back to the early eighties (Mann, 1980; Mann & Repp, 1981), a new facet of this context sensitivity has been discovered more recently. Gaskell and Marslen-Wilson (1996) showed that phonological assimilations are processed by a context-sensitive mechanism. Phonological assimilations are rather drastic context-driven distortions that arise in speech production. In case of a phonological assimilation, a phoneme is not slightly modified by a surrounding phoneme as by coarticulation, but seems to lose its 'identity' completely. In English, the /n/ in 'lean,' for instance, may be pronounced rather like an /m/ in 'lea(n/m) bacon.' That is due to the rule of Coronal Place assimilation that is applied in English, which allows coronal word-final nasals and stops to take over the place of articulation of the onset of the following word. Gaskell and Marslen-Wilson showed that such deviations from a canonical form as 'lean' are treated differently in different contexts. The deviation 'leam' is accepted as an instance of 'lean' if, and only if, the next word starts with a phoneme that can trigger the /n/-to-/m/ change. This is the case in 'leam bacon' but not in *'leam gammon.' This result has now been replicated with diverse techniques ranging from cross-modal priming to automatically evoked event-related potentials (Gaskell & Marslen-

Wilson, 1996, 1998; Gow, 2002a, 2003; this dissertation, chapters three and six). Given that the observed pattern in perception is the inverse to the process of assimilation in production, we will coin the effect 'compensation for assimilation.'

In contrast to the generally accepted view that the perception of phonological changes is context sensitive (but see Lahiri & Reetz, 2002), quite different mechanisms have been proposed in order to account for this context sensitivity. First, Gaskell and Marslen-Wilson (1998) proposed that the listener 'learns' to tolerate certain deviations due to co-variant learning on a phonological level during language acquisition. That means that the listener 'knows' that a /m/ before a /b/ may signal a /nb/ sequence. This process was called 'phonological inference.' In contrast, Gow (2002a) proposed a feature-parsing account. According to this view evidence for phonological features is grouped by gestalt principles (as proposed by Bregman, 1990) on a phonological-feature level. This view builds on detailed acoustic investigations of assimilated segments (e.g., Gow & Husami, 1999; Nolan, 1992), which show that assimilated segments often carry evidence for the intended place of articulation and the place of articulation of the next phoneme. In the example given, the /m/ in 'lean bacon' still bears some resemblance to the originally intended /n/ and differs from an intended /m/ as in 'some bacon.' The problem for the listener is then reduced to the problem of parsing the feature cues correctly. The cues for a labial nasal in 'lean bacon' must be parsed from the assimilated segment and ascribed to the assimilating segment in 'bacon.' Then, only the residual cues for an alveolar nasal remain, and the token can be identified as an instance of 'lean.'

A third possibility has been proposed in chapter six. There, we argued that the context sensitivity might also be accounted for by a perceptual integration of target and context on early, auditory levels of processing. This proposal builds on the work of Macmillan and Kingston (Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickley & Bartels, 1999). According to this framework, a phonological categorization of a segment is achieved in (at least) two processing steps. First, auditory processes lead to an intermediate representation in an auditory perceptual space. Then, decision rules associate certain regions of this perceptual space with certain phonological categories. In this framework, a context effect can arise by means of either perceptual integration of target and context or a context-sensitive decision rule. What do these possibilities imply if applied to the perception of assimilated segments? If there is a context-sensitive decision rule, the auditory processing of the assimilated segment is independent of the context supplied by the assimilating segment. However, the classification function which associates regions from this perceptual space with phonological categories is context-sensitive. Assuming a perceptual integration of target and context, however, implies that the assimilating

segment influences the perception of the assimilated segment on early, auditory processing levels. Then, the context already influences the position of the assimilated segment in an auditory perceptual space and the same decision rule leads to a different categorization of the same stimulus in different contexts.

Whether the context-sensitive end product of the perceptual process is due to either a perceptual integration of target and context or due to a context-sensitive decision rule can be probed by the complementary use of an identification and a discrimination task. If the context effect arises as a consequence of a context-sensitive decision rule, intermediate representations in a perceptual space should be independent of their context. Hence, the perceptual difference between 'leam' and 'lean' should not depend on the following context. It follows that discrimination performance should not be influenced by context. However, there should be an influence of context on identification performance, because the context influences the decision rules, which associate regions of perceptual space with phonological categories. In the case of compensation for phonological assimilation, this translates to the prediction that both 'lean bacon' and 'leam bacon' are perceived as containing instances of 'lean' in an identification task, while 'leam gammon' is rejected as an instance of 'lean.' In contrast, discrimination between 'lean' and 'leam' should be relatively easy in both the viable and the unviable context. If, however, there is a perceptual integration of target and context on early auditory levels of processing, context effects are expected in both the identification task and the discrimination task. In this case, discrimination between 'leam' and 'lean' should be more difficult if these are followed by '...bacon' than if followed by '...gammon.'

In our study on the perception of Hungarian assimilated utterances (chapter six), we tested these predictions using an identification task and a discrimination task by Hungarian and Dutch listeners. The Hungarian word 'bal' (Engl. 'left') was presented to listeners in either its original form or changed to 'bar.' According to the rule of liquid assimilation in Hungarian phonology (see Olsson, 1992, p.57; Siptár & Törkenczy, 2000, p. 182), this change is allowed in 'balról' → 'barról' (Engl. 'from the left') but not in 'balnál' → *'barnál.' (Engl. 'at the left'). In the identification task, instances of 'barról' were often (mis-)identified as instances of the utterance 'balról. However, the same token of 'bar' was almost never identified as an instance of 'bal' in *'barnál,' which is illegally derived from 'balnál' (Engl. at the left). In a second experiment, the same tokens were also used in a discrimination task. Participants made more errors in discriminating 'bal' from 'bar' if followed by 'ról' than if followed by 'nál.' Furthermore, this result was to a great extent independent of experience with the rule of Hungarian liquid

assimilation. Dutch listeners were neither familiar with the Hungarian assimilation rule nor is there a similar rule in Dutch. Nevertheless, they produced similar results as Hungarian listeners.

Together, these results lend support for the assumption that compensation for assimilation is achieved by a perceptual integration of target and context. The same results challenge the feature-parsing and the phonological-inference account. Only a perceptual-integration account predicted an effect of context in a discrimination task. In the framework of Kingston and Macmillan (Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Dickey & Bartels, 1999), the feature-parsing and the phonological-inference account seem to advocate a context-sensitive decision rule. Consequently, these accounts predict that a context effect should only be evident in the identification task, but not in the discrimination task.

In addition, the absence of an effect of native-language experience indicates the involvement of general perceptual principles in the processing of assimilated word forms. This challenges the phonological-inference account, which assumes that implicit knowledge of the assimilation rules of a language is crucial in order to develop adequate compensation for assimilation. However, the results obtained in chapter six clearly contradict the predictions of the phonological-inference account.

However, to make a clear distinction between a perceptual-integration account and a feature-parsing account is more difficult. Although we described the feature-parsing account as assuming context-sensitive decision rules, it is possible to re-formulate a feature-parsing account so that it can account for the context effect in the discrimination data. One only has to assume that the perceptual space in which feature-parsing is supposed to happen cannot be probed by means of a discrimination task.

In this context, it is interesting to note that the feature-parsing account bears more than a superficial resemblance to the theory of direct perception (see, e.g., Fowler, 1996). Within the framework of the theory of direct perception of speech, it is also assumed that the speech-perception system 'parses' context-dependent variation from a target and assigns it to its source (Fowler & Brown, 2000). In comparison, Gow (2003) argued that the evidence for the place of articulation of the assimilating segment is parsed from the assimilated segment and assigned to the assimilating segment. However, to what extent Gow's account is to be looked upon as a special case of the theory of direct perception of speech is not of prime importance here. What is important is that the theory of direct perception assumes that only the outcome of the parsing process is accessible for conscious report as necessary for a discrimination task. Failures to detect context-dependent variation in an appropriate context in a discrimination task (Fitch, Halwes, Erickson, & Liberman, 1980) are assumed to lend support for the assumption that speech

gestures are the objects of perception (see Fowler, 1996, p. 1740). Within the framework of feature parsing, it could similarly be argued that phonological features are the objects of perception. If one additionally assumes that phonological features undergo an obligatory parsing process, it may be argued that the feature-parsing account is supported and not falsified by the observation that performance in a discrimination task is context-sensitive.

Are the feature-parsing account and the perceptual integration account then two different versions of the same theory? No, these two accounts can be distinguished along similar lines as the direct-perception theory of speech perception and the auditory enhancement theory (see Fowler et al., 2000 and Lotto & Kluender, 1998, respectively). According to the perceptual-integration account, the auditory processing of target and context integrates on early non-linguistic levels of processing. Therefore, the 'currency' in which integration occurs is an auditory one. That is, there must be an identifiable *auditory feature* of the context phoneme that leads to a perceptual integration. Whether this auditory feature occurs in a speech sound or in a non-speech sound should not matter. In contrast, the feature parsing account assumes that evidence for the *phonological feature* of the assimilating segment is parsed from the assimilated segment in order to be assigned to the assimilating segment. Therefore, the context must contain the appropriate phonological feature in order to be able to 'catch' the evidence for the assimilating feature that is parsed from the assimilated segment (see Gow, 2003, Exp. 3). Thus, the feature-parsing account and the perceptual-integration account differ in their assumption of the speech-specificity of the processes involved.

In this chapter, we are aiming to gather relevant evidence on this issue by extending our earlier work of (chapters six and seven). We already showed that Dutch listeners react to phonological changes in Hungarian utterances in a context-sensitive way. If a phonological change is presented in a context in which the change is, according to Hungarian but not according to Dutch phonology, viable, Dutch, as well as Hungarian, listeners have more problems in detecting this change than in a context in which the phonological change is unviable. We will therefore test whether this perceptual bias, which seems to be independent of language experience, is better explained in terms of a feature-parsing account than in terms of a perceptual-integration account. To this end, we test whether non-speech context sounds are able to generate similar context effects as the speech contexts used in chapter six.

Experiment 1

In this experiment, we use the same target stimuli as in chapter six. These were tokens from a speech-sound continuum ranging from [bɔl] to [bɔr]. /bɔl/ is the Hungarian word for

left, /bɔr/ is not a word in Hungarian. These tokens were concatenated with their appropriate case suffix for the Relative case (/ro:l/ indicating a movement from a surface) and with the case suffix for the Adessive (/nal/, indicating a position to the side of something). The change from a lateral /l/ to a trill /r/ is possible in the bi-morphemic word /bɔlro:l/ (Engl. from the left), but not in the word /bɔlnal/ (Engl. at the left). Hungarian and Dutch listeners perceived the target sounds ([bɔl] vs. [bɔr]) more categorically¹ and discriminated them more easily in the context that did not allow the assimilation ([nal]) than in the context that allowed for the assimilation ([ro:l]).

The Hungarian target words [bɔl] and [bɔr] differ in the manner of articulation of the coda, which is an apical lateral in [bɔl] and an apical trill in [bɔr]. The main acoustic feature differentiating an apical trill /r/ from an apical lateral /l/ is the presence or absence of amplitude modulation (AM). In addition, the two context sounds also differed with regard to the presence of AM at the onset. There is an AM modulation in the onset of the viable context [ro:l], but not in the unviable context [nal]. It is conceivable that the context effects observed in chapter six were due to a perceptual integration of the AM in the target with the AM in the viable context. If this is the case, any sound carrying AM at the same rate as the speech context might give rise to a similar context effect. To this end, we tested the impact of a non-speech context with and without AM on the perception of a [bɔl] to [bɔr] speech-sound continuum. In adherence to our methodology as applied in chapter six, perception was probed by means of an identification and a discrimination task. The discrimination task was the four-interval oddity (4I-oddity) task. In this task, four stimuli presented at a constant inter-stimulus interval (ISI) are presented to the listener, of which three are identical (the standard) and one is different (the 'odd'). The oddball stimulus always appears in the second or the third position. After hearing the train of four stimuli, the listener has to indicate whether the second or third stimulus was the odd. In order to prevent a combinatorial explosion of conditions, we did not probe all possible contrasts in the discrimination task. As we are interested in how far listeners are able to detect a deviant from a canonical pronunciation, only the original stimulus with a lateral /l/ was used as standard in the 4I-oddity task, while the stimuli with differing degrees of AM served as odds. This discrimination task was chosen because it seems to be a task in which performance is not influenced by a

¹ It should be noted here that we use the term 'categorically' in a purely descriptive sense in order to convey a difference in steepness of the identification functions. That is, 'less categorical' means that participants were overall more uncertain in their responses. We do not mean to suggest that the perception is more categorical as implied by a strong theory of categorical perception.

phonological processing stage. Gerrits (2001, chapter 3; see also Schouten, Gerrits, & van Hessen, 2003) showed that discrimination performance in a 4I-odddity task does only depend on the acoustic difference between two stimuli. Whether the two stimuli do belong to the same phonological category (within-category discrimination) or not (between-category discrimination) does not impede or enhance discrimination.

Method

Participants

Ten students (nine female, one male) of the University of Maastricht participated in the experiment and were paid for participation. All participants were native speakers of Dutch, free of any known hearing problems, and right-handed. Participants were aged from 19 to 24 (mean: 20.5). All participants were fluent speakers of the English language. In addition, some of the participants also spoke some German and French. One participant also spoke Indonesian. However, no participant had any knowledge of Hungarian.

Materials

In order to generate a continuum from [bəl] to [bər], a female native speaker of Hungarian was recorded uttering multiple tokens of the canonical form [bəlnal] (Engl. 'at the left') and an unviable variant *[bərnal] plus the canonical form [bəlro:l] (Engl. 'from the left') and a viable alternative [bərro:l]. The sample frequency for the recording of the natural utterances was 22050 Hz. Recordings were band-pass filtered from 130 to 8000 Hz. The first syllable from an utterance of [bəlnal] was extracted to serve as a template for all target stimuli. This syllable was then edited with the software package PRAAT 4.0 (Boersma & Weenink, 2002) in order to create a continuum of speech sounds from the original apical lateral to an apical trill.

An LPC analysis with sixteen predictors yielded a stable solution and was used in order to estimate source and filter for this utterance. Given the identical place of articulation, the primary cue for the lateral/trill distinction is the presence of amplitude modulation (AM) in the trill. For the edited sound stimuli, one cycle of AM (20 Hz) was added to the estimated source using five steps from 0 dB (no AM) to 12 dB (strongest AM). This was done by editing the intensity function of the source as estimated by PRAAT.

Formant estimations showed similar formant transitions from the vowel part to the consonant for both apicals and trills, but a slightly (200 Hz) lower third formant for the resonant period of the trill. Therefore, the third formant was also lowered in five steps of 0.1 bark. Filtering the edited sources with the edited filters yielded five stimuli, in which the first stimulus ended in an unaltered apical trill and the last stimulus imitated a final apical trill. In order to convey a casual speaking style, pre-voicing was shortened from 40 ms to 27 ms by cutting two complete cycles of prevoicing.

The non-speech context sounds were created in the following way. A 2 s long stretch of white noise was generated and then convolved with speech context sounds [nal] and [ro:l] used in the studies reported in chapter six. This generates a sound which approximates the long-term spectrum of the speaker, but does not contain any speech like information. A 300 ms piece was extracted from the middle part of the edited 2 s noise and a linear fade-out was applied to the last 30 ms of the stimulus. The noise was then equated in mean amplitude with the speech-context sounds. This stimulus was used as the non-speech analogue of the context speech sound [nal], which does not allow the phonological change from [bəl] to [bər]. This sound will be called -AM in order to indicate that this sound does not contain AM. In order to create a non-speech analogue of the context [ro:l], which allows the change to occur, the same AM technique as for the target stimulus was applied to the noise to create a +AM stimulus that was identical to the -AM stimulus in every respect but the presence of AM. These context sounds were concatenated with the speech stimuli in order to generate the experimental stimuli. These stimuli were presented to the participants sitting in a sound attenuated booth over headphones (Sennheiser HMD 25-1) using the ERTS program (Behringer, 1996).

In the discrimination task, four stimuli, each consisting of a speech stimulus and a constant context sound, were presented at a rate of 1 Hz. The standard speech stimulus was always the original speech stimulus with no AM and the odd stimulus in the second or third position, was any of the stimuli with AM ranging from 3 to 12 dB. These stimuli were presented in concatenation with one of the non-speech contexts.

Procedure

Participants were seated in a sound-attenuated booth and faced a computer monitor. Instructions were given in written form via the computer monitor. Participants first completed the identification task and then the discrimination task. For the identification task, participants were instructed that they were going to hear the Hungarian word for left 'bal' spoken by a speaker who sometimes makes an error, and pronounces 'bal' as 'bar.' Participants were asked to indicate, after hearing a speech sound, whether the speech sound was 'bal' or 'bar.' 500 ms before hearing a stimulus, the computer screen displayed the two answer alternatives ('bal' versus 'bar') on the left and right of the screen, corresponding to the response key allocation. After hearing a stimulus participants had 2.5 s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, the word on the screen corresponding to the response alternative was moved up and to the margin of the computer screen by 6 pixels while the other alternative disappeared. This indicated to the participants that the computer had registered their answer.

In order to familiarize the participants with the stimuli, they first judged the stimuli in isolation. In this training block, each stimulus was presented five times in a random order. Then the experimental conditions started. The Context conditions (+AM noise, -AM noise) were blocked. The order of presentation within blocks was randomized. The presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants first heard the stimuli with the +AM context, while the other half heard the stimuli first with -AM context.

After completing the identification task, participants were handed a short questionnaire asking for hearing problems, age, handedness, and foreign-language knowledge. After completion of the questionnaire, the discrimination task followed. The instructions for the discrimination task were presented via the computer monitor. The instruction informed participants that they are going to hear four sounds on each trial, of which three were identical and one was different. Participants were informed that their task was to pick the odd one out, which was be either the second or third of the stimuli. Instruction stressed that the odd stimulus might be the same word as the standard, only pronounced differently. In every trial of the discrimination task, the computer screen displayed the two answer alternatives (the digits 'two' and 'three') on the left and right of the screen, corresponding to the response key allocation, 500 ms before hearing a stimulus consisting of four speech sounds. After hearing a train of four speech sounds, participants had 3s to respond. If no response was given in this time, a feedback screen asked participants to respond faster. If a response was given, feedback indicated whether the choice was correct or not.

The Context conditions (\pm AM) were blocked in four blocks of 40 trials each. The order of presentation within blocks was randomized. The presentation order of the blocks was counterbalanced over participants. In every participant group, half of the participants heard the stimuli with the +AM context in the first and third block and the stimuli with -AM context in the second and fourth block. For the other half, the targets with the -AM context were presented in the first and third block and the targets with the +AM context in the second and third block.

Design

The design of the identification task entails two independent within-subject variables. The first independent variable is the AM in the target syllable, to be identified as 'bal' or 'bar,' with five levels from 0 to 12 dB AM. The second independent variable is the Context with the levels +AM and -AM. Each of the ten cells of this design was presented to each participant 16 times and the dependent variable is the percentage of 'bal' responses for every cell of the design.

The design of the discrimination task entails two independent variables. The first independent variable is Δ AM between standard and odd stimulus, with four levels from 3 to 12 dB, and the second independent variable is the type of context stimulus with the levels +AM and -AM. For each of the eight cells of this design, ten trials each were presented with the odd stimulus in the second and third position. From these 20 trials per cell, the dependent variable d' was calculated. We arbitrarily defined the second position as the target. Hence, a correct response in a trial with the odd stimulus at the second position was counted as a hit, while an error in a trial with the odd stimulus at the third position was counted as a false alarm.

Results

Identification task

Table 8.1 shows the mean percentage of 'bal' responses for every cell of the design. The descriptive data show that the identification function is steeper in the -AM condition than in the +AM condition. A repeated-measure ANOVA performed on the these data revealed a significant effect of AM depth ($F(4, 36) = 40.77, p < 0.001$), but no main effect of Context ($F <$

Table 8.1 Percentages of 'bal' identifications in Experiment 1

Context	AM depth in target				
	0 dB	3 dB	6 dB	9 dB	12 dB
+AM	73.6%	59.6%	40.8%	29.1%	22.2%
-AM	94.4%	81.7%	41.3%	23.2%	21.5%

1). However, there was a significant interaction of AM depth and Context ($F(4, 36) = 4.76, p < 0.01$), which shows that the +AM context led to a significantly shallower identification function.

Discrimination task

Table 8.2 shows the mean d' scores for every cell of the design. A repeated-measure ANOVA performed on these data revealed a significant effect of ΔAM ($F(3, 27) = 12.71, p < 0.001$), a significant main effect of Context ($F(1, 9) = 16.38, p < 0.005$), and the interaction between these factors ($F(3, 27) = 3.78, p < 0.05$). The interaction is due to the fact that the effect of context was significant for all levels of the factor ΔAM (6 dB: $t(9) = 2.97, p < 0.05$, 9 dB: $t(9) = 3.04, p < 0.05$, 12 dB: $t(9) = 3.62, p < 0.02$), but the 3 dB level ($t < 1$).

Discussion

The non-speech sounds clearly influenced the perception of the speech-sound continuum. Identification was more uncertain and discrimination more problematic when the context sound carried an AM. This mirrors the findings obtained earlier (chapter six), in which the AM carrier was a speech sound. Therefore, the present results support the assumption that the perception of phonological changes rests to a rather substantial degree on general perceptual principles and is not speech-specific.

Table 8.2 Mean d' scores in Experiment 1

Context	ΔAM				
	3 dB	6 dB	9 dB	12 dB	mean
+AM	0.32	1.13	1.52	1.33	1.08
-AM	1.26	3.56	4.06	3.90	3.20

The present results are difficult to reconcile with the assumptions of a feature-parsing account. This account assumes that the evidence for the trill in the first /r/ in the assimilated form [bʊrro:l] is parsed from this segment and assigned to the trill in the case suffix [ro:l]. This parsing is supposed to happen on a phonological-feature representation of the input (see the interpretation of Exp. 3 in Gow, 2003). As the current experiment used a non-speech target that probably does not carry evidence for phonological features, there should be no context to which the feature cues of the assimilated target can be assigned. Accordingly, there should be no effect of the non-speech sounds on the perception of the phonologically changed forms. However, the present results show that there is a clear context effect with non-speech sounds that mirrors the context effects caused by speech sounds.

One possibility to counter this argument is to argue that the context sound was too much speech like, and therefore sufficient to lead to a preliminary detection of the features for the apical trill in the context. By design, the noise stimuli used had a similar long-term spectrum as speech. Therefore, it may be argued that such a noise, with added AM, sounds similar to a whispered /r/. Such a sound might lead to an activation of a phonological feature, which, in turn, would allow phonological feature-parsing. To rule out such an alternative explanation, and in order to test the generality of the results of Experiment 1, we ran Experiment 2 with a different non-speech context sound.

Experiment 2

In this experiment, we tested whether the results of Experiment 1 depend on the spectral composition of the carrier sound to which the AM is applied. In the previous experiment, the spectrum of the noise to which AM was applied was deliberately speech-like. Therefore, the question arises whether the effects observed in Experiment 1 are due to a spectral overlap of the carriers of the AM in the target and in the non-speech context. To this end, we replicated Experiment 1 with different non-speech contexts. In this experiment, we used a pure tone with a frequency of 400 Hz. The spectral composition of this sound is thus completely unlike speech. If the pattern of context effects as observed in Experiment 1 is replicated, this indicates that AM occurring in non-speech sounds is perceptually integrated with the AM in the speech sounds, independent of the spectral composition of the context. Again, we tested the impact of these non-speech sounds on the identification and discrimination of the target speech sounds.

Method

Participants

9 students of the University of Maastricht and one non-scientific staff member participated in the study. They were paid for participation. All participants were female; and all participants but one were right-handed. Nine of the participants were aged 20 to 24 and one participant was aged 50. All participants were native speakers of Dutch and free of any known hearing impairment. All participants spoke English fluently. In addition, some of the participants spoke some German or French or both of these languages. One participant also spoke Spanish fluently. However, none of the participants was familiar with Hungarian.

Materials

The targeted speech stimuli were the same as in Experiment 1. The -AM context sound was a 0.3 s long pure tone with a frequency of 400 Hz, which was linearly clipped for the initial and final 10 ms. The +AM sound was created by multiplying the -AM sound with the amplitude envelope of the +AM sound from Experiment 1. These non-speech sounds were concatenated with the target speech sounds.

Procedure and Design

Procedure and Design were the same as in Experiment 1.

Results

Identification

The results of the identification experiment, expressed in percentage 'bal' responses are shown in Table 8.3. The results replicate the pattern observed in Experiment 1, showing that the identification function is shallower in the case of the +AM context. A repeated-measure ANOVA revealed a significant effect of AM depth ($F(4, 36) = 33.9, p < 0.001$), but no main effect of Context ($F < 1$). However, the interaction between the two factors is significant ($F(4, 36) = 3.08, p < 0.05$), replicating a similar effect in the previous experiment.

Discrimination

The results of the discrimination experiment, expressed in mean d' scores are shown in Table 8.4. A similar pattern as in Experiment 1 is observed; discrimination performance was worse in the +AM condition. A repeated-measure ANOVA revealed a significant effect of Context ($F(1, 9) = 7.56, p < 0.05$) and a significant effect of ΔAM ($F(3, 27) = 17.52, p < 0.001$). The interaction between these two factors was not significant ($F < 1$).

Table 8.3 Percentages of 'bal' identifications in Experiment 2

Context	AM depth in target				
	0 dB	3 dB	6 dB	9 dB	12 dB
+AM	67.4%	48.6%	23.3%	13.2%	11.3%
-AM	78.9%	55.1%	15.0%	4.3%	3.1%

Table 8.4 Mean d' scores in Experiment 2

Context	Δ AM				mean
	3 dB	6 dB	9 dB	12 dB	
+AM	0.53	2.37	2.80	3.58	2.32
-AM	1.65	4.34	4.48	4.51	3.82

Discussion

The results of the present experiment show that the context effect found in Experiment 1 does not depend on an overall spectral overlap between the AM in the speech target and the AM in the non-speech context. In both the identification task and the discrimination task, remarkably similar results have been obtained in both experiments. The present results therefore indicate a perceptual integration of the AM in the speech-sound target and the AM in the pure tone. It seems that amplitude modulations are perceptually integrated irrespective of the frequency range in which they occur.

However, there is one difference between the results of Experiment 1 and Experiment 2. In the previous discrimination experiment, there was an interaction between Context and Δ AM: The effect of Context was significant at all but the lowest level of Δ AM. In the present experiment, the overall significant effect of Context did not interact with Δ AM. It should be noted that the overall level of discrimination performance was better in Experiment 2. Therefore, the interaction between the independent variables Context and Δ AM in Experiment 1 may be due to a floor effect. At the lowest level of Δ AM, discrimination performance was near chance, which may have made it more difficult to detect a significant effect of Context.

Nevertheless, Experiment 1 and 2 show that non-speech stimuli are able to influence the perception of phonological changes. While the non-speech sounds in Experiment 1 may still have led to a weak activation of phonological-feature detectors, it is difficult to see how a phonological-feature detector could be triggered by the presentation of an amplitude-modulated pure tone, as in Experiment 2. Hence, these results support a perceptual-integration account for compensation for assimilation. The results are difficult to explain in terms of the feature-parsing account (Gow, 2002a, 2003). This latter account assumes that context effects only arise if the context carries evidence for phonological features. This might have been possible for the noise stimuli used in Experiment 1. However, it is difficult to see how the speech perception system could mistake the AM in the pure-tone context used in Experiment 2 to indicate the presence of phonological features. This buttresses the assumption that the perception of phonological changes is strongly influenced by basic auditory processes and does not reflect a speech-specific

process. In the next experiment, we aim to find converging evidence for the supposed auditory locus of the compensation-for-assimilation effects.

Experiment 3

In this experiment, we modulated the frequency of a non-speech context sound instead of its amplitude. Single-cell recordings in early cortical auditory areas have revealed that cells reacting selectively to a particular frequency of AM show a similar response to frequency modulated tones (Wang, Liu, & Liang, 2003). This result seems counterintuitive at first glance. If one considers the tonotopic organization of auditory cortex, however, it is evident that frequency modulation (FM) and AM lead to similar effects on a given single cell. Consider a cell with a characteristic frequency of f Hz responding to a FM tone with f Hz at the center of the FM. Given a sine-wave form of the FM, this cell gets more or less the same input as for an AM tone with a modulation frequency of $2f$. This is due to the fact that the center of the FM is passed through twice within one cycle of FM. For cells with a characteristic frequency at the lower or higher limits of the FM, the input will be similar as for an AM tone and a FM tone with a modulation frequency of f . This follows from the fact that the upper and lower frequency limits are only passed through once within one cycle of FM. Therefore, we expect that FM may in principle also cause the same context effects as AM on the speech-sound continuum. This was tested in this third experiment.

Method

Participants

12 participants (nine female and three male) from the same subject pool as in the previous experiments participated in this experiment. They were paid for their participation. Two of the participants were left-handed, the rest was right handed. The participants were aged 18 to 25 (mean 21.8 years). All participants were native speakers of Dutch and free of any known hearing impairment. All participants spoke English fluently. All but one participant had additional foreign-language proficiencies. Eight participants spoke French, four of them fluently. Seven participants spoke German, of which three fluently. Three participants spoke Spanish; one of these also spoke Surinamese. In addition, one participant spoke Frisian, a Germanic dialect still spoken in the northern part of The Netherlands. However, none of the participants was familiar with Hungarian.

Materials

The same speech targets and the stationary -AM stimulus as in the previous experiments were used. This sound was a steady pure tone with a frequency of 400 Hz, and thus a tone without frequency modulation (FM). Hence, we call this sound the -FM sound. In order to create a +FM sound, FM at the rate of 50 Hz and a depth 100 Hz in a sinusoidal fashion was added. That leads to a minimum of 300 Hz and a maximum of 500 Hz, which leads a 0.95 bark range between center (400 HZ) and minimum and a 0.89 bark deviation from center to maximum. This is a rather strong FM depth, which is not too strongly skewed due to the non-linear relation between the Hertz scale to perceived height.

Procedure and Design

The procedure and Design were the same as in Experiment 1, with the only difference that the independent variable context now has the levels \pm FM instead of the \pm AM.

Table 8.5 Percentages of 'bal' identifications in Experiment 3

Context	AM depth in target				
	0 dB	3 dB	6 dB	9 dB	12 dB
+FM	83.9%	65.2%	33.3%	16.7%	9.9%
-FM	93.8%	71.4%	24.5%	6.8%	6.3%

Results

Identification task

The results of the identification experiment, as percentage 'bal' responses are shown in Table 8.5. The results show that the identification function is shallower in the case of the +FM context. A repeated-measure ANOVA revealed a significant effect of AM depth ($F(4, 44) = 79.74, p < 0.001$), but no main effect of Context ($F < 1$). However, the interaction between the two factors is significant ($F(4, 44) = 4.20, p < 0.01$), replicating a similar effect in the previous experiments.

Discrimination task

The results of the discrimination experiment, expressed in mean d' scores are shown in Table 8.4. As in the previous experiment, discrimination performance is worse in the context with a modulation, which is a FM here. A repeated-measure ANOVA revealed a significant effect of Context ($F(1, 11) = 10.63, p < 0.01$) and a significant effect of ΔAM ($F(3, 27) = 8.99, p < 0.001$). The interaction between these two factors was not significant ($F < 1$).

Discussion

The present experiment succeeded in inducing a similar context effect as the earlier experiments. On a more speculative note, the present result may be taken as evidence for the hypothesis that compensation for assimilation might occur in early cortical areas. This is, because

Table 8.6 Mean d' scores in Experiment 3

Context	ΔAM				
	3 dB	6 dB	9 dB	12 dB	mean
+FM	0.71	2.09	2.00	2.37	1.74
-FM	1.51	2.63	2.58	3.49	2.61

FM and AM lead to similar effects in tonotopically organized areas. Both AM and FM lead to an increase and decrease in energy at a particular frequency. An interpretation favoring an auditory basis of the effect is fostered by our electrophysiological studies (chapter seven). There we found that context effects in assimilated utterances could be detected with event-related potentials (ERPs) during passive listening. An ERP component called MisMatch negativity (MMN), which indicates a perceptual distance between two stimuli (see, e.g., Näätänen & Winkler, 1999), displayed context sensitivity when presented with the phonological change [bɔl] → [bɔr].

The MMN was smaller if the change occurred in a phonological context that allowed the change (i.e., [bɔrro:l]) than if the change occurred in a context that did not allow the change (i.e., *[bɔrnal]). Especially interesting for the current purposes is the fact that the MMN inverted polarity at the mastoid electrodes, indicating a generation in the temporal cortex. Although exact localization is difficult with ERP, the polarity reversal of the MMN indicates a generation of the effect in early cortical auditory areas, in accordance with the results of this experiment. Although both interpretations are admittedly speculative, they are buttressed by the report of Scott and Wise (2003) who report that results from functional-imaging studies indicate a similar location for the processing of AM.

General Discussion

In this series of three experiments we investigated the effect that following non-speech context sounds have on the perception of phonological changes in a preceding speech sound. Gaskell and Marslen-Wilson (1996) showed that the perception of phonological changes is context-sensitive. A phonological change does not impede the recognition of the changed word, if the context allows the phonological change to happen as a consequence of a lawful phonological assimilation. Therefore, the Hungarian word /bɔl/ (Engl. 'left') is recognized in [bɔrro:l] but not in *[bɔrnal], because /l/ may be changed to /r/ in Hungarian before /r/ but not before /n/. Previously (see chapter six), we showed that this context sensitivity is not dependent on experience with such an assimilation rule in one's native language. Comparable effects were observed for Hungarian and Dutch listeners, notwithstanding the fact that Dutch has no phonological rule that changes /l/ into /r/. In this chapter, we evaluated whether this context sensitivity displayed by Dutch listeners responding to Hungarian words can also be elicited by non-speech sounds. This was clearly the case. Experiment 1 showed that the critical feature is the presence of AM in the context, and that it does not matter whether this AM occurs in a

speech sound, as in the study reported in chapter six, or whether the AM occurs in non-speech noise as in Experiment 1. Experiment 2 validated this finding by showing that this holds true even if the speech sound and the context sound are spectrally dissimilar. Finally, Experiment 3 showed that a similar effect was obtained if a frequency-modulated context sound was used instead of an amplitude-modulated sound. Combining this result with human electrophysiological (chapter seven), animal electrophysiological (Wang et al., 2003), and functional imaging results (Scott & Wise, 2003) suggests that the effects may possibly be located in early cortical auditory areas that specifically deal with AM.

These results favor the assumption that compensation for assimilation in speech perception is due to a perceptual integration of an assimilated target and assimilating context at early auditory levels. This is supported by the fact that context effects were present in both, discrimination and identification, tasks. Rival theories as a feature-parsing account, or a direct-realist adaptation of such an account, however, are difficult to reconcile with the present results. These accounts rest on the assumption that articulatory or phonological features associated with the assimilated target are parsed from this target and associated with the assimilating segment. Parsing can, however, only be successful if the context bears articulatory or phonological features. It is difficult to see how articulatory or phonological features can be ascribed to the non-speech segments used here. Therefore, these two accounts are challenged by the finding that non-speech sounds, not resembling speech, induced similar context effects as speech sounds.

The present results seem to indicate that the human auditory system is innately equipped with very powerful processing mechanisms that can deal with the invariance problem in speech perception. However, we want to put forward an alternative view. Instead of attributing 'powerfulness' to the auditory system, and to view it as a 'clever' adaptation to environmental forces (i.e., speech production), it may be argued that it is not the system that has adapted but rather the environment, that is, speech production, that adapts to the inherent properties of the perceptual system. Consider the following analogy (from Clark, 1997): A space traveler with an ultra-adaptationistic view observing mankind might report that humans have evolved in order to sit comfortably on chairs. What this imaginary space traveler failed to notice is that chairs are artifacts that humans have adapted to fit them. Similarly, language, and thus speech production, is in some respect also an artifact, which is amenable to change. Therefore, it is conceivable that the system of phonological changes in a given language adapts to the auditory abilities of the listener (see also the chapters in Hume & Johnson, 2001). In this view, we need not assume that aspects of the auditory system are extremely well adapted to the environment, but rather that languages use existing properties of the perceptual system. A final affirmation of this conclusion

would require evidence that such context sensitivities as reported here could also be found in non-human species. Context effects—as compensating for speaking rate (Stevens, Kuhl, & Padden, 1988) and for coarticulation (Lotto, Kluender, & Holt, 1997)—have already reported for non-human species. It remains to be seen whether similar effects can be obtained for compensation for phonological changes.

To summarize, the current experiments clearly show that regressive context effects in the perception of phonological assimilation can be triggered by non-speech sounds. This indicates that compensation for assimilation is not a speech-specific process but rather arises at auditory levels of processing.

In Search of the Auditory, Phonetic and/or Phonological Problems in Dyslexia¹

There is a growing consensus that developmental dyslexia is associated with a phonological core deficit. One symptom of this phonological deficit is a subtle speech-perception deficit. The assumedly auditory basis of this deficit is however still hotly debated. If dyslexics, however, do not have an auditory deficit and accordingly perceive the underlying acoustic dimensions of speech as well as normal reading subjects, why do they exhibit, e.g., a categorical perception deficit? A possible answer to this conundrum lies in the possibility that dyslexics do not adequately handle the context-dependent variation that speech signals typically contain. A mathematical model simulating such a sensitivity deficit mimics the speech-perception deficits attributed to dyslexia. Therefore, we empirically examined whether dyslexic children handle context-dependencies in speech differently than normal-readers. Contrary to our initial hypothesis, dyslexics did not show less context sensitivity in speech perception than normal readers at auditory, phonetic, and phonological levels of processing nor did they reveal any categorization deficit. It is therefore worthwhile to consider the possibility that time course aspects of the processes necessary to achieve lexical access, and not *phonological representations* per se, may be impaired in dyslexia.

¹ This is an modified version of Blomert, L., Mitterer, H. & Paffen, C. (submitted). In Search of the Auditory, Phonetic and/or Phonological Problems In Dyslexia: Context Effects in Speech Perception.

Developmental dyslexia refers to poor reading and writing in spite of adequate intelligence and the absence of any sensory disorder or neuropsychological signs. Dyslexics are, however, not only impaired in reading, but also show subtle deficits on a range of tasks that probe 'phonological processing' and do not involve written material. Dyslexics have been shown to be subtly impaired in tasks such as repeating words and nonwords (Brady, Shankweiler, & Mann, 1983; Brady, Poggio, & Rapalla, 1989; Snowling 2000), the retention of verbal material in working memory (Snowling, Nation, Moxham, Gallagher, & Frith, 1997; Witrok, Ho, & Schuster, 2002), rapid naming (Denckla & Rudel, 1976; Bowers & Swanson, 1991; Wolf, Bowers, & Kiddle, 2000), object naming (Katz, 1986), and meta-linguistic tasks that involve the manipulation of phoneme size units, so-called phoneme awareness tasks (Adams, 1990; Bradley & Bryant, 1983; Wagner & Torgesen, 1987). These lines of research have led to a general acceptance of the assumption that the core problem in dyslexia is best described as a phonological deficit (Catts, 1989, Shaywitz, 1998).

In search for the cause of this phonological deficit, some have implicated a significant role for auditory and speech perception. McBride-Change (1996), for instance, conducted a structural linear equations analysis, which showed that the variance in phoneme-awareness tasks could be accounted for by the performance on categorical-perception tasks. In a categorical-perception task, participants have to classify a continuum of speech sounds as, for instance, either /ba/ or /da/. Usually, most stimuli of such a continuum are perceived consistently as either /da/ or /ba/, and only a small range of stimuli is perceived as ambiguous. Dyslexics, however, tend to deviate from this pattern. They tend to show less categorical perception of short speech sounds. That is, they perceive a wide range of speech-sound continua as potentially ambiguous. Several studies investigating categorization performance found dyslexics to exhibit such problems, particularly with stop consonants. Godfrey, Sydral-Lasky, Millay and Knox (1981) showed that phoneme categories of reading disabled children were less well separated than for normal reading children, if they were presented with synthetic speech /ba/ - /da/ and /ba/ - /ga/ continua. Also, the children were less able to discriminate between /ba/ and /da/, even at the extremes of the continua. Reed (1989) used /ba/ - /da/ stimuli which differed in F2 and F3 frequency during the first 35 msec. Under these conditions, reading disabled subjects were impaired in discriminating between /da/ and /ba/. In sum, however, the results of categorization studies are equivocal. Some studies found phoneme discrimination and categorization deficits in dyslexics (e.g., Tallal 1980, Werker & Tees, 1987), whereas others did not (e.g., Snowling, Goulandris, Bowlby & Howell, 1986). The different results may in part be explained by the different stimuli used. Van Hessen and Schouten (1999), for example, found

that more natural stimuli were generally perceived more categorically than less natural stimuli. Blomert and Mitterer (in press) showed that dyslexics did not differ from controls when performing a phoneme-identification task with semi-natural stimuli, but did differ if the stimuli were synthetic.

The nature of the phonological-awareness and the categorization problem is still fiercely debated (e.g. Studdert-Kennedy, 2002). Some (Stein, & Walsh, 1997; Tallal, 1980; Tallal, Miller, & Fitch, 1993) have argued that the problems in handling speech sounds may be attributed to a general perceptual deficit in temporal processing, especially in the perception of brief auditory events. Indeed, a number of studies reported that dyslexics show subtle anomalies in the perception of brief acoustic events, as found in stop-vowel syllables (for a review, see McBride-Chang, 1995). Such observations have led to the theory that dyslexics are impaired in the processing of rapid auditory events. However, other research revealed that dyslexics not only have discrimination problems with brief formant transitions, but also showed impaired performance with longer transition times. In an MEG study Heim, Eulitz, Weinbruch and Elbert (1999) presented /ba/ and /da/ stimuli with CV transition times of 43 and 93 milliseconds respectively, modeled after the Tallal & Piercy (1975) stimuli. The electrophysiological data revealed no differences between dyslexics and controls, but the behavioral results revealed significant differences for both, the short and long transition times. This result clearly poses a problem for a temporal processing deficit account, which predicts unimpaired performance for long transition times.

The temporal processing deficit theory has also been challenged on theoretical grounds. Studdert-Kennedy and Mody (1995) argued that the construct 'temporal processing' is not well defined. They argued that the evidence in favor of impaired temporal perception in dyslexia is inconclusive. Results showing impaired processing of rapid sequences of auditory events do not necessarily imply a temporal deficit, but can also be explained by assuming that auditory recognition of the subcomponents of rapid auditory sequences is impaired. Moreover, findings of subtly deviant speech perception in dyslexics do not necessarily imply that there is a perceptual deficit. An alternative view is that the perception of acoustic cues is unimpaired, but that the utilization of these cues for the categorization of a speech sound differentiates dyslexics from normal-reading controls. This assumption gains credibility from the fact that there is no one-to-one correspondence between an acoustic cue and a phoneme category. For example, what differentiates the phonemes /b/ and /p/ is not only Voice-Onset Time (VOT), but, among other cues, also length and f_0 of the adjacent vowel. Therefore, categorizing acoustic events into phonological categories implies more than the application of a simple decision rule (if

VOT > -20 ms then /b/), but rather requires the application of a multi-dimensional nonlinear function to a multi-dimensional stimulus pattern.

In an insightful paper, Mody, Studdert-Kennedy, and Brady (1997) tried to tease apart whether the differences between dyslexics and normal-reading controls on categorical-perception tasks are caused by a general auditory or a speech-specific deficit. To this end, they used a synthetic /ba/-/da/ continuum that varied in F2 onset frequency. In line with earlier results, they found that dyslexics differ from normal-reading controls in the perception of this continuum. Then, they presented the isolated F2 transitions of the speech sounds, which were completely unlike speech and cannot be categorized into any speech categories. If dyslexics were impaired in the perception of short transitions in acoustic events, they should also be impaired in perceiving these isolated formant transitions. However, dyslexics and normal-reading controls performed equally well when discriminating isolated second formant transitions taken from the speech continuum. Mody et al. conclude that dyslexics may have trouble identifying phonetically similar, but phonologically contrastive synthetic syllables. A similar result was reported by Rosen and Manganari (2001). Moreover, Nitttrouer (1999) showed that dyslexics more strongly rely on formant transitions than on spectral cues for speech identification in certain paradigms. These results are not compatible with the assumption that dyslexics have a general temporal-auditory deficit. If this was the case, dyslexics should also be impaired in discriminating isolated second formant transitions and generally use formant transitions to a lesser degree.

But if dyslexics perceive the underlying dimensions of speech as well as normal-reading participants, why do they show a categorical-perception deficit at all? An alternative to an auditory-perception account is the hypothesis that dyslexics have less well defined phonological categories (e.g., Snowling, 2000). This might explain why dyslexics show less categorical perception in spite of adequate perception of the acoustic cues. In addition, Brady (1997) argued that such an assumption may also explain other phonological deficits that are associated with dyslexia, such as impaired phoneme-awareness and verbal short-term memory. However, it is yet unclear why dyslexics develop less sharply defined phonological categories. Even Mody et al. (1997, p. 201) leave open this question: "*speech categories may be for unknown reasons, broader and less sharply separated in reading disabled than in normal children* (p. 201). *The nature, origin, and extent of the perceptual deficit remains to be determined*" (p.227).

In this chapter we hypothesize that dyslexics may be impaired in coping with the enormous variation inherent to fluent speech (see, e.g., Farnetani, 1997). Such a failure would almost automatically lead to less well-defined phonological categories. Figure 9.1 aids in making this point. Consider the width-dimension of Figure 9.1A to be one acoustic dimension that

differentiates the syllables /ta/ and /ka/, for example, F3 onset frequency. The height-dimension represents the probability of a certain phoneme; the white function gives the probability for /ta/ and the gray-shaded function the probability for /ka/. The depth-dimension represents a context variable, which moderates how strongly a given F3 value is associated with a certain syllable. A prime candidate for this depth-dimension is the place of articulation required by the previous segment. If the previous phoneme has an alveolar place of articulation—with a high F3—, a following medium F3 is more likely to be associated with the velar /k/, due to the coarticulation of the preceding alveolar segment and the velar segment. This context dependency is illustrated in Figure 9.1A.

From these distributions, we now calculate the categorization function with and without taking the context dimension into account. The two distributions are quasi-normal distributions using the derivative of the logistic function with a $\mu_t = -2$ and $\mu_k = 2$. This makes the equations and the integration in step 3 more manageable than the using the Gaussian distribution. The probability distribution is then given by equation [1]. Note that the function is only defined for $[-5 < x < 5]$ and $[-2 > y > 2]$. This is a reasonable assumption given the articulatory nature of the width- and depth-dimension.

$$[1] \dots p_i(x, y) = \frac{\exp(-x + y + \mu)}{(1 + \exp(-x + y + \mu))^2}$$

The optimal categorization function for the two overlapping distributions is given by $p_t/[p_t+p_k]$ (Rojas, 1996), the probability for /t/ divided by the sum of probabilities, and *mutatis mutandis* for /k/. To calculate the optimal categorization function, we have to compute $p_t/[p_t+p_k]$, in which p_t and p_k are given by putting in the appropriate mean in [1]. Because the overlap of the three-dimensional function is identical for all y , it suffices to calculate the categorization function for any value of y . For $y = 0$, the categorization function is given by [2], which is the dotted line in Figure 9.1B.

$$[2] \dots C_t(x) = \frac{\frac{\exp(-x + 2)}{(1 + \exp(-x + 2))^2}}{\frac{\exp(-x + 2)}{(1 + \exp(-x + 2))^2} + \frac{\exp(-x - 2)}{(1 + \exp(-x - 2))^2}}$$

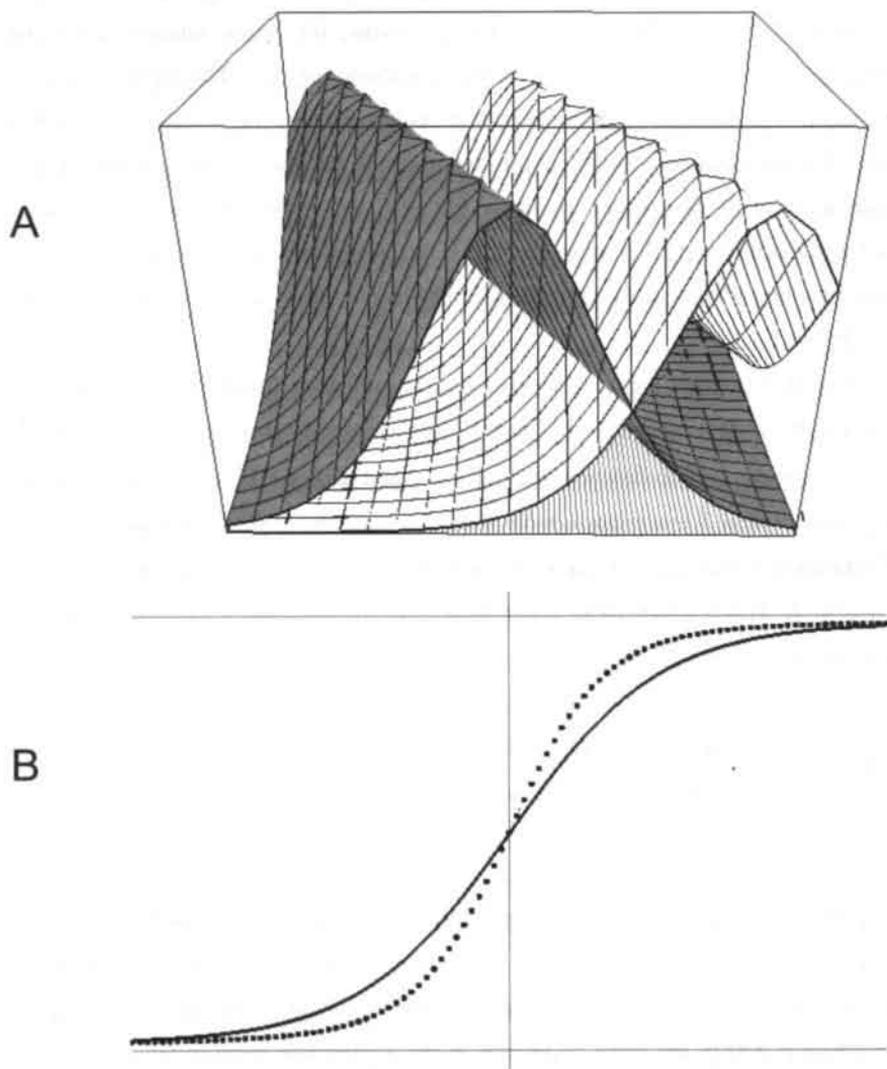


Figure 9.1 (A) The two functions represent a theoretical probability distribution of /t/ and /k/ with probability represented in the z-dimension. The x-axis represents a major cue differentiating /t/ and /k/ (e.g. F3 onset), and the y-axis represents a context sound, influencing the F3 onset. (B) The two classification functions resulting from an optimal classification of the data in (A). The dotted line represents the classification when context is used, the continuous line represents the best classification when the context dimension is disregarded.

Consider what would happen if the context variable (the depth-dimension) was not used in categorization. Then, the probability that a certain F3 (i.e., x) value is associated with /t/ or /k/ would be given by its mean value over all context positions. For a given value of x , this is the area under $f(x,y)$ in the interval $y = [-2,2]$ divided by the length of the interval, as calculated in [3].

$$\begin{aligned}
 [3] \dots p_2(x) &= \frac{1}{2 - (-2)} * \int_{-2}^2 \frac{\exp(-x + y + \mu)}{(1 + \exp(-x + y + \mu))^2} dy \\
 &= \frac{1}{4} * \frac{1}{1 + \exp(x + \mu - 2)} - \frac{1}{4} * \frac{1}{1 + \exp(x + \mu + 2)}
 \end{aligned}$$

If we use this function, to calculate the optimal categorization function, we get equation [4]. This categorization function leads to more overlap between the two categories. Consequently, as shown in Figure 9.1B, this leads to a shallower identification function than the identification function that takes the context dimension into account.

$$\begin{aligned}
 [4] \dots C_2(x) &= \frac{\frac{1}{4} * \frac{1}{1 + \exp(x + 2 - 2)} - \frac{1}{4} * \frac{1}{1 + \exp(x + 2 + 2)}}{\frac{1}{4} * \left(\frac{1}{1 + \exp(x + 2 - 2)} - \frac{1}{1 + \exp(x + 2 + 2)} + \frac{1}{1 + \exp(x - 2 - 2)} - \frac{1}{1 + \exp(-x - 2 + 2)} \right)} \\
 &= \frac{\frac{1}{1 + \exp(x + 1)} - \frac{1}{1 + \exp(x + 3)}}{\frac{1}{1 + \exp(x - 4)} - \frac{1}{1 + \exp(x + 4)}}
 \end{aligned}$$

This brings us back to the earlier problem: Why do dyslexics show a shallower categorization function, although they seem able to perceive the underlying acoustic distinction at normal levels? One possible answer to this question is that dyslexics handle context-dependent variation in speech inadequately. Therefore, they show a shallower identification function, comparable to the continuous line versus the dotted line in Figure 9.1B.

In this chapter, we aim to investigate whether this conjecture can be supported. Context effects in speech perception have been stipulated at multiple levels of processing. First of all, context dependencies have been observed at auditory levels of perception (Lotto & Kluender,

1998; Delgutte, 1997). Nevertheless, some context-sensitivities seem to arise at speech-specific processing levels (see Fowler, Brown, & Mann, 2000), which we will coin here phonetic context effects. Finally, Gaskell and Marslen-Wilson (1996, 1998) proposed that some context sensitivities in speech perception arise at a phonological processing level. Therefore, we will examine in how far speech perception in dyslexic and normal-reading children is susceptible to influences from the auditory, phonetic, and phonological contexts in Experiments 1 to 3.

Experiment 1

In this first experiment, we compare to which extent dyslexic and normal-reading children use acoustic context information in their identification of a speech sound. Hereby, we fall back on the framework of *spectral contrast* (see Holt & Lotto, 2002; Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997). According to this framework, the perceptual system takes into account the acoustic parameters of the context when evaluating a speech cue. A prototypical example is the contrast between an alveolar stop (/d/) and a velar stop (/g/). In prevocalic position, the main cue for this contrast is the onset of the third formant (F3), which is higher in a /da/ syllable than in a /ga/ syllable. According to the contrast framework, a preceding pure tone should influence the perception of a syllable with an ambiguous 'medium' F3 onset contrastively: If the preceding tone is lower than the F3 onset, the F3 onset is contrastively 'high' in comparison to the tone, and hence should be perceived as /da/, which is characterized by a high F3 onset. Vice versa, the same syllable should be perceived as /ga/ if the preceding tone is higher than the F3 onset of the syllable. Lotto and Kluender (1998) showed that this was the case.

It should be noted here that such spectral-contrast enhancement would be functional in speech perception. Due to coarticulation, contrasts between adjacent speech sounds diminish. This is especially the case if two adjacent sounds require quite different positions of the articulators (e.g., tongue tip, jaws, etc. leading to non-canonical productions of the speech sounds involved). A mechanism enhancing spectral contrasts would then in effect increase the—in production diminished—contrast between two adjacent segments. However, in the case of two speech sounds, which require similar positions for the articulators, the speech sound will be produced in a more or less canonical manner. Because two speech sounds require similar positions of the articulators, no spectral contrast arises. Therefore, a mechanism enhancing spectral contrast will 'correct' uncanonical productions of speech sounds while leaving canonical productions unaffected. This leads, in effect, to a more stable representation of the speech sounds over different instances.

In order to investigate the use of acoustic context by dyslexic and normal-reading children, we adopted the paradigm of Lotto & Kluender (1998, Experiment 4) and investigated whether dyslexic and normal-reading children differ in the way they use auditory context when identifying speech sounds. Given that Dutch does not have a velar voiced plosive, we used a continuum ranging from the unvoiced velar to an unvoiced alveolar plosive (/t/ vs/ /k/). In contrast to the Lotto and Kluender study, we used words instead of nonsense syllables in order to make the task requirements more transparent for the young participants. In addition, we used semi-natural stimuli. With this design, we make two predictions. First, we expect that the dyslexics will have shallower identification functions than normal-reading children. Second, we expect that auditory context will influence speech-sound identification more strongly in normal-reading children than in dyslexic children.

Method

Participants

42 second- and third-grade children between the ages of 7 years and 3 months and 9 years and 10 months participated in the study. 14 participants were recruited from the Regional Institute for Dyslexia (RID) in Arnhem and Maastricht, The Netherlands. All had undergone an extensive cognitive diagnosis including two reading tests, two phonological tests (a standardized auditory synthesis test from the Taaltest voor Kinderen [language test for children], van Bon, 1982, and an unstandardized phoneme-deletion task devised by the RID). These children were selected on the following criteria: Performance had to be at least one standard deviation below the age-appropriate mean on two standardized reading tests (the Een-minuut-leestest [one-minute-reading test], Brus & Voeten, 1972, and the Klepel, a pseudoword reading test devised by van den Bos, Spelbert, Scheepstra, & de Vries, 1994). Moreover, performance had to be within one standard deviation of the age-appropriate mean on the Wechsler Intelligence Test (WISC-R) and a standardized visual-form perception test. In addition, all children had to pass a hearing test at 0.5, 1, 2, and 4 kHz at 20 dB and to perform within 15 dB of the norm level in a speech audiogram in order to be included in the study.

28 normal-reading children were tested on the two reading tests, two phonological tests (a standardized auditory synthesis test from the Taaltest voor kinderen [language test voor kinderen]), and an unstandardized phoneme-deletion task devised by the RID). Control subjects performed the same hearing tests as the dyslexic children. Five children showed evidence of small hearing problems, and an additional 5 children were excluded because they performed more than one standard deviation below the age-appropriate mean on one of the reading tests. The descriptive data for the two groups are shown in Table 9.1.]

Design

The experiment consisted of two tasks. First, we assessed whether participants were able to discriminate between the endpoints of a speech-sound continuum using a two-alternative forced-choice (2AFC) discrimination task with feedback. If participants made more than four errors on the first 20 stimuli, another 20 stimuli were presented with a maximum of 60 training stimuli. Then, subjects listened to the whole continuum of seven speech sounds and were asked to identify the speech sounds as either [tart] 'cake' or [kart] 'card'. The speech sounds were first presented in isolation, and then preceded by a high sine-wave or a low sine-wave tone. This leads to a three-by-seven factorial structure.

Stimuli

A female natural speaker of standard Dutch was recorded on a DAT-Tape (Sony PCM-R500) using a Sennheiser K6 table microphone. The speaker was instructed to say /tart/ [cake] and /kart/ [card]. The recordings were digitized with a computer with a sampling frequency of 44.2 kHz, filtered at 6.5 kHz and resampled to 22.1 kHz. A /tart/ utterance was used as a template for the construction of the continuum. The formants of the utterance were estimated using Linear Predictive Coding (LPC) analysis as implemented on the software package PRAAT 3.8 (Boersma & Weenink, 1999) using 14 prediction coefficients. The LPC coefficients were used to estimate the glottal source by inverse filtering of the original sound.

Table 9.1 Descriptive data for dyslexic and normal-reading group

	dyslexic children	normal-reading children	t
Age (yr.month)	9.1	8.6	
WISC-R	105 (9.2)	—	
Reading level words ^a	6.0 (2.7)	11.6 (2.6)	5.9
Reading level nonwords ^a	6.6 (2.8)	12.7 (2.6)	6.1
Phoneme deletion ^b	17.2 (6.4)	24.9 (4.09)	3.9
Auditive synthesis ^a	11.0 (2.9)	10.9 (3.7)	0.1

Note: a) standardized scores, $m = 11$ $sd = 2$, b) number of correct items (maximal 28).

As the estimated source did not contain any aperiodic noise, broad-band noise was added to the source for the first 25 ms of the source, emulating a 25 ms VOT observed in the original utterance. The LPC-based formants were edited: A continuum of seven stimuli was created by using seven different formant transitions. In the original utterance, the third formant started at 3.4 kHz, six variants were created with a step size of 0.330 bark for the start of the formant transitions, preserving the non-linearity of the original formant transition. This led to F3 onsets at 3.40 kHz, 3.23 kHz, 3.07 kHz, 2.92 kHz, 2.77 kHz, 2.63 kHz, 2.50 kHz for the seven filters. With these edited formant filter and the estimated source, seven semi-natural speech sounds were generated.

The sine-wave tones used as context sounds had frequencies of 2.58 and 3.15 kHz, respectively just above the lowest F3 onset and just below the highest F3 onset. The sine-wave sounds had a duration of 125 ms and were clipped linearly for the first and last 25 ms in order to prevent clicks. There was a 25 ms gap of silence between the tone and the speech sound.

Apparatus

The experiments were done with a mobile lab. Stimulus presentation was controlled from a laptop by a DELPHI 5.0 (Bohrlund, 1999) program. The sound output of the computer was amplified (Radio Design Labs, ST-PH 1) and played to the participants via a headphone (SONY MDR-V900). Experiments were done in a quiet room with only the experimenter and the participant present.

Procedure

All children were first acquainted with the 2AFC procedure using the salient pair 'huis'-'muis' [house-mouse]. Two pictures representing the two words were displayed on the computer screen on the upper left and upper right corner. Participants were instructed that, upon hearing a word, they should indicate which word they heard by pressing an upper left ('q') or an upper right ('p') key of the computer's keyboard. (A custom-made cover left only these two keys available for pressing.) After each reaction, a 'smiley' face indicated whether the choice was correct. None of the children displayed any problems in understanding the 2AFC task. After this, children performed a 2AFC task with feedback with the two endpoints of the semi-natural [tart]-[kart] continuum. Two pictures on the computer screen indicated which key was associated with which word. If more than four errors were made on the first 20 trials, another 20 stimuli were presented. After 60 stimuli, this training phase was terminated. Then all seven semi-natural stimuli were judged in a block without any precursor sound. In this and the following phase the 'smiley' face only 'looked' in the direction of the picture corresponding to the child's response. This procedure does not provide performance-related feedback but achieved some interactivity of the experimental task. Each stimulus was presented 10 times in a random order. Then, the semi-natural stimuli were presented with one of the two precursor tones. Each of the 14 (seven words by two precursor tones) stimuli was presented 15 times in a random order. After every 70 trials, there was a short break. The experiment lasted about 20 minutes.

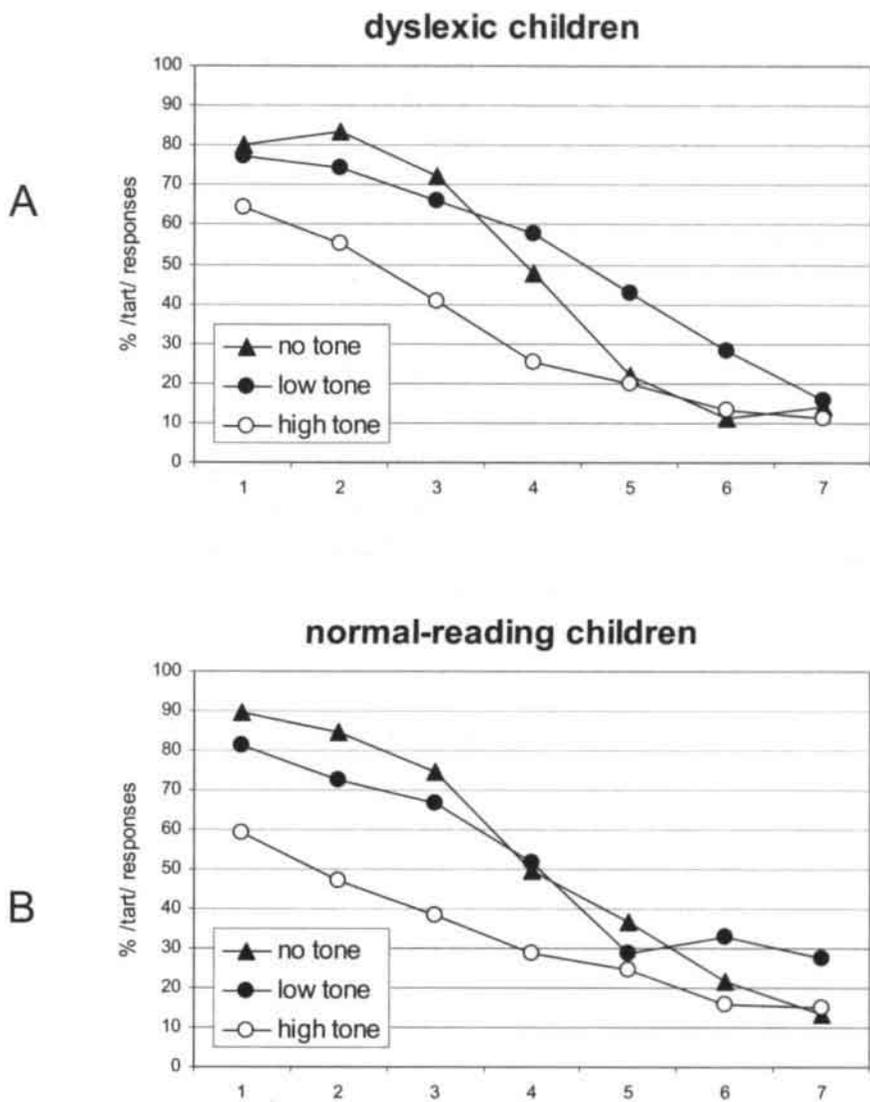


Figure 9.2. Mean percent of /tart/ responses in Experiment 1 as a function of acoustic context and stimulus continuum. Figure 2A shows the data of the dyslectic, Figure 2B the data of the normal-reading group. The ordinate co-ordinate represents the seven-step /tart/-/kart/ continuum. In both groups, there is a clear acoustic context effect

Results

The individual data were transformed into percent /t/-responses (see Figure 9.2). These data were subjected to an analysis of variance (ANOVA) with two within-subject variables (Context: none, high tone, and low tone; Continuum: the seven semi-natural stimuli) and Group (dyslexic and normal-reading) as between-subject variable. The ANOVA revealed a significant effect of the Continuum ($F(3, 75) = 161.5, p < 0.001$), and a main effect of the Context ($F(2, 48) = 21.9, p < 0.001$). Post-hoc Bonferoni tests ($p < 0.05$) showed that participants gave overall less /t/-responses in the high-tone condition (32.8%) than in the low-tone (53.2%) and in the no-context condition (50.4%), while these latter two conditions did not differ significantly. There also was a significant interaction of the two within-subjects factors ($F(6, 182) = 7.8, p < 0.001$). To investigate this interaction, seven ANOVAs were performed comparing the three context conditions at all levels of the continuum with each other. While Context influenced the responses at all levels, post-hoc Bonferoni tests showed that the high-tone condition differed from the two other conditions on the first four levels of the continuum (the '/t/-end'). The low-tone condition differed from the other two conditions at the last 3 levels of the continuum (the '/k/-end'), with the exception of the comparison between low-tone and no-context at the seventh level (the most /k/-like stimulus). However, neither the between-subjects factor ($F < 1$) nor its interaction with any of the other factors was significant (continuum-by-group $F(3, 75) = 1.14, p = 0.332$, all other $F_s < 1$).

The absence of an effect of Group might possibly be an artefact of averaging over participants. Therefore, we fitted a logistic function with two free parameters ($1/[1+\exp(-ax + b)]$) to the three individual identification functions in the three different context conditions (see Table 9.2).

Table 9.2 Slope parameters for normal-reading and dyslexic group in Experiment 1

Group	Context			mean
	no tone	low tone	high tone	
Dyslexic	4.033	2.314	1.905	2.751
Normal-reading	3.272	1.717	2.020	2.130

These estimations were used in an ANOVA with the between-subject variable Group and Context as within-subject variable. Context influenced the slope parameter ($F(2, 48) = 15.0$, $p < 0.001$). Post-hoc tests showed that the slope in the no-context condition ($m = 3.65$) was larger than in the high-tone context ($m = 2.01$) and in the low-tone context ($m = 1.96$), while these two latter conditions did not differ significantly. The between-subjects factor did not have a significant effect nor interacted with the Context factor ($F_s < 1$). It is important to note that the dyslexic participants did have a larger mean slope value in the no-context condition. Therefore, the fact that we did not find a 'speech-perception deficit' cannot be attributed to lack of statistical power.

Discussion

The experiment revealed the expected main effects: Changing the F3 onset influenced the perception of the initial stop as either /t/ of /k/. Second, the precursor sounds influenced the perception of the speech sounds in line with the frequency-contrast model. Participants were more likely to give a /t/ response, associated with a high F3 onset, after a low precursor tone than after a high precursor tone. Therefore, the experimental stimuli and procedure seem adequate. However, both groups performed similar in this experiment. Neither could we find any evidence for a 'speech-perception deficit,' nor did we find that the dyslexic group made less use of the context. One possibility why we did not find a 'speech-perception deficit' might be that the formant transition of the stop was too long (80 ms). Tallal (1980) claimed that dyslexics' speech-perception problems are restricted to brief transitions, shorter than 50 ms. There is, however, evidence showing that speech-perception problems can be found with long-lasting stimulus differences. de Gelder and Vroomen (1998) found a speech-perception deficit using a /ba-/da/ continuum with a transition time of 80 ms, equal to the transition time used here (see also Heim et al., 1999). Masterson, Hazan, and Wijazatilake (1995) report problems with steady-state fricatives and Post, Foorman, and Hiscock (1997) even reported problems in vowel perception. Therefore, the relatively long transition time in the formant-transition times does not seem to be responsible for the present failure to find a speech-perception deficit in the dyslexic group. One possibility to explain the differences between our results and previous reports of impaired speech perception may be related to the stimulus material. All studies mentioned in the McBride-Chang (1995) review relied on completely synthetic speech. Therefore, it is conceivable that a speech-perception deficit in dyslexia may be restricted to synthetic speech. In this experiment, however, semi-natural speech was used.

The other result contradicting our initial hypotheses was that dyslexic and normal-reading children did not differ in their use of acoustic context information. However, this does

not rule out our initial hypothesis that dyslexics may make less use of context in the identification of speech sounds. It is conceivable that the use of the acoustic context may not be impaired, but that the use of speech-specific context on higher levels of processing may be impaired. Such an argument is buttressed by the findings of Fowler et al. (2000). They showed that speech-specific context effects may dissociate from auditory context effects. Fowler et al. also argued that spectral contrast effects result from forward masking. If auditory context effects are indeed based on different mechanisms than more speech-specific context effects, then it is possible that dyslexic and normal-reading children differ in their use of speech-specific context. This was tested in Experiment 2.

Experiment 2

In this experiment, we investigated the influence the phonetic context information on speech sound identification in dyslexic and normal-reading children. Here we make use of the 'compensation-for-coarticulation' paradigm, which is one of the corner stones of the direct-perception theory for speech perception (Fowler, 1996). As in experiment 1, the target sounds were drawn from a continuum ranging from /ta/ to /ka/. As context sounds, we used the syllables /al/ and /ar/. Mann (1980) showed that such pre-cursor syllables influence the perception of the velar-alveolar distinction in stops. Participants are more likely to hear a speech sound as /ta/ after /ar/ than after /al/. This may be explained in terms of the direct-perception theory: Participants should find the interpretation of an ambiguous stop as /t/ more likely after /ar/ than after /al/, because /r/ requires a tongue position very different from the prototypical tongue position for a /t/. In contrast, a /k/ requires a similar tongue position as /r/. Therefore, it is more likely that the ambiguous stop after /ar/ is an uncanonical version of /t/ than an uncanonical version of /k/. Vice versa, the same ambiguous stop is interpreted as a /k/ after /al/, because an /lt/ sequence affords less movement from the articulators than an /lk/ sequence.

Similarly to the auditory context effects, this compensation for coarticulation is functional in speech perception, as it compensates for context-dependent variance that arose in speech production. This, in turn, allows a more narrowly defined representation of the speech sounds in question. Most importantly for current purposes, it seems that such phonetic context effects are distinct from auditory context effects (Fowler et al., 2000). Therefore, we may expect that dyslexics and normal-reading children differ in their use of phonetic context information, although they did not differ in the use of an auditory context.

Method

Participants and Apparatus

Participants and apparatus were the same as in Experiment 1.

Design

The experiment consisted of two tasks. First, we briefly acquainted the participants with the stimuli. In a 2AFC task with feedback, the children had to discriminate the endpoints of the continuum preceded by one of the two precursor syllables. After twenty trials, the second part started. In this part, participants had to categorize all stimuli of the continuum preceded by one of the precursor syllables. No feedback was given. This leads to a two (/ar/ vs. /al/ as precursor syllable)-by-seven (continuum) factorial structure.

Stimuli

The same /tart/-/kart/ continuum was used as in Experiment 1. As the speaker was not available for further recordings, the precursor syllables were synthesized by applying different filters to the source of the vocal portion of the /tart/ utterance used for the continuum. Therefore, the precursor syllables and the stimuli of the continuum sounded as coming from the same speaker in a single utterance. The precursor syllables had a duration of 300 ms. The filters for both syllables were identical for the first 170 ms with $F1 = 900$, $F2 = 1500$, $F3 = 3200$ and $F4 = 4900$. At $t = 170$ ms, the filter settings started to differ between the /al/ and /ar/ syllable. For the /al/ syllable, the formants changed linearly to reach $F1 = 630$, $F2 = 1020$, $F3 = 3740$ at $t = 0.3$ s. For the /ar/ syllable, the formants changed linearly to reach $F1 = 585$, $F2 = 2060$, and $F3 = 2360$ at $t = 0.3$ s. These formant transitions indicate a movement of the tongue to the front of the mouth in the case of the /al/ syllable and a movement to the back of the mouth for the /ar/ syllable (cf, e.g., Ladefoged, 1997). The inter-stimulus interval (ISI) was adjusted to achieve a natural sounding bi-syllable with an appropriate closure time, which was given at 50 ms ISI.

Procedure

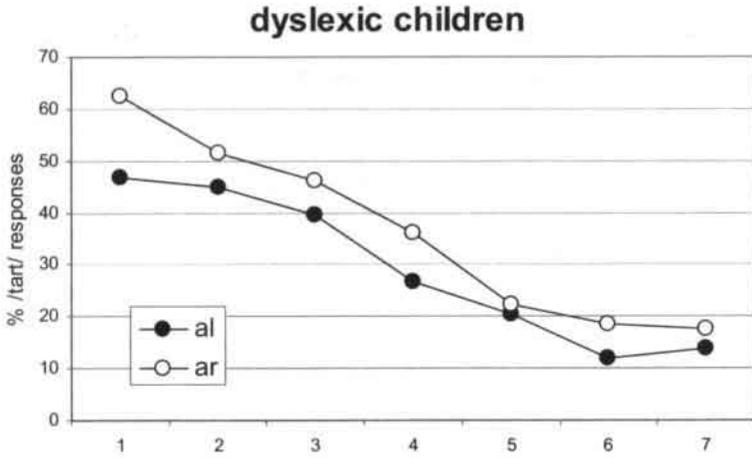
This experiment was performed in a second session at least one week after Experiment 1. All children were therefore familiar with the display and the mode of reacting. The experiment started with a 20-trial training part. In this part, participant had to classify the endpoints of the /tart/-/kart/ continuum, which were preceded by one of the precursor syllables. A 'smiley face' indicated whether the classification was correct. In the experimental part, all stimuli of the continuum were used preceded by one of the precursor syllables. In this part, the 'smiley face' looked into the direction of the picture corresponding to the response of the child. This was done in order to achieve some interactivity of the experimental task. Each of the 14 experimental stimuli was presented 15 times in a randomized order. After this part, subjects did a 20-trial post-test on the identification of the precursor syllables as /al/ or /ar/ in a 2AFC task with feedback. The whole experimental session lasted about 15 min.

Results

The post-test data showed that all participants perceived the precursor syllables as intended. The mean percentage correct was 94.4% (dyslexic: 93.6%, normal-reading: 95%), which is equivalent to one error on the twenty-trial test. 24 of the 32 participants did not make more than one error; the maximum number of errors was four. It can be concluded, therefore, that the precursor syllables were perceived as intended as /al/ and /ar/.

The data from the experimental session were transformed to percent-/tart/ responses for each individual in each condition. These data (see Figure 9.3) were subjected to a repeated-measure ANOVA with Context (/al/ vs. /ar/) and Continuum (the seven stimuli of the continuum) as within-subjects factors and Group (dyslexic vs. normal-reading) as between-subjects factor.

A



B

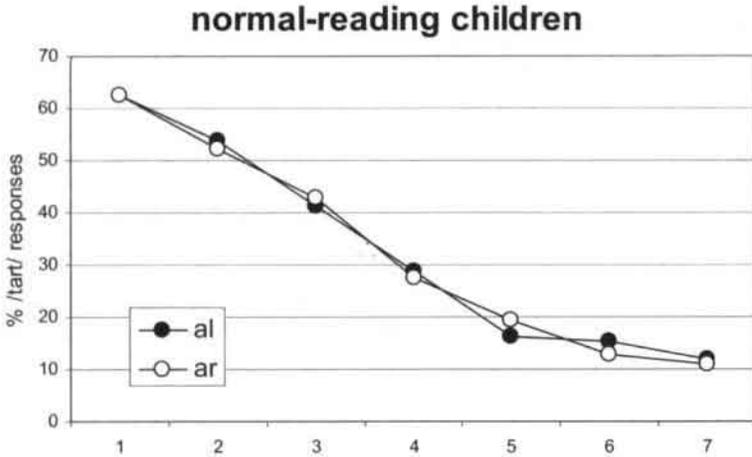


Figure 9.3 Mean percent of /tart/ responses in Experiment 1 as a function of phonetic context and stimulus continuum. The ordinate co-ordinate represents the even-step /tart/-/kart/ continuum. A phonetic context effect is only evident in the dyslexic group.

Table 9.3 Percent correct responses in the training part of Experiment 2

Group	Context			
	/al/		/ar/	
	Stop		Stop	
	/t/	/k/	/t/	/k/
Dyslexic	52.9	90.0	68.7	82.9
Normal-reading	68.8	98.8	74.1	83.5

Both within-subjects factors were significant (Context: $F(1, 30) = 5.5, p < 0.05$, Continuum: $F(2, 60) = 41.0, p < 0.001$), while their interaction was not significant ($F < 1$). Moreover, the interaction between Context and the Group factor was significant ($F(1, 60) = 5.7, p < 0.025$). To investigate the nature of this interaction, we tested the effect of Context in both groups separately. Context did affect responses in the dyslexic group ($F(1, 13) = 8.4, p < 0.025$), but not in the normal-reading group ($F < 1$). In the dyslexic group, the context effect was in the expected direction. After /al/ there were more /kart/ responses than after /ar/, indicating compensation for coarticulation. The between-subject factor and the three-way interaction were not significant ($F_s < 1$).

Given the absence of a context effect in the normal-reading group, we also analyzed the training data for context effects. Table 9.3 shows the mean percentage of correct responses for the four stimuli. An ANOVA shows that all participants generally made more errors for /tart/ (66.0% correct) than for /kart/ (88.9% correct, $F(1, 30) = 28.1, p < 0.001$), while the precursor phoneme did not have an influence ($F < 1$). Most important, the interaction between target and context was significant ($F(1, 30) = 8.8, p < 0.01$). As Table 9.3 shows, /tart/ was more often recognized correctly in the /ar/ context than in the /al/ context, while /kart/ was more often recognized correctly in the /al/ context than in the /ar/ context. This is in line with compensation for coarticulation. An /al/ biases perception towards /kart/, while an /ar/ biases towards /tart/. These effects did not interact with the Group factor ($F_s < 1$).

Table 9.4 Slope parameters for normal-reading and dyslexic group in Experiment 2.

Group	Context		mean
	/al/	/ar/	
Dyslexic	3.089	1.870	2.480
Normal-reading	2.239	2.876	2.557

In addition, we analyzed the individual classification functions by fitting a logistic curve (see Table 9.4). The individual slope parameters were then analyzed with an ANOVA using Context and Group as factors. Both factors did not influence the slopes ($F_s < 1$), nor did the interaction reach significance ($F(1, 30) = 2.36, p < 0.138$). It is noteworthy, however, that the mean slope in this experiment is rather small, although this is only a trend compared to the no-context conditions in Experiment 1 ($t(31) = 1.8, p < 0.1$). In addition, participants were overall more likely to respond /kart/ in both context condition of this experiment than in the no-context condition of Experiment 1 ($t(31) = 5.3, p < 0.001$).

Discussion

As in the previous experiment, dyslexics did not show shallower identification functions in a speech-sound classification task. That is, we again fail to find evidence to support the hypothesis that dyslexia is associated with a speech-perception deficit. However, dyslexic and control children differed in the way phonetic context information was used. Contrary to our expectations, Context influenced identifications more strongly in the dyslexic than in the control group. This result parallels a report by Nittrouer (1999). She found that dyslexic children used the range of acoustic cues for a given phonemic distinction differently than normal-reading children. In a 2AFC task with two words differing in the place of articulation of an initial fricative, the contribution of different cues—spectral composition of the fricative noise and formant transitions to the vowel—to place perception were tested. It turned out that dyslexic children placed a heavier weight on the context-dependent cue, the formant transition, than on the local cue, the spectral composition of the noise, while normal-reading children showed the opposite pattern. Similarly, the control group in the present experiment did put less weight on a contextual cue than the dyslexic group. The pattern of the dyslexic group here resembles the results that have been obtained with pre-school children, as described by Nittrouer (e.g., 1992) in her developmental-weighting-shift model. This may indicate that dyslexic children fail to develop optimal speech-recognition routines.

Most importantly, however, the present results fail to support our initial hypothesis about context effects in speech perception. Similar as in the Experiment 1, the dyslexic group used of the context information in speech-sound classification a way that is appropriate for compensation for coarticulation.

Experiment 3

Up to this point, we investigated the use of acoustic and phonetic context information in the identification of speech sounds. The results did not show the expected effect of less context sensitivity in the dyslexic group. Such an effect might nevertheless be found when a context effect is examined that arises at a higher level of processing. An example of such a high-level context sensitivity is the 'phonological-inference' mechanism proposed by Gaskell and Marslen-Wilson (1996, 1998). This mechanism is supposed to aid the recognition of words that have undergone phonological assimilation. Phonological assimilations are similar to coarticulation in that the acoustic manifestation of a phoneme is influenced by adjacent phonemes. In contrast to coarticulation, this influence is so strong that it blurs phonological contrasts—at least superficially. To give an example of a phonological assimilation, consider the utterance 'garden bench.' The /n/ in 'garden' may be assimilated by the /b/ in bench in order to lead to the utterance 'gardembench.' This change of the final nasal in 'garden' to become an /m/ occurs only if the next segment is a labial obstruent. That is, the /n/ does not change in the utterance 'garden chair.' Hence, *'gardemchair' is a forbidden form.

According to the model of phonological inference, the perceptual system evaluates assimilated word forms by using the phonological context of the assimilated segment. Hence, in perception, assimilated forms as 'gardem' are only accepted as instances of the respective canonical form if the phonological context allows the assimilation. Therefore, 'gardem' is accepted as an instance of 'garden' in 'gardembench' but not in *'gardemchair.' Note that phonological inference is thus an instance of a context effect. The context determines whether the assimilated form is recognized as an instance of a canonical form.

This context effect is conceptually different from the compensation for coarticulation effect that was the subject matter of the previous experiment. In 'compensation for coarticulation,' it is the phonetic gesture of the context sounds that drives the context sensitivity. In the model of phonological inference, it is not the phonetic gesture, but the abstract phonological features, which are crucial. Only if the phonological feature values of the context are such that the assimilation is allowed, the assimilation is compensated for. Given these conceptual differences, we may expect the phonological context effects caused by phonological

inference to dissociate from acoustic and phonetic context effects. Therefore, we tested whether this phonological context effect was different in dyslexic and normal-reading children.

To this end, we applied a method of testing phonological inference established in the third chapter of this dissertation. They showed that the context sensitivity in the perception of phonological assimilation could be probed with a 2AFC task. The results showed that, if participants were asked to indicate whether they hear the Dutch word 'tuin' [garden] pronounced properly or wrongly with a final /m/, they are well able to do so. Performance is also near ceiling level if the composite 'tuinstoel' [garden chair] is used. However, the results were different when the composite 'tuinbank' [garden chair] was used. In this case, the pronunciation with /m/ might be the consequence of the application of place assimilation. Indeed, it turned out that participants made a significant number of errors by misperceiving the stimulus 'tuinbank' as 'tuinbank.' That is, they compensated for the possible assimilation. The change from /n/ to /m/ in production is countered by a change from /m/ to /n/ in perception. As shown in chapter three, this effect cannot be attributed to a lexical top-down effect or to decision-making processes. Rather, the effect seems to be pre-lexical. Here, we used the same 2AFC paradigm to compare phonological context sensitivity in dyslexic and normal-reading children.

Method

Participants and apparatus

The participants were the same as in the previous experiments. The experiment was conducted in the same session as Experiment 2. The apparatus was the same as in the previous experiments.

Design

The design entailed three factors, two within-subjects factors and a between-subjects factor, that is, Group (dyslexic vs. control). The first within-subject factor was the identity of the word-final nasal of the target word 'tui...', which could be /n/ or /m/. The second within-subject factor was the context following the target word. There could be no context, a viable context ('bank'), in which assimilation ('tuinbank') may occur in natural speech, and an unviable context ('stoel') in which assimilation is not allowed to occur. This leads to a two-by-two-by-three design with the factors Group, Nasal murmur (/n/ vs. /m/), and Context (none, viable, that is, 'bank,' and unviable, that is, 'stoel').

Stimulus material

A male native speaker of Dutch was recorded uttering 'tuinbank' [tœynbank], 'tuinbank' [tœymbank], 'tuinstoel' [tœynstu:l], and 'tuimstoel' [tœynstu:l] several times. The speaker was chosen because of his low f₀ (75–80 Hz), which did not vary greatly between utterances. Therefore, cross-spliced utterances sounded natural. The context words 'stoel' [stu:l] and 'bank' [bank] were spliced from two other utterances. By cutting 5 ms of the friction noise of the /s/ in [stu:l], the length of this sound was made equivalent to the length of the [bank] sound. This includes a silent period of 25 ms before the onset of the /b/, an appropriate closure duration in natural utterances of [tœynbank]. The stimuli differed slightly in length (< 1ms) due to the constraint of splicing at zero-crossings. Then, the onset and nucleus of [tœyn] were spliced from the nasal murmur, using the lowest amplitude between the vowel and nasal part as cutting point. This 'tui' [tœy] utterance was then concatenated with a /n/ and /m/ nasal murmur spliced from two other tokens, resulting in tokens of [tœyn] and [tœym] only differing in nasal

murmur. These tokens were then concatenated with 'stoel' [stu:l] and 'bank' [baŋk], leading to three [tæyn] and [tæym] stimulus pairs: one without context, one with a 'stoel' [stu:l], and one with a 'bank' [baŋk] context.

Procedure

Participants looked at the computer screen displaying a large 'n' in the upper left corner and a large 'm' in the upper right corner. They were instructed to press the upper left key of the computer keyboard ('q') upon hearing 'tɔin' and the upper right key ('p') upon hearing 'tuim'. A custom made shell left only these two keys available for pressing. Participants first went through a short training phase with explicit feedback given by a 'smiley face' on the computer screen in order to clarify any problems with the task instruction. However, all children understood the nature of the task without any problems.

In a second phase, no explicit feedback was given; instead the 'smiley face' looked in the direction of the picture associated with the reaction. All participants started with the no-context condition ([tæyn] 'tuin' and [tæym] 'tuim'). After presenting the block with the targets in isolation, the order of presentation of the viable-context block ('tuinbank' vs. 'tuimbank') and unviable-context block ('tuinstoel' vs. 'tuimstoel') was counterbalanced. One block consisted of 50 stimuli presented in a randomized order. The experiment lasted about 10 minutes.

Results

Table 9.5 shows the average results of both groups coded as percentage correct responses. An ANOVA over these data revealed significant main effects of Nasal murmur ($F(1, 30) = 11.7, p < 0.005$) and Context ($F(3, 43) = 100.3, p < 0.001$). The main effects were, however, qualified by a significant interaction ($F(1, 42) = 23.3, p < 0.001$). This indicates 'phonological inference.' In the viable-context condition, participants show a bias towards (mis)perceiving 'tɔinbank' as 'tuinbank.' Neither the main effect for Group ($F < 1$) nor any of the interactions with a within-subject variable attained significance (Nasal murmur by Group: $F(1, 30) = 2.7, p > 0.1$, all other F s < 1).

Table 9.5 Percentage of correct responses in Experiment 3

Group	Context					
	none		unviable: /stoel/		viable: /bank/	
	Final nasal		Final nasal		Final nasal	
	n	m	n	m	n	m
Dyslexic	81.7	78.3	79.7	77.1	68.9	34.3
Normal-reading	81.6	84.6	75.3	81.1	67.8	44.7

Discussion

The results replicate the patterns observed with adult participants (cf. chapter three). The identification task is more difficult in the viable context than in the unviable context and without context. In addition, there is a bias towards responding with 'tuinbank,' the canonical form, in the viable condition. However, the Group factor did not have any measurable influence on the reactions, just as in the previous experiments. The dyslexics did not make more errors in classifying the two nasal murmurs. Note that this is not due to a ceiling effect. Actually, most participants found the task not easy, which is reflected in just over 80% correct performance in the no-context condition. In addition, there was a similar bias towards the canonical form 'tuinbank' in both the dyslexic and the control group.

The present failure to find a dissociation of phonetic and phonological context effects in the Experiment two and three may also have some repercussions for theories about context effects in speech perception. The present results might be taken to indicate that the phonological-inference mechanism probed here is in fact not a separate mechanism at a phonological processing stage. This fits with recent proposals that the compensation for assimilation, which was assumed to occur at a phonological level, may indeed occur at auditory/phonetic processing stages (Gow, 2002a; chapters three, five, six, seven, and eight of this dissertation). If that is the case, the result of Experiment 3, testing phonological inference, may be expected to be similar to the results of Experiment 1 and 2, testing auditory and phonetic context effects.

General Discussion

A number of studies suggested that dyslexics suffer from a subtle 'speech-perception deficit,' which becomes evident in categorical-perception tasks (Godfrey et al., 1981; Hurford & Sanders, 1990; Reed, 1989). Recent results (Mody et al., 1997; Nittrouer, 1999; Rosen & Manganuari, 2001) indicated that the sensitivity to speech cues is not impaired in dyslexics. Therefore, the nature of the 'speech-perception deficit' remains to be explored. We investigated whether the 'speech-perception deficit' displayed by dyslexic children could be a consequence of insufficient context sensitivity in speech perception. Although it is not immediately apparent how insufficient context sensitivity could lead to a 'speech-perception deficit,' a mathematical model, simulating the effects of a 'context sensitivity deficit' on the performance in categorical-perception task, showed that this is the case (see Figure 9.1). Therefore, we performed three experiments probing acoustic, phonetic, and phonological context sensitivity (respectively

Experiment 1 to 3) in dyslexic and control children. The results showed no evidence for either a speech-perception deficit or insufficient compensation for context-dependent variation by dyslexic children in any of the three experiments. Nevertheless, one group difference was found in Experiment 2, pertaining to a phonetically based context effect. In this experiment, dyslexics were influenced more strongly by the context than their normal-reading peers. This is in line with a result obtained by Nittrouer (1999). In her view, dyslexics do not integrate speech cues properly in order to achieve phonological categorization. According to such an interpretation, dyslexics may weigh the coarticulatory cue in the context more strongly than the cues in the segment itself.

It should be noted that the present null results seem rather informative. As noted by Mody et al. (1997), the nature of the speech-perception deficit in dyslexics is not yet explored. As the hypothesis that dyslexics suffer from insufficient sensitivity to basic acoustic cues does not well in explaining recent data, we explored the possibility of insufficient context sensitivity and obtained 'conclusive' null results. The data show that the dyslexic group did not show less context sensitivity than the normal-reading control group. Therefore, the conclusion seems warranted that a speech-perception deficit in dyslexia cannot be explained by insufficient compensation for variance in the speech signal. Dyslexics adequately use acoustic, phonetic, and phonological context information when identifying a given speech sound.

In this context, it is important to note that McQueen and Cutler (2001) argued that normalization processes in speech perception indicate that speech perception uses some form of intermediate representations. By filtering variance out of the input signal, normalization processes yield a to some degree abstract code that is instrumental in lexical access. While many different versions of this abstract code are possible (phonological features, phonemes, demi-syllables, etc.), these abstract representations are much closer to a representation of phonemes than the raw acoustic input. Hence, the present results indicate that dyslexics might in principle be able to develop non-impaired intermediate representations based on adequate normalization processes. Although there might be other possibilities why dyslexics may not develop adequate phonological representations—for instance, poor filtering of background noise—it is also worthwhile to consider the possibility that it is the *phonological processing* of phonological information to achieve lexical access, and not the *phonological representations* per se, which is impaired in dyslexics. If some sub-process of phonological processing is delayed, this may just be as devastating for the global phonological processing as inadequate phonological representations.

A second finding of the current study is that—in contrast to a large number of reports of less categorical speech perception in dyslexia—we did not obtain such a result. The dyslexic

group did on average perceive the continuum in Experiment 1 *more* categorically than the normal-reading group. Although this mean difference is far from significant, it rules out the possibility that the failure to obtain less categorical speech perception is due to the lack of statistical power. Therefore, the question arises how the apparent discrepancy between the present and previous results can be explained. As already mentioned, one difference between the current study and previous studies is the fact that the majority of studies have employed synthetic speech. It is conceivable that the alleged speech-perception deficit associated with dyslexia is not as pervasive as previously assumed. The difference of results might be associated with the familiarity with the stimulus material. The semi-natural speech used here sound is of a quality that is within the range of speech stimuli that is frequently encountered in everyday live (e.g., in telephone conversations). This cannot be said for synthetic speech stimuli, especially if they do not incorporated the typical distortions of natural voice sources (see Klatt & Klatt, 1990). This difference in stimulus familiarity might account for the account for the difference in results between our and, for instance, the studies by Reed (1989) or Hurford and Sanders (1990).

In summary, the present study explored two aspects of speech perception in dyslexics. First, we tested the acuity of speech perception and, secondly, the ability to use acoustic, phonetic, and phonological contexts in speech-sound identification. Using natural speech, dyslexics did not show any speech-perception deficit, replicating Blomert and Mitterer (in press). Moreover, dyslexics showed an equivalent amount of context sensitivity to acoustic and phonological contexts. In addition, they used the phonetic context to a greater extent than normal-reading controls. This seems in line with a finding by Nittrouer (1999), showing that dyslexics weigh contextually-dependent cues more strongly than local, more or less context-independent, cues. Most importantly, however, the current study indicates that the quality of the phonemic representations used in speech perception may be similar in dyslexic and normal-reading children, because the context compensation processes giving rise to phonemic representations seem to function properly. Although there are other possibilities that may lead inadequate phonemic representations, it seems also worthwhile to consider the possibility that the phonological core deficit in developmental dyslexia may be caused by a phonological-processing deficit and not by a phonological-representation deficit.

Epilogue

This dissertation examined how listeners are able to understand spoken words even if they have been changed by a phonological assimilation. The question stated was: How can the input [tœym] be mapped on the lexical representation /tœyn/? Chapters two and three evaluated the two rather general and opposite proposals: “context sensitivity” versus “underspecification.” The gist of these two chapters is easy to summarize: We did not find any evidence for the assumption of phonological underspecification, but found strong evidence in favor of the context sensitivity assumption. Hence the form [tœym] is only mapped on the lexical representation /tœyn/ if it occurs in a phonological context that allows the specific /n/-to/m/ assimilation to occur.

We coined this context sensitivity ‘compensation for assimilation,’ following the lead from Mann’s (1980) compensation for coarticulation. Accepting the existence of a ‘compensation-for-assimilation’ mechanism then raises the question about the nature of this mechanism. In the introduction, we stated four questions that we were to investigate. These were:

1. Is compensation for assimilation a general phenomenon or restricted to NPA in Dutch?
2. At what kind of processing level does the compensation for assimilation arise?
3. Is compensation for assimilation acquired during language acquisition?
4. If this compensation for assimilation is an important process in speech comprehension, is it impaired in populations, which have a (subtle) speech-comprehension deficit?

It has been eight chapters ago since these questions were introduced. Against the background provided by these chapters, I will now try to derive answers to these questions.

Questions answered

The *first* question was whether compensation for assimilation is a general phenomenon or restricted to NPA in Dutch. We found evidence for a ‘compensation-for-assimilation’ mechanism in two quite different languages and with rather different assimilation rules. In addition, we found evidence for ‘compensation-for-assimilation’ with quite diverse methods as identification tasks and the MMN, actually producing a good example of the method of converging operations. So there is evidence for ‘compensation for assimilation’ in multiple languages and with multiple methods.

The *second* question aimed at localizing the processing level at which compensation occurs. The first strategy to answer this question was to investigate the role of phonetic detail.

The rationale was that, if it can be shown that phonetic detail contains information for the perception of assimilated utterances and that listeners can use such subtle information, it is unlikely that compensation for assimilation occurs at a level of abstract phonological features (as assumed by Gaskell, Hare, & Marslen-Wilson, 1995). Chapter five provided evidence affirming the role of phonetic detail in both the perception and the production of assimilated utterances. At first glance, this provides an argument against the assumption of a phonological processing level for compensation for assimilation. However, a model working on abstract phonological features may accommodate such a result. If phonological feature input is continuous instead of categorical, the strength of activation of a phonological feature may convey information about the acoustic details of an utterance. Gaskell (in press) recently constructed such a model. However, there is now evidence that the length of a segment (this dissertation, chapter five), the spectral composition of a segment (Gow, 2002a) and the precise amplitude envelope of a segment (this dissertation, chapter seven) influence the compensation process. Therefore, it may be argued that graded phonological categories may be a Procrustean bed in which these variables have to be molded to fit. This pleads for a locus of the context effects in a multi-dimensional space, in which phonetic detail is still available *as such*, rather than in a projection on a single, phonological-feature dimension.

A second strategy to evaluate the locus of the compensation process was to test whether the context effects indicating compensation for assimilation still persist if participants are asked to discriminate between assimilated and canonical forms. The rationale is that, should compensation occur at a higher level of processing, participants should well be able to discriminate between assimilated and canonical forms. If compensation occurs at a higher, language-specific processing level, participants should be able to access some auditory processing level at which no compensation has yet occurred. On this level, assimilated and canonical utterances should be clearly discriminable. Such a result was not obtained. Instead, we observed (chapter six) that participants had problems to distinguish a canonical from an assimilated form in a discrimination task. The pervasiveness of the context effect is underscored by the fact that the discrimination task used seems to tap auditory representations. Therefore, it seems inevitable to conclude that the processing of assimilated utterances is influenced by context already at auditory processing levels.

In addition, we examined whether compensation for assimilation is an exclusively prelexical process. If this was the case, compensation should occur equally well for words and nonwords. This was consistently the case (this dissertation, chapters three, six, and seven). Earlier evidence to the contrary (Gaskell & Marslen-Wilson, 1998) has to be attributed to the

task used, being the phoneme-monitoring task. This task seems to involve higher cognitive processes, as phonemes, whether or not derived implicitly before lexical access, are certainly not the objects of perception (Morais, Bertelson, Cary, & Alegria, 1986; Warren, 1999).

To sum up, our results showed that, even if compensation for assimilation might occur at a phonological level, at least this phonological level will have to allow graded representations (chapters five and seven). More strongly, though, a crucial part of the compensation already seems to arise in auditory processing (chapter six and eight).

However, the result also revealed that compensation for assimilation is certainly not a purely auditory phenomenon. This was one of the results of our efforts to elucidate whether language experience is necessary in order to compensate for assimilation, which constitutes the *third* of our four questions. The results showed that to some degree, compensation for assimilation is independent of experience with specific assimilation rules. This was evident in the results of chapter six, in which there was a context effect in the discrimination task that did not depend on experience with the assimilation rule. This was also evident in the results obtained with event-related potentials in chapter seven. Concerning the results obtained with the identification task, it was always evident that both native and non-native listeners had more problems to make a phonemic distinction in a context in which this phonemic distinction could be blurred by a phonological assimilation. This was the case for the results of Dutch and Portuguese listeners presented with Dutch utterances in chapter four; *and* it was the case for the results of the Dutch and Hungarian group presented with Hungarian utterances in chapter six. This part of the results, which is independent of experience with specific assimilation rules, can best be interpreted as a result of a context effect at an early auditory level.

However, the results of the identification task did reveal an effect of language experience (chapters four and six). It turned out that listeners, who have experience with an assimilation rule, tend to have a bias toward perceiving the phoneme class that might have been subjected to assimilation. Again, this result is consistent in both cross-linguistic comparisons. Given that this effect was evident only in the identification task but not in the discrimination task, we may assume that this effect arises at a phonetic or phonological level of processing.

Therefore, we may conclude that compensation for assimilation is not a unitary process. Compensation seems to arise, first and foremost, from context sensitivities at an early auditory level. In addition, language experience creates a bias at a language-specific level of processing. However, the fact that compensation for assimilation seems to arise partly as a consequence of processes independent of language-experience—and hence is probably innate—strikes one as counterintuitive. Because assimilation rules are to some degree language-dependent,

compensation for assimilation seems to represent a specific problem in speech perception. Solutions to such specific problems are unlikely to be innate (see chapter six). So how is it possible that there is some compensation for unknown assimilation rules in non-native languages?

There are two possible answers to this question. First of all, compensation for assimilation could be a linguistic universal, which is part of a Universal Grammar. Secondly, it could be the case that assimilation rules are perceptually constrained, as a consequence of a quasi-evolutionary process of the language itself (Deacon, 1997, de Boer, 2000). These alternatives can be distinguished with regard to their predictions about the effects of non-speech sounds. Given that UG assumes a 'language module,' non-speech sounds should not be able to trigger similar context effects than speech sounds. Contrasting this prediction, chapter eight showed that non-speech sound *can* induce similar context effects as those induced by speech sounds. This constitutes evidence in favor of the view that assimilation rules are perceptually, and more precisely, auditorily perceptually constrained.

This leads us to the *fourth* question. If this compensation for assimilation is an important process in speech comprehension, is it impaired in populations, which have a (subtle) speech-comprehension deficit, namely developmental dyslexics? The answer is clearly no, as chapter nine showed. This indicates that probably higher-level phonological processes are impaired in dyslexics, while basic perceptual processes are well in place. This, in turn, indicates a higher-level deficit associated with dyslexia.

Questions unanswered

Although a bit cliché, this dissertation still needs to address the question what "more research is necessary." At least two things spring to mind. First of all, I was more interested in proving the auditory basis of the effects observed than in investigating contributions of higher levels. Therefore, it was not investigated thoroughly what triggers the labeling bias that arises with language experience. Recall that the results showed (chapters four and six) that listeners with experience with a certain assimilation rule tend to have a bias towards perceiving the phonemes that may undergo assimilation. There are two general possibilities how this bias may arise. For the sake of presentation, I will use the example of the bias of Dutch listeners towards an alveolar nasal /n/, which may assimilate to /m/. First of all, the labeling bias may arise in any situation in which the phoneme identity is unclear. That is, if Dutch listeners do not know whether a segment is an /n/ or an /m/, they always tend to be biased to perceive an /n/. In this case, it does not matter whether the uncertainty arises as a consequence of a following /b/ in the

same utterance—which this dissertation shows to lead to problems in perceiving the place or articulation of the nasal—, or whether the uncertainty about the place of articulation of the phoneme may be due to background noise. This solution was hinted at in chapter four, in which we proposed that the likelihood of /n/ in the coda position may be responsible for biasing the Dutch listeners. However, even a theory such as the Underspecification theory (see chapter two for details) may come back via the backdoor here. It may be the case that, if in doubt, listeners always tend to perceive the ‘default,’ and hence underspecified, phoneme. Note that these explanations may not be mutually exclusive. It may be the high frequency of a certain feature that may lead to its lexical underspecification. Therefore, it will be a quite a challenge to disentangle these two proposals.

On a more general level, another, yet different account may also be proposed. In contrast to the previous proposals, the bias towards /n/ may also be context-sensitive, so that it only arises in situation in which assimilations may occur. In this case, the bias towards perceiving a certain phoneme *only* arises if the phonological context of this segment allows assimilation to occur. However, other noise sources, such as background noise, should not lead to a bias toward one phoneme. Let me note here that I did some pilot experiments on the effects of noise masking on the identification of a nasal as either /n/ or /m/. However, the results only—rather uninterestingly—seemed to indicate that the spectral composition of the noise itself may lead to an acoustic bias towards perceiving one or the other nasal. Therefore, the challenge will be to find ways to mask phoneme identity without introducing an acoustic bias for the perception of one or the other nasal. Then, we may indicate whether the bias toward perceiving /n/ over /m/ is due to a ‘phonological inference’ or due to some context-independent labeling bias.

In addition, there is at least one more loose thread with regard to the assumption of an auditory contribution to compensation for assimilation. If this assumption holds, then at least two more participant groups without experience with a given assimilation rule should show some degree of compensation for assimilation. These are, first, infants, and secondly, non-human species. It should be noted, though, that this supposition—that any of these groups have any chance at compensating for assimilation—would have met with utter disbelief at the time this endeavor was started. Therefore, raising this question in the first place may be regarded as a merit and not a shortcoming. Nevertheless, the proposal of an auditory locus clearly begs the question whether infants and non-human species may show some evidence of compensation for assimilation. The jury is still out.

Summary

Samenvatting

This thesis investigates how listeners are able to recognize spoken words even if they are modified by the phonological context in which they occur. Such modifications are called phonological assimilation: Due to such phonological assimilation the word 'garden' may be pronounced 'gardem' if followed by 'bench' but not if followed by 'chair.' First, chapter two shows that changed spoken words such as 'gardem' are probably not generally recognized as an instance of the original, canonical form 'garden.' Based on the theory of phonological underspecification, it was argued that features that may assimilate in certain contexts are not used in order to achieve lexical access. Then, it would not matter for word recognition whether the last phoneme in the word 'garden' is pronounced as /n/ or /m/. Earlier studies had already cast some on this assumption. The electrophysiological experiment reported in chapter two converges on these earlier critical reports. Thus, it seems that 'gardem' is not generally recognized as an instance of 'garden.' This leaves open the question how 'garden' is eventually recognized in the assimilated utterance 'gardembench.'

Chapter three then shows that 'gardem' is only accepted as an instance of 'gardem' if followed by 'bench,' which allows the assimilation; but not if followed by 'chair,' which does not allow the assimilation. This chapter represents a methodological development over earlier, similar reports by applying two new methods to the study of compensation for phonological assimilation. First, it is shown that a relatively simple word-identification task can be used to assess the compensation-for-assimilation process. In addition, this chapter represents the first application of event-related potentials (ERPs) to the study of context effects in speech perception. This method is especially useful as it allows us to evaluate the speech-perception process without imposing unnatural meta-linguistic tasks upon the participants. Instead the automatic perception process can be studied with possibly interfering effects of attention and decision processes: An ERP component called mismatch negativity (MMN) can be observed in a passive-listening situation if the participant pre-attentively detects a mismatch between two stimuli. Our results show that the same acoustical difference between 'garden' and 'gardem' evokes such a mismatch response only if the context does not allow the assimilation (e.g., '...chair'). However, no MMN is observed if the context allows the assimilation (e.g., '...bank').

These results complement the finding of chapter two that possibly assimilated word forms (such as 'gardem') are not generally accepted as instances of their canonical forms (i.e., 'garden'). Instead, it seems that modified forms are only accepted as instances of their canonical forms if they occur in a context that allows the assimilation.

Chapter four investigates whether the ability to compensate for assimilation is dependent on experience with an assimilation rule. To this end, Dutch utterances in which nasal place

In dit proefschrift wordt onderzocht hoe gesproken woorden herkend kunnen worden ondanks het feit dat zij door hun fonologische omgeving van klank veranderd zijn. Dergelijke veranderingen kunnen ontstaan als gevolg van fonologische assimilaties: Zo kan het woord 'tuin' als 'tuim' worden uitgesproken als het door het woord 'bank' wordt gevolgd, maar niet als het door het woord 'stoel' wordt gevolgd. Na een inleidend hoofdstuk laat hoofdstuk twee zien dat dergelijk veranderde vormen als 'tuim' niet automatisch als een instantie van het woord 'tuin' worden herkend. De theorie van 'fonologische onderspecificatie' veronderstelt dat wel. Deze theorie stelt dat fonologische kenmerken die in bepaalde contexten assimileerbaar zijn geen rol spelen bij de woordherkenning. Het zou dus niets uit moeten maken of 'tuin' met een /n/ of een /m/ wordt uitgesproken. Onze negatieve bevinding met betrekking tot deze theorie past in een reeks van eerdere bevindingen die ook al problematisch waren voor de theorie van 'fonologische onderspecificatie'. Als de theorie van 'fonologische onderspecificatie' niet houdbaar blijkt, blijft wel de vraag open staan hoe veranderde woordvormen uiteindelijk wel worden herkend.

Hoofdstuk drie laat zien dat 'tuim' alleen als een instantie van het woord 'tuin' wordt herkend als het in een fonologische context voorkomt, waarin deze verandering volgens de fonologische regels is toegestaan. Dus wordt 'tuim' herkend als een instantie van 'tuin' als het in 'tuimbank' voorkomt, maar niet als het in '*tuimstoel' voorkomt. Dergelijke resultaten zijn al eerder gevonden. Toch levert hoofdstuk drie twee nieuwe methodes op om de perceptuele processen, die bijdragen aan de herkenning van geassimileerde woordvormen, te bestuderen. Ten eerste laten de resultaten zien dat de compensatie voor assimilatie ook te bestuderen valt met een redelijk simpele woord-identificatie taak. Ten tweede wordt hier voor het eerst gebruik gemaakt van metingen van elektrische hersenpotentialen in een onderzoek naar context effecten bij spraakperceptie. Deze methode is vooral vruchtbaar omdat het perceptieproces bestudeerd kan worden zonder dat de proefpersoon een onnatuurlijke taak hoeft uit te voeren. In plaats daarvan kan men tijdens het luisteren naar een reeks stimuli een Mismatch Negativity (MMN) observeren als een proefpersoon in staat is een onderscheid te maken tussen twee stimuli. Onze resultaten laten zien dat het verschil tussen 'tuin' en 'tuim' onopgemerkt blijft als deze stimuli gevolgd door het woord '...bank' worden aangeboden. Het verschil tussen 'tuin' en 'tuim' wordt wel onderscheiden als deze woorden door '...stoel' gevolgd worden aangeboden.

Deze resultaten vormen een aanvulling op die van hoofdstuk twee. Uit die resultaten bleek dat veranderde vormen zoals 'tuim' niet altijd als een instantie van het oorspronkelijke woord 'tuin' werden herkend. Nu blijkt dat vormen zoals 'tuim' alleen als een instantie van het oorspronkelijke woord 'tuin' worden herkend als ze in een context voorkomen waarin de verandering van /n/ naar /m/ mogelijk is.

assimilation had been applied were presented to Portuguese listeners. The phonology of Portuguese does not know a rule of nasal place assimilation. Nevertheless, Portuguese listeners show a similar, though not identical, context sensitivity as listeners with experience with a rule of nasal place assimilation. They find also perceive 'gardembench' as similar to 'gardenbench.' In contrast to Dutch listeners, however, they do mistake canonical forms for assimilated forms as often as they mistake assimilated forms for canonical forms. Dutch listeners pre-dominantly make the latter mistake.

Chapter five investigates whether the compensation mechanism works in a categorical, all-or-nothing way or is able to use fine-grained acoustic details. The latter seems to be the case. In addition, the argument is made that such fine-grained mechanisms are better able to make use of the available information in the speech signal: Production studies indicate /m/s in assimilated utterances as 'gardembench' subtly differ from intended word-final /m/s as in 'armchair.' This information cannot be used by a mechanism that rests on phonologically categorized input.

Chapter six then starts a series of more thorough cross-linguistic investigations, in which assimilated utterances are presented to native-language listeners and listeners, who are neither familiar with the language of the utterances, the kind of assimilation rule, *nor* the language family of the utterances. This is achieved by presenting Hungarian utterances to Hungarian and Dutch listeners. In addition, two Hungarian utterances, a Hungarian word and a nonword are used. In chapter six, listeners are asked to either identify the phonological make-up of the utterances or to discriminate two utterances. Results show that discrimination between a phonologically changed utterance and a canonical form is more difficult in a context that allows the phonological change as a phonological assimilation. This is the case independent of the wordness of the stimulus and the experience of the listeners with the assimilation rule: Hungarian and Dutch participants performed comparably with the Hungarian utterances; and Hungarian participants showed equivalent results with the Hungarian words and nonwords. However, a small difference due to native-language experience is observed in the identification task. While context influences identification performance independent of native-language experience, Hungarian listeners are more likely to perceive the Hungarian canonical forms. Whether the stimulus in question is an existing Hungarian word or not, however, does not matter here. This replicates the results obtained in the chapters three and four. There, context influenced identification of phonologically modified Dutch utterances in Dutch, German, and Portuguese participants. All participants, independent of experience with the kind of phonological changes applied to the stimuli, have problems to make a clear distinction between a canonical and a changed form in a context in which the change is a lawful phonological assimilation. However, only Germanic

In hoofdstuk vier wordt de vraag behandeld of deze 'compensatie voor assimilatie' alleen plaatsvindt als de luisteraar aan dergelijke fonologische assimilaties is blootgesteld. Om dit te onderzoeken hebben wij dezelfde Nederlandse woorden met en zonder veranderingen aan Portugese luisteraars voorgelegd. Het Portugees kent geen fonologische regel die /n/ in /m/ verandert. Toch laten de Portugese luisteraars een gelijkend, maar niet identiek patroon aan resultaten zien als luisteraars die dit soort assimilaties uit hun 'moedertaal' kennen. Alle hebben luisteraars meer moeite om een verschil tussen 'tuin' en 'tuim' te maken als deze in een context voorkomen waarin het /n/ in een /m/ mag veranderen—zoals in 'tui(n/m)bank'—dan wanneer 'tuin' en 'tuim' in een context voorkomen die geen assimilatie toestaat—zoals 'tui(n/m)stoel'. Taalervaring speelt daarbij wel een rol: Alleen de Nederlandse proefpersonen, en niet de Portugese, hebben de neiging om vaker het 'oorspronkelijke' woord te horen. Terwijl Portugese luisteraars even vaak 'tuinbank' als 'tuimbank' en 'tuimbank' als 'tuinbank' horen, maken Nederlandse luisteraars voornamelijk het tweede type fouten.

In hoofdstuk vijf wordt bekeken of het compensatie mechanisme op een categoriale—alles of niets—manier of meer continu werkt. In het laatste geval zou de mogelijkheid bestaan dat fonetisch detail een rol speelt in het compensatie-proces. De resultaten laten zien dat dit het geval is. Verder wordt aangetoond dat een meer continu mechanisme zinvol is omdat fonetisch detail het gedeeltelijk mogelijk maakt een geassimileerde /m/ zoals in 'tuimbank' te onderscheiden van een 'geïntendeerde' /m/ zoals in 'armstoel'.

In hoofdstuk zes komt een nieuwe reeks crosslinguïstisch experimenten aan bod. Hierbij worden geassimileerde uitspraken aan luisteraars aangeboden die geen ervaring hebben met de taal en de taalfamilie waarin die uitspraken gedaan zijn, en evenmin met het type assimilatie dat in de uitspraken is toegepast. Dit wordt bereikt door Hongaarse uitspraken aan Nederlandse luisteraars voor te leggen. Verder worden Hongaarse woorden en pseudo-woorden gebruikt. Ten eerste moeten luisteraars de woorden in eerste instantie als 'origineel' of 'veranderd' herkennen, en daarna proberen een onderscheid tussen twee uitspraken te maken. Onafhankelijk van de taalervaring kostte het discrimineren meer moeite als de woorden in een context voorkwamen, die een assimilatie toeliet dan in het geval dat de context geen assimilatie toeliet. Een klein effect van taalervaring werd echter zichtbaar in de 'identificatietaak'. Hoewel de Nederlandse en Hongaarse proefpersonen een meer categoriale identificatiefunctie hebben wanneer er sprake is van een context die geen verandering toelaat, hebben alleen Hongaarse proefpersonen de neiging om het oorspronkelijke Hongaarse 'woord' te horen, dus de assimilatie in de perceptie ongedaan te maken. Dit was echter onafhankelijk ervan of de veranderde uitspraak een Hongaars woord of een Hongaars pseudo-woord was. Dit repliceert de bevindingen van hoofdstuk drie en vier.

listeners have a response bias that leads them to recover the canonical form when hearing the assimilated form. Similarly, Hungarian listeners are more likely to perceive the canonical forms in Hungarian words and nonwords.

Chapter seven then again uses the methodological innovation introduced in chapter three. Context effects in speech perception are measured with ERPs, namely the MMN. Results with ERPs prove to be more sensitive to acoustic details of the stimuli than the behavioral measurements in chapter six. A context sensitivity of the MMN is only observed if the phonological change is not strong. This is observed in Hungarian and Dutch listeners. Strong changes lead to a strong MMN, even if the context would allow the change to occur. This replicates the finding from chapter five that the compensation-for-assimilation process is influenced by acoustic detail.

Chapter eight then tests whether compensation of assimilation is to be attributed to a speech-specific mechanism or may be explained by a general-perceptual mechanism. It is assessed whether non-speech sounds can trigger similar context effects as speech sounds. This is the case. Therefore, it seems that general-perceptual principles are involved in compensation for phonological assimilation. Nevertheless, it should be remembered that the results reported in chapter four and six converge in showing that experience with an assimilation rule in one's (native) language contribute to the perceptual compensation for phonological assimilation.

Chapter nine finally reports an applied study. It is asked whether the compensation mechanisms, as evident throughout this dissertation, may be impaired in developmental dyslexics. The chapter falls back on the methodological contribution of chapter three by using a—for the participant—relatively simple word-identification task. Otherwise, it would have been difficult to measure compensation for phonological assimilation in children. A mathematical model shows that the 'categorical-perception deficit' associated with developmental dyslexia may be attributed to faulty compensation mechanisms. However, results do not bear out this hypothesis. Compensation for assimilation seems to work similarly in dyslexics and normal-reading children.

A final chapter then reviews the commonalities between the different chapters. It is concluded that with different material and different designs, the results show that compensation for assimilated is not a unitary process. Instead, there seems to be a general perceptual bias, which contributes crucially to compensation for assimilation. In addition, native language experience seems to be built upon these general perceptual mechanisms to lead to a complete compensation for assimilation.

In die hoofdstukken bleek dat Nederlandse uitspraken met een assimilatie die mogelijk is volgens de fonologie van het Nederlands moeilijk te onderscheiden zijn van een onveranderde versie van de uitspraak. Dit was het geval voor Nederlandse, Duitse, en Portugese luisteraars. Verder bleek echter dat luisteraars die wel ervaring hebben met de assimilatie (Nederlandse en Duitse luisteraars) vaker de neiging hebben de 'onveranderde' vorm waar te nemen. Dit is vergelijkbaar met de neiging van de Hongaarse proefpersonen om vaker de 'onveranderde' vorm te horen dan Nederlandse luisteraars als het om Hongaarse uitspraken gaat.

In Hoofdstuk zeven wordt de perceptie van Hongaarse woorden door Nederlandse en Hongaarse luisteraars verder onderzocht. Hier wordt gebruik gemaakt van de methodologische vernieuwing die eerder is geïntroduceerd: de meting van hersenpotentialen, met name de MMN. De resultaten laten vooral zien dat akoestische details van de geassimileerde uitspraken bepalen in hoeverre de assimilatie in de perceptie ongedaan wordt gemaakt. Dat wil zeggen dat bij duidelijk veranderde vormen de context niet meer beïnvloedt of de verandering wel of niet opgemerkt wordt. Een vergelijkbaar resultaat was in al hoofdstuk vijf gevonden. Net als in hoofdstuk zes zijn er geen verschillen tussen Nederlandse en Hongaarse luisteraars.

In hoofdstuk acht wordt onderzocht of compensatie voor assimilatie veroorzaakt wordt door een taalspecifiek mechanisme of algemene perceptuele mechanismen. Om dit te onderzoeken, wordt getest of compensatie verschijnselen ook gevonden kunnen worden bij niet-talige stimuli. Dit is het geval. Het blijkt dus dat algemene perceptuele processen een rol spelen bij compensatie voor assimilatie. Toch moet benadrukt worden dat eerdere resultaten al lieten zien dat ervaring met een taal wel de perceptie van geassimileerde uitspraken beïnvloedt.

Hoofdstuk negen ten slotte geeft verslag van een toegepaste studie. De vraag die gesteld wordt is of de in dit proefschrift gevonden compensatie-voor-assimilatie processen wel goed functioneren bij kinderen met ontwikkelingsdyslexie. Een wiskundig model laat zien dat de problemen van dyslectici goed verklaard kunnen worden als compensatie processen niet goed zouden werken. Om dit bij dyslectici te onderzoeken wordt gebruikt gemaakt van een van de methodologische innovaties van hoofdstuk drie. De daar gebruikte woord-identificatie taak maakt het mogelijk compensatie processen makkelijk bij kinderen te meten. De resultaten laten echter zien dat compensatie processen bij dyslectici net zo werken als bij niet-dyslectici.

In een epiloog wordt dan terug geblikt op de acht empirische hoofdstukken. Op basis van de verzamelde data wordt geprobeerd antwoord te geven op de vragen die in de inleiding werden gesteld. De conclusie is dat compensatie voor assimilatie door het samenspel van verschillende processen tot stand komt. Hierbij blijken zowel algemene perceptuele principes als specifieke taalervaring een rol te spelen.

References

- Adams, M.J. (1990). *Beginning to read: Thinking and learning and about print*. Cambridge, MA: MIT Press.
- Adams, S.G., Weismer, R. & Kent, R.D. (1993) Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, and Hearing Research*, 36, 41-54.
- Allen, J.S. & Miller, J.S. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics*, 63, 798-810.
- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Balota, D.A. & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340-357.
- Beddor, P.S. & Krakow, R.A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106, 2868-2887.
- Behringer, J. (1996). Experimental Run Time System (Version 3.18) [Computer Software]. Darmstadt, Germany: Technische Hochschule Darmstadt.
- Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J.C. Goodman and H.C. Nusbaum, (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.
- Bezooijen, R. van (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38, 253-266.
- Blomert, L. & Mitterer, H. (in press). The fragile nature of the speech-perception deficit in dyslexia: natural vs. synthetic speech. *Brain & Language*.
- de Boer, B. (2000). Self-organisation in vowel systems. *Journal of Phonetics*, 28, 441-465.
- Boersma, P. & Weenink, D. (1999). Praat 3.9: a system for doing phonetics with the Computer [Computer software]. Amsterdam, The Netherlands: University of Amsterdam.
- Boersma, P. & Weenink, D. (2001). Praat 4.0. Amsterdam, The Netherlands: University of Amsterdam.
- Bon, W.H.J. van (1982). *Handleiding Taaltests voor kinderen* [Manual for the language tests for children]. Lisse, The Netherlands: Swets & Zetinger.
- Booij, G. (1995). *The phonology of Dutch*. Oxford, UK: Clarendon Press.
- Borland (1999). Delphi 5.0 [Computer programming language]. Scotts Valey, CA: Authors.
- Bos, K.P. van den, Lutje Spelberg, H.C., Scheepstra, A.J.M., & de Vries, J.R. (1994). *De Klepel, een test voor de leesvaardigheid van pseudoworden. Verantwoording, handleiding, diagnostiek en behandeling* [Klepel: A test for the reading proficiency in reading pseudowords: justification, manual, diagnostics and treatment.] Nijmegen, The Netherlands: Berkhout.
- Bowers, P.G & Swanson, L.B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology*, 51, 195-219.
- Bradley, L. & Bryant, P.E. (1983). Categorizing sounds and learning to read: a causal connection. *Nature*, 301, 419-421.
- Brady, S. (1997). The ability to encode phonological representations: An underlying difficulty of poor readers. In B.A. Blachman et al. (Eds.), *Foundations of reading acquisition and dyslexia: Implications for early intervention*. (pp. 21-47). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, 35, 345-367.
- Brady, S., Poggie, E., & Rapalla, M. (1989). Speech repetition abilities in children who differ in reading skills. *Language and Speech*, 32, 109-122.
- Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brus, B.T. & Voeten, M. (1972). *Een-minuut leestest* [one-minute reading test]. Nijmegen, The Netherlands: Berkhout.
- Cavalli-Sforza, L.L. (2000). *Genes, Peoples and Languages*. New York, NJ: North Point Press.

- Catts, H.W. (1989). Defining dyslexia as a developmental language disorder. *Annals of Dyslexia*, 39, 50-65.
- Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York, NJ: Harper & Row.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.
- Coenen, E., Zwitsersloot, P. & Bölte, J. (2001). Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes*, 15, 535 – 564.
- Cole, R.A. (1973). Listening for mispronunciation: A measure of what we hear during speech. *Perception & Psychophysics*, 13, 153-156.
- Collins, B. & Mees, I.M. (1996). *The phonetics of English and Dutch*. Leiden, The Netherlands: E.J. Brill.
- Darcy, I. (2002). Online processing of phonological variation in speech comprehension: The case of assimilation. In S. Hawkins & Noël Nguyen (Eds.). *Temporal integration is the perception of speech* (p. 32). ISCA.
- Deacon, T.W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York, NJ: Norton.
- Delgutte, B. (1997). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 507-538). Oxford, UK: Blackwell.
- Denckla, M., Rudel, R. G. (1976). Rapid automatized naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14, 471-479.
- Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch* [dissertation]. Utrecht, The Netherlands: LOT.
- Evans, B. & Iverson, P. (2002). Vowel normalisation for dialect. In S. Hawkins & N. Nguyen (Eds.). *Temporal integration is the perception of speech* (p. 37). ISCA.
- Farnetani, E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle (Ed.), *Handbook of phonetic sciences* Oxford, UK: Blackwell.
- Fitch, H.L., Halwes, T.E., Erickson, D.M., Liberman, A.M. (1980). Perceptual equivalent of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 27, 343-350.
- Fowler, C.A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, 20, 143-165.
- Fowler, C.A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730-1741.
- Fowler, C.A., Best, C.T., & McRoberts, G.W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559-570.
- Fowler, C.A. & Brown, J.M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, 62, 21-32.
- Fowler, C.A., Brown, J.M., & Mann, V.A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 877-888.
- Gaskell, M.G. (in press). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*.
- Gaskell, M.G., Hare, M., & Marslen-Wilson, W.D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, 19, 407-439.
- Gaskell, M.G. & Marslen-Wilson, W.D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144-158.
- Gaskell, M.G. & Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380-396.
- Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.
- de Gelder, B. & Vroomen, J. (1998). Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain and Language*, 64, 269-281.
- Gerrits, E. (2001). *The categorization of speech sounds by adults and children* [dissertation]. Utrecht, The Netherlands: LOT.
- Godfrey, J.J., Sydral-Lasky, A.K., Millay, K.K., & Knox, C.M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology*, 32, 401-424.

- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133-159.
- Gow, D. W. (2002a). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163-179.
- Gow, D. W. (2002b). Integrating feature cues over time to resolve ambiguity. In S. Hawkins & Noël Nguyen (Eds.), *Temporal integration is the perception of speech* (p. 37). ISCA.
- Gow, D.W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575-590.
- Gow, D. W. & Hussami, P. (1999, November). Acoustic modification in English place assimilation. Paper presented at the meeting of the Acoustical Society of America, Columbus, OH.
- Gussenhoven, C. & Jacobs, H. (1998). *Understanding phonology*. London: Arnold.
- Greenberg, S., Carvey, H., Hitchcock, L. & Chang, C (2002) Time frames of spoken language. In S. Hawkins & N. Nguyen (Eds.), *Temporal integration in the perception of speech* (p.13). ISCA.
- Hauser, M.D., Newport, E.L., & Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, 78, B53-B64.
- Hobbes, Th. (1983, first published 1651). *Leviathan*. Penguin Books Ltd.: Harmondsworth, UK.
- Heim, S., Eulitz, C., Wienbruch, C., & Elbert, T. (1999). Auditory processing assessed by the mismatch field in dyslexic children. *Psychophysiology*, 36 (suppl. 1), 59.
- Hessen, A.J. van & Schouten, M.E.H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, 56, 56-72.
- Holt, L.L. & Lotto, A.J. (2002) Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167, 156-169.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: a case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, 109, 764-774.
- Huckvale, M. (1999). Opportunities for re-convergence of engineering and cognitive science accounts of spoken word recognition. *Speech, Language, and Hearing: work in progress*, 11, 62-75. London, UK: University College London.
- Hume, E. & Johnson, K. (2001). *The role of speech perception in phonology*. New York, NJ: Academic Press.
- Hume, E. & Johnson, K. (2001). The interplay of perception and phonology. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 3-26). New York, NJ: Academic Press.
- Hura, S.L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language & Speech*, 59-72.
- Hurford, D.P. & Sanders, R.E. (1990). Assessment and remediation of a phonemic discrimination deficit in reading disabled second and fourth graders. *Journal of Experimental Child Psychology*, 50, 396-415.
- Jufrasky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7* (pp. 3-34). Berlin: Mouton de Gruyter.
- Keppel, G. (1991). *Design and analysis: a researcher's handbook* (3rd edition). Upper Saddle River, NJ: Prentice Hall.
- Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 451-470). Cambridge, UK: Cambridge University Press.
- Kingston, J. & Macmillan, N.A. (1995). Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants - a detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, 97, 1261-1285.
- Kohler, K.J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In: W.J.Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modelling* (pp. 69-92). Dordrecht: Kluwer
- Ladefoged, P. (1996). *Elements of acoustic phonetics* (2nd edition). Chicago: University of Chicago Press.
- Ladefoged, P. & Maddieson, I. (1996). *Sounds of the world's languages*. Oxford: Blackwell Publishers.
- Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition*, 38, 245-294.
- Lahiri, A. & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 637-676). Berlin: Mouton de Gruyter.

- Lerner, A.J. & Loewe, F. (1956). *My fair lady*. Musical stage production.
- Liberman, A.M., Harris, K.S., Kinney, J.A., & Lane, H. (1957). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*, 379-388.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In: W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling*, (pp. 403-439). Dordrecht: Kluwer.
- Los, S.A. (1999). Identifying stimuli of different perceptual categories in mixed blocks of trials: Evidence for cost in switching between computational processes. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 3-23.
- Lotto, A.J. & Kluender, K.R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*, 602-619.
- Lotto, A.J., Kluender, K.R., & Holt, L.L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, *102*, 1134-1140.
- Macmillan, N.A., Kingston, J., Thorburn R., Dickey, L.W., & Bartels C. (1999). Integrality of nasalization and F-1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions. *Journal of the Acoustical Society of America*, *106*, 2913-2932.
- Mann, V.A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, *28*, 407-412.
- Mann, V.A. & Repp, B.H. (1981). Influence of preceding fricative on stop-consonant perception. *Journal of the Acoustical Society of America*, *69*, 548-558.
- Marr, D. (1982). *Vision*. W. H. Freeman, San Francisco, CA.
- Marslen-Wilson, W.D., Moss, H.E., & Van Halen, S. (1996). Perceptual distance and competition. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376-1392.
- Marslen-Wilson, W.D. & Warren, P. (1994). Levels of perceptual representations and processes in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653-675.
- Massaro, D. (1998). *Perceiving talking faces: from speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- Massaro, D.W. & Cohen, M.M. (1983). Categorical or continuous speech perception: a new test. *Speech Communication*, *2*, 15-35.
- Masterson, J., Hazan, V., & Wijayatilake, L. (1995). Phonemic processing problems in developmental phonological dyslexia. *Cognitive Neuropsychology*, *12*, 233-259.
- McBride-Chang, C. (1995). Phonological processing, speech perception, and reading disability: an integrative review. *Educational Psychologist*, *30*, 109-121.
- McBride-Chang, C. (1996). Models of speech perception and phonological processing in reading. *Child Development*, *67*, 1836-1856.
- McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.
- McQueen, J.M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21-46.
- McQueen, J.M. & Cutler, A. (2001). Spoken word access processes: An introduction. *Language and Cognitive Processes*, *16*, 469-490.
- Mendeloff, H. (1969). *A manual of comparative romance linguistics: Phonology and morphology*. Washington, D.C.: The Catholic University of America Press.
- Mitterer, H. (2000). Are connected-speech processes selected for communicative consequences? In B. de Boer & P. Vogt (Eds.), *Proceedings of the Benevolang Workshop 2000* (pp. 31-34). Brussels: Vrije Universiteit Brussel.
- Mitterer, H. & Blomert, L. (in press). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics*.
- Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology*, *58*, 199-231.
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, *24*, 45-64.
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Erlbaum.

- Näätänen, R. (1995). The Mismatch Negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing*, 16, 6-18.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vaino, M., Alki, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Ahlo, K. (1997). Language specific phoneme representation revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport*, 4, 503-506.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125, 826-859.
- Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.
- Nearey, T.D. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustic Society of America*, 85, 2088-2113.
- Nolan, F. (1992). The descriptive role of segments: evidence from assimilations. In G.J. Docherty & R.D. Ladd (Eds.), *Papers in Laboratory Phonology: II. Gestures segment, prosody* (pp.261-280). Cambridge, UK: Cambridge University Press.
- Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 20, 1-32.
- Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research*, 39, 278-297.
- Nittrouer, S. (1999). Do temporal processing deficits cause phonological problems? *Journal of Speech, Language, and Hearing Sciences*, 42, 925-942.
- Norris, D. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*, 62, 714-719.
- Norris, D., McQueen, J.M., Cutler, A., & Butterfield, S. (1997). The Possible-Word Constraint in the Segmentation of Continuous Speech. *Cognitive Psychology*, 34, 191-243.
- Norris, D., McQueen, J.M. & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209-1228.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-324.
- Ohala, J. J. 1990. The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech* (pp. 258-275). Cambridge, UK: Cambridge University Press.
- Olsson, M. (1992). *Hungarian phonology and morphology*. Lund: Lund University Press.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12, 1038-1055.
- Pitt, M.A. & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.
- Post, Y.V., Foorman, B.R., & Hiscock, M. (1997) Speech perception and speech production as indicators of reading difficulty. *Annals of Dyslexia*, 47, 3-27.
- Psychology Software Tools. E-Prime (Beta 5.0). Pittsburgh, PA.
- Pulvermüller, F., Kujala, T., Shryrov, Y., Simola, J., Tiitinen, H., J., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage*, 14, 607-616.
- Reed, M.A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, 48, 270-292.
- Repp, B.H. & Svastikula, K. (1988). The perception of the [m]-[n] distinction in VC syllables. *Journal of the Acoustical Society of America*, 83, 237-247.
- Rietveld, A.C.M. & van Heuven, V.J. (1997). *Algemene fonetiek* [General phonetics]. Bussum, The Netherlands: Coutinho.

- Rojas, R. (1996). A short proof of the posterior probability property of classifier neural networks. *Neural Computation*, 8, 41-43.
- Rosen, S. & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Sciences*, 44, 384-388.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493-515.
- Samuel, A.G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, 125, 28-51.
- Schouten, B., Gerrits, E. & van Hesse, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41, 71-80.
- Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments & Computers*, 30, 131-145.
- Schulze, H. & Scheich, H. (1999). Discrimination learning of amplitude-modulated tones in Mongolian gerbils. *Neuroscience Letters*, 261, 13-16.
- Scott, S.K. & Wise, R.J.S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, 41, 23-24.
- Seo, M. (2001). A perception-based study of sonorant assimilation in Korean. *OSU Working Papers in Linguistics*, 55, 43-69.
- Sharma, A. & Dorman, M.F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, 106, 1078-1083.
- Shaywitz, S. E. (1998). Current concepts: Dyslexia. *New England Journal of Medicine*, 338, 307-312.
- Sinnott, J.M. & Saporita, T.A. (2000). Differences in American English, Spanish, and monkey perception of the say-stay trading relation. *Perception & Psychophysics*, 62, 1312-1319.
- Siptár, P. & Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford, UK: Oxford University Press.
- Snowling, M. (2000). *Dyslexia* (2nd ed.). Malden, MA: Blackwell Publishers.
- Snowling, M., Goulandris, N.B., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology*, 41, 489-507.
- Snowling, M., Nation, K., Moxham, P., Gallagher, A., Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading*, 20, 31-41.
- Steels, L. (1999). *The Talking Heads Experiment: Vol. I Words and Meaning* (Special Pre-edition). Brussels: Vrij Universiteit Brussels.
- Stein, J. & Walsh, V. (1997). To see but not to read: the magnocellular theory of dyslexia. *Trends in Neuroscience*, 20, 147-152.
- Steriade, D. (2001). Directional asymmetries in place assimilation: a perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219-250). New York, NJ: Academic Press.
- Stevens, K.S. (1998). *Acoustic Phonetics*. Cambridge, MA.: MIT Press.
- Stevens, K. N. (2002). Towards a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872-1891.
- Stone, G.O. & Van Orden, G.C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 744-774.
- Studdert-Kennedy, M. (2002). Deficits in phoneme awareness do not arise from failures in rapid auditory processing. *Reading and Writing*, 15, 5-14
- Studdert-Kennedy, M. & Mody, M. (1995). Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychonomic Bulletin and Review*, 2, 508-514.
- Suomi, K., McQueen, J.M. & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36, 422-444.
- Summerfield, Q. & Assmann, P.F. (1989). Auditory enhancement and the perception of concurrent vowels. *Perception & Psychophysics*, 45, 529-536.
- Sussman, H.M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: The orderly output constraint. *Behavioral and Brain-Sciences*, 21, 241-299.
- Tallal, P. (1980). Auditory temporal perception, phonics, and the reading disabilities in children. *Brain and Language*, 9, 182-198.

- Tallal, P., Miller, S., Fitch, R. (1993). Neurobiological Basis of Speech: A case for the Preeminence of Temporal Processing. In Tallal, A. M. Galaburda, R. R. Llinas and C. von Euler (Eds), *Temporal Information Processing in the Nervous System: Special Reference to Dyslexia and Dysphasia* (pp. 27-47), Annals of the New York Academy of Sciences.
- Tallal, P. & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, 13, 69-74.
- Thierry, G., Doyon, B., & Démonet, J.F. (1998). ERP mapping in phonological and lexical semantic monitoring tasks: A study complementing previous PET results. *NeuroImage*, 8, 391-408.
- Tobin, Y. (1997). *Phonology as human behavior*. Durham: Duke University Press.
- Traunmüller, H. & Eriksson A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107, 3438 - 3451.
- Utman, J.A., Blumstein, S.E., & Burton, M.W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics*, 62, 1297-1311.
- Van der Heijden, A.H.C., Hagenaar, R. & Bloem, W. (1984). Two stages in postcategorical filtering and selection. *Memory and Cognition*, 12, 458-469.
- Van Heuven, V.J. & Van Berg, I. (1982). Sprekerstrategie en de perceptieve dominantie van het woordbegin [Speakers' strategies and the perceptual dominance of word beginnings]. *Glott*, 5, 31-48.
- Wagner, R. K. & Torgesen, J.K. (1987). The nature of phonological processing and its causal role in learning to read. *Psychological Bulletin*, 101, 192-212.
- Wang, X., Lu, T., & Liang, L. (2003). Cortical processing of temporal modulations. *Speech Communication*, 41, 107-121.
- Warren, R.M. (1971). Identification times for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, 9, 345-349.
- Warren, R.M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge, UK: Cambridge University Press.
- Warren, R.M. (2000). Phonemic organization does not occur: Hence no feedback. *Behavioral and Brain Sciences*, 23, 350-351.
- Warren, R.M. & Bashford, J.A., Jr. (1976). Auditory contralateral induction: An early stage in binaural processing. *Perception & Psychophysics*, 20, 380-386.
- Warren, R.M., Bashford, J.A., & Gardner, D.A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. *Perception & Psychophysics*, 47, 345-349.
- Werker, J.F. & Tees, R.C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology*, 41, 48-61.
- Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Ahlo, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7, 357-369.
- Witruk, E., Ho, C.S. H., & Schuster, U. (2002). Working memory in dyslexic children: How general is the deficit? In Witruk, Evelin (Ed); Friederici, Angela D. (Ed); et-al., *Basic functions of language, reading and reading disability. Neuropsychology and cognition, Vol. 20* (pp. 281-297). Dordrecht, Netherlands: Kluwer.
- Wolf, M., Bowers, P. Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities*, 33, 387-407.
- Woltjer, J.J. (2001). *Recent verleden: de geschiedenis van Nederland in de twingste eeuw* [Recent past: the history of the Netherlands in the 20th century]. Amsterdam, The Netherlands: Munitga Maarten.
- Wunderlich, J.,L. & Cone-Wesson, B.K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *Journal of the Acoustical Society of America*, 109, 1526-1537.

Curriculum Vitae

Holger Mitterer was born in Hanau, Germany, on January 4th 1973. He attended the Hohe Landesschule in Hanau and obtained his Abitur diploma in 1992. From 1992 to 1993, he was working as caretaker for elderly and severely handicapped as a replacement for military service.

In 1998, he graduated as psychologist (Diplom-Psychologe) after studying psychology at the Universities of Bielefeld, Germany, and Leiden, The Netherlands. During this time, he also worked as a research assistant at the University of Bielefeld and as a piano teacher at the Jugendmusikschule in Brackwede, Germany.

From 1998 until 2003 he was employed as a Ph.D. student and junior lecturer at the University of Maastricht, The Netherlands.

The research described in this book was conducted during this period at the University of Maastricht and the Hungarian Academy of Science in Budapest, Hungary.

Currently, Holger Mitterer is a member of the speech-comprehension group at the Max-Planck Institut für Psycholinguistik in Nijmegen, The Netherlands.

In Dutch, English, German, and many other languages an "n" may be pronounced as an "m" if it is followed by, for instance, a "b". "Garden bench" may hence be pronounced as "gardem bench". These kind of changes in fluent speech are called *phonological assimilations*. Most speakers are unaware of such changes, and speakers of German even often reject the notion that their speech may contain such 'impurities'. Nevertheless, such assimilations happen.

Inadvertently, these assimilations pose a problem for the listener. How can the listener accept that "gardem" means "garden", while maintaining that "gun" and "gum" are different things? This book tries to answer this question. It is shown that the listener's perception *compensates* for assimilations.