How neuronal oscillations code for temporal statistics

HOW NEURONAL OSCILLATIONS CODE FOR TEMPORAL STATISTICS

Cover	Sanne ten Oever
Production	Proefschriftmaken.nl || Uitgeverij BOXPress
ISBN	978-9462-954-89-2

HOW NEURONAL OSCILLATIONS CODE FOR TEMPORAL STATISTICS

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday 9th of June 2016, at 14.00 hours

by

Sanne ten Oever

Promotor
Prof. Dr. A.T. Sack

Copromotor
Dr. N.M. van Atteveldt

Assessment Committee

Prof. Dr. B. Jansma	Chair
Prof. Dr. R. Goebel	
Prof. Dr. J. Obleser	University of Lübeck, Germany
Prof. Dr. J.H.M. Vroomen	University of Tilburg, The Netherlands

Table of content

CHAPTER 1

GENERAL INTRODUCTION

In daily life the environment is full of abundant sensory information. A system that has to process all this input needs to be able to extract and organize this information in a meaningful way. Accordingly, information can be grouped as belonging to either the same or to a different event. Grouping two separate sensory input sources to one event is achieved by using statistical regularities in the environment [see e.g. (Perruchet & Pacton, 2006)]. For example, we know that a specific voice belongs to a specific person because the vision and the sound of the two virtually always occur together. The automaticity of this binding process becomes very clear if we for the first time see a person we only heard before, such as a radio presenter; somehow the voice and the vision of the person do not seem to belong together.

The environment is full of these regularities and without them making sense of the world would be a complicated task. The current thesis focusses on a special form of such regularities, namely temporal statistics. Temporal statistics refers to the consistent relation between two separate inputs in time. Music is a great example as it is clear that it has a specific temporal structure and subsequent tones are systematically presented with a specific delay. But temporal information is more omnipresent. For example in speech there are systematic temporal regularities in how spoken words follow each other. We even use temporal information to localize sounds in the environment.

The examples described above show that temporal information can be implicitly acquired through experience. This is fundamentally different from making explicit judgments about the duration of an event. While implicit temporal information is used to optimize integration processes during perception; extracting temporal information is a goal in itself during explicit temporal judgments (Coull & Nobre, 2008). In daily life, temporal cues will be mostly used to guide our perception and this is also the focus of the current thesis.

Most sensory information gets conveyed to the brain as a consequence of a cascade of different active neurons, initiated through the activation caused by a physical change in sensory neurons. For example, sound pressure changes induce movements of hair-cells in the cochlea, subsequently activating downstream neurons in the auditory pathway (Hudspeth, 1989). Temporal information is qualitatively different from these types of sensory information as it by itself is not

conveyed through any physical medium. Instead, temporal information represents the experience of the succession of different events. Since temporal information relates two different sensory events to each other, it always needs a reference to the external sensory world [(James, 1886; Pöppel, 1997) but see (Newton, Motte, & Cajori, 1987) for another account]. This property makes studying the process of extracting and coding temporal regularities in the sensory input challenging. The work in the current thesis investigates the behavioral benefits afforded by extracting the temporal relationships from the environment and how these learned temporal associations might be encoded in the brain.

Temporal statistics in the environment

Rhythms and temporal cueing

Temporal associations require a consistent temporal relationship between two re-occurring events. These associations can be formed as soon as the two stimuli systematically occur in the environment. Two basic forms of associations can be distinguished: rhythms (e.g. in music) and temporal cueing (e.g. the delay that exists between dropping a ball and the sound of the ball). For the most basic rhythm one unique temporal delay exists between subsequent stimuli that are presented repeatedly. Therefore, three events are sufficient to understand the temporal relationship between these repeating stimuli (see figure 1). For temporal cueing one specific stimulus acts as a cue how long the time delay for a subsequent target stimulus will be (also called foreperiod). However, the arrival time of the initial cue is unknown. To understand these types of associations at least four stimuli are required (cue-target, cue-target). For both types of associations the stimulus modality is in principle irrelevant and stimuli could even be from different modalities.

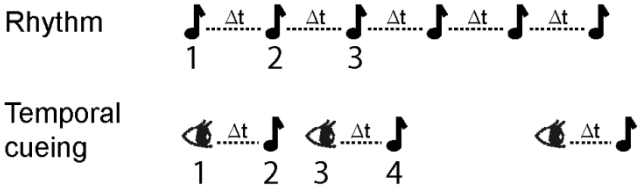


Figure 1. Two different types of temporal associations, rhythms and temporal cueing. Rhythms establish their relation after 3 stimuli (2 delay periods), while temporal cueing only establishes the relation after 4 stimuli (2 delay periods).

The behavioral benefits afforded by either rhythms or temporal cueing have been shown in various studies (Correa, 2010; Jones, Moynihan, MacKenzie, & Puente, 2002; Los, Knol, & Boers, 2001; Niemi & Näätänen, 1981). On the one hand, temporal information guides the motor system to prepare a motor response at the time point of the expected stimulus (Los & Van der Burg, 2013; Niemi & Näätänen, 1981), thereby responding faster. This will have maximal benefit if the time interval for motor preparation is shorter as the temporal cued interval. On the other hand, perceptual processes at the time point of expected stimulus arrival are enhanced, improving the detection (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Lasley & Cohn, 1981; Rohenkohl, Cravo, Wyart, & Nobre, 2012) and discrimination of stimuli (Ellis & Jones, 2010; Jones, et al., 2002; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010). For both motor and perceptual processes behavioral benefits decrease for too long intervals as temporal estimates are less reliable for longer intervals (Luce, 1986; Niemi & Näätänen, 1981; Woodrow, 1914).

From the previous paragraph it is clear that for more reliable temporal estimates behavioral benefits are stronger. However, the reliability of an estimate is not only limited by the inaccuracies of our internal estimation, but also by the external temporal consistency in which sensory input is provided. For example, when throwing a ball at the wall, the time point of arrival is variable, caused by difference in the speed of the ball. In this situation, temporal estimates reflect an interval at which it is likely that the ball will arrive (Luce, 1986; Niemi & Näätänen, 1981). The wider the estimation interval, the less benefit one has from the estimation.

Although influencing the temporal interval of stimulus occurrence decreases the benefit, it is important to realize that temporal estimations can be updated as time passes. This can drastically change the benefits afforded by them. This is nicely demonstrated in experimental settings that have two time points at which targets can occur (Woodrow, 1914; Zahn & Rosenthal, 1966). In these experiments a stimulus can for example occur at either 200 or 1000 ms after a temporal cue. There is a 50% chance that the stimulus arrives at 200 ms or at 1000 ms after the cue. However, as time passes and the stimulus did not occur at 200 ms, the chance of stimulus occurrence in the late interval increases to 100%.

Response times at this late interval are typically faster than for the early interval (Correa, Lupiáñez, Milliken, & Tudela, 2004). This demonstrates that temporal estimations are updated online and can be nicely displayed in a hazard function that shows over time the chance that a stimulus can occur (Correa, et al., 2004; Coull & Nobre, 1998). To counteract temporal estimation updating, it is necessary to introduce catch trials (Correa, Lupianez, & Tudela, 2006) or to incorporate these probabilities in the study design.

Although rhythmicity and temporal cues both require the orienting of attention to a specific point in time, it is not yet clear whether they operate with the same temporal mechanisms (Correa & Nobre, 2008). On the one hand, electroencephalogram (EEG) recordings point to similar neuronal mechanisms as the both temporal structures show similar electrophysiological responses (Correa & Nobre, 2008). Additionally, behavioral benefits are generally in similar magnitude (also see Chapter 2). On the other hand, patients with right frontal lesions seem to be able to perform temporal orienting tasks based on rhythms but not on temporal cueing (Triviño, Arnedo, Lupiáñez, Chirivella, & Correa, 2011). Moreover, rhythmic information seems to overrule temporal cueing when they are both presented at the same time (Ellis & Jones, 2010). Although the exact mechanisms are not clear it thus seems that rhythmic cues have a stronger and dominating effect. It might be easier to infer the temporal relationship in rhythms (as less stimuli are required to understand the temporal statistics in a rhythm, figure 1). Consequently, they might require less higher order control processing (which would explain why patients with right frontal lesions can perform this rhythmic cueing tasks).

Short or long term

Much research has investigated temporal associations by manipulating temporal relations between different stimuli in the lab. This implies that these associations can be learned relatively fast and can be updated almost instantaneously. However, some temporal associations are very systematic in the environment, for example the dropping of a ball with the same weight: if a ball is dropped from the same distance, the time it

takes to fall on the floor is always exactly the same. It is an open question whether the exact timing of these associations are stored in the brain.

There has been extensive research investigating the malleability of processing temporal information during paradigms where participants explicitly have to judge the simultaneity of two presented stimuli [see for a review (Vroomen & Keetels, 2010)]. There is a specific interval in which stimuli are judged to be presented at the same time. This time interval varies for different stimulus types, that is, for simple stimuli such as beeps and flashes the interval is relatively narrow, while for more complex and natural stimuli such as speech this interval is much wider (Vatakis & Spence, 2006; Zampini, Shore, & Spence, 2003). However, the width of the interval and the exact delay at which participants mostly judge the stimulus to be simultaneous can be influenced. For example, the width of the interval can be changed by explicitly giving feedback about the performance to the participants (Powers III, Hillock, & Wallace, 2009). This narrowing of the perceived simultaneity window was maintained even a week after the training. Also during development the integration window significantly narrows (Hillock-Dunn & Wallace, 2012; Lewkowicz & Flom, 2014). Moreover, the point of most perceived simultaneity can be shifted if participants are presented with a constant lag of auditory and visual stimuli (so-called temporal recalibration; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004). This adaptive mechanism seems important in the real world when objects are presented at a distance: in that case visual information is faster as the transduction time through the air is faster for visual than for auditory information (although a definite answer that this mechanism is in place for this reason is lacking). Again, the strength of the recalibration seems stronger for more complex stimuli as for simple stimuli (Roseboom, Kawabe, & Nishida, 2013). Collectively, these studies seem to suggest that temporal associations can be changed rapidly. However, in temporal recalibration experiments the temporal shift of perceived simultaneity does not tremendously change. Typically, the exposure delay is around 240 ms while the shift in perceived simultaneity is in the order of 20-60 ms.

Although explicit temporal associations have been investigated extensively, there has been strikingly little research investigating the flexibility of implicitly learned temporal associations. While it has been

shown that newly presented complex temporal associations can be learned within one training session [see e.g. (Fiser & Aslin, 2002)], it is unknown whether real-world temporal structure – which can be highly consistent – bears the same flexibility. For example, in speech there are systematic delays between mouth movements and speech sounds (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009) that most likely are learned over the lifetime (see Chapter 6). One would predict that these consistencies are coded in the brain and have less plasticity as newly presented stimulus pairs.

Conclusively, research has shown temporal associations can be made very fast and are used to aid perception. However, there is very limited research investigating whether the flexibility by which temporal associations are updated depends on the consistency that these associations have in the environment. Part I of this thesis focusses on the benefits afforded by learning new temporal dynamics within one experimental session. In contrast, part II of this thesis shows that the consistent temporal relation existing between mouth movements and speech sounds seems to be coded more rigidly as this learned temporal association biases the identification of speech.

Temporal statistics in audiovisual settings

Experimental findings

When investigating temporal properties of cross-modal stimuli one has to take into account characteristic temporal differences between sensory modalities in both brain processing time and the transduction time through the air. On the one hand, auditory information is processed much faster as visual information in the brain (Musacchia & Schroeder, 2009), that is, in the maquace brain auditory information arrives in primary auditory cortex 10 ms post-stimulus and visual information arrives in primary visual cortex 35 ms post-stimulus. On the other hand, visual information is transduced through the air much faster as auditory information. Furthermore, auditory information seems to be sampled by the brain at a higher rate as visual information and auditory perception seems to dominate in temporal processing (Fendrich & Corballis, 2001; Repp & Penel, 2002). For example, participants are much better at

judging the simultaneity of two auditory compared to two visual stimuli (Virsu, Oksanen-Hennah, Vedenpää, Jaatinen, & Lahti-Nuuttila, 2008).

Although there are intrinsic and extrinsic timing differences between different senses, cross-modal temporal statistics are easily learned since we quickly adapt to the natural existing timing delays. As mentioned before, the point of perceived simultaneity shifts adaptively to an induced timing delay in the environment (Fujisaki, et al., 2004; Vroomen, et al., 2004). Moreover, participants can learn about specific audio-visual delays and subsequently optimize perception and action (Niemi & Näätänen, 1981). Interestingly, there is higher sensitivity for events in which visual cues precede auditory cues (Fujisaki, et al., 2004; Vroomen & Stekelenburg, 2011). For example, maximal integration in brain responses in superior colliculus have been reported when visual cues precedes auditory stimuli with 50 ms (Wallace, Wilkinson, & Stein, 1996). Moreover, the width of the window in which participants perceive an audio-visual stimulus pair to be synchronous is typically broader on the side at which visual information precedes auditory information and the highest synchrony of audiovisual input is perceived when vision precedes audition (Vroomen & Stekelenburg, 2011). Yet, almost exclusively the visual preceding side of the temporal integration window can be narrowed (Fujisaki, et al., 2004) while for auditory preceding stimuli this is less easy. This seems logical as in the natural environment we are typically exposed to auditory lagging events, partly because auditory information transduces slower, but also because usually some physical motion (which is perceived by vision) has to occur before any sound pattern can be produced.

Audio-visual Speech

Speech is full of temporal dynamics that contain unimodal rhythmic as well as cross-modal temporal cues. For example, the production of syllables typically lasts 200-250 milliseconds and therefore speaking seems to occur at a 3-9 Hz rhythm, so-called theta. The adding of visual information significantly improves speech perception; it has been shown multiple times that seeing a speaker improves the intelligibility of a noisy auditory stream (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). Visual mouth movements seem to typically precede auditory speech

sounds and therefore act as a cue for the onset time of the auditory speech [(Chandrasekaran, et al., 2009) but also see Schwartz and Savariaux, 2014]. Consequently, the listener can optimally focus at time points in which auditory speech is most informative.

Simultaneity judgments of syllables indicate that the temporal binding window of audiovisual speech is rather broad [up to 400 ms wide (Dixon & Spitz, 1980; Massaro, Cohen, & Smeele, 1996)]. This seems to indicate that audiovisual speech is temporally not very specific. Conversely, syllables have a unique visual-to-auditory delay that is syllable specific (Chandrasekaran, et al., 2009). This delay seems to fit the width of the binding window, ranging between 50 and 250 ms. It is therefore conceivable that the width of the temporal binding window is a consequence of the temporal statistics in the environment in which different syllables can occur. This however does not directly imply that all temporal information within this window is lost, but merely the percept of synchrony. This is partly confirmed by a shorter width of this window when participants are forced to judge whether vision of auditory occurred first (Vroomen & Stekelenburg, 2011). Moreover, it seems that automatic perceptual processes have more access to temporal information compared to our conscious knowledge (Repp, 2000).

If information about the unique visual-to-auditory delay of specific syllables is available it should contain cues about the content of the upcoming stimuli. Indeed, some studies show that participants can learn content representations dependent on the time point of presentation relative to a cue (Hamid, 2014; Hamid, Wendemuth, & Braun, 2010). However, there is only limited research especially in the field of speech and multisensory processes investigating the hypothesis that temporal information about cue-target delays could constrain the specific content in the target stimulus based on their stimulation history. In Chapter 6 we show that indeed participants use the unique delay between mouth movements and speech sounds to guide their syllable perception.

Using oscillations to code temporal statistics

Neurons function through electrical activity. This activity is governed by positively and negatively charged ions present within the cells and in the

extracellular space that transfer in and out of the cell. Neurons actively maintain an electrical potential of around -65 mV with the extracellular space. Communication between neurons is initiated when many positively charged ions enter the cell and the potential reaches -20 mV. This is the threshold for an action potential, which is a cascade of events in which the neuron first depolarizes (getting closer to 0 mV potential and even above it), then repolarizes to the resting potential, hyperpolarizes (< -65 mV), and then turns back to the original resting potential (figure 2A). In this process the neurons sends an electric signal ('fire') to other neurons through its axon, thereby changing the potential of the receiver neuron. This change can either be positive (excitatory cell) or negative (inhibitory cell), moving the resting potential closer or further away from -65 mV.

Neuronal ensembles seem to naturally operate in an oscillatory pattern. These oscillations reflect the collective shifts in membrane potentials of cells relative to the extracellular space moving around the resting potential (Buzsáki & Draguhn, 2004). Let's imagine a neuronal population of excitatory neurons that has a stable resting potential and a

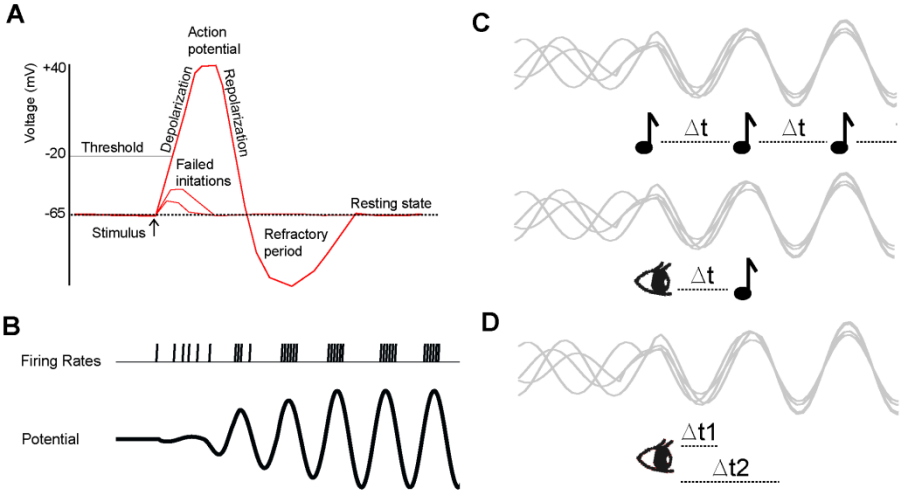


Figure 2. Oscillatory mechanisms. A) Action potential. B) Firing rates are clustered at the most excitable phase of an oscillation. C) Rhythms (top) and temporal cues (bottom) are better perceived when aligning the stimuli with excitable phases of ongoing oscillations. D) Temporal associations might cause natural categorization of specific stimuli by aligning specific phases with specific stimulus types.

constant in input to the full population. Due to small random fluctuations one neuron might reach an action potential (figure 2B). This will induce a short increase in the membrane potential (due to the depolarization) and will subsequently be followed by hyperpolarization. Directly surrounding cells are influenced by this firing as more positively charged ions will be present in the extracellular space during the hyperpolarization, effectively also partly hyperpolarizing the surrounding cells. Therefore, the resting potential for a collection of cells will be in a hyperpolarized (more negative) state and cells are less likely to reach the threshold for an action potential during this period if the input is not strong enough. Consequently, surrounding cells are more likely to fire after this inhibitory period is finished (if enough input is provided). Once cells start collectively firing after the inhibitory period is over the excitatory and inhibitory strength of the neuronal ensemble will grow. This will make the firing of cells even more clustered after the next inhibitory period. This collective inhibition and excitation makes any oscillation a self-sustaining process as long as there is no disruptive strong input.

Although in theory, excitatory neurons (as described above) could induce an oscillation due to the properties of the action potential, the action potential only influences a minimal amount of neighboring cells and interactions with input patterns in a dynamic system of firing cells disrupts the sustainment of this type of oscillation. Instead, oscillations predominantly occur when there is stronger and more sustained inhibition that is more widespread (Van Vreeswijk, Abbott, & Ermentrout, 1994). Therefore, in most settings inhibitory cells are the primary force of oscillations. When inhibitory cells fire, they directly create a hyperpolarized membrane potential at receiving (mostly excitatory) neurons. This is the main inhibitory force which makes the excitatory cells cease to fire. This inhibition causes a much stronger hyperpolarization as the hyperpolarization caused by the action potential of the excitatory cells alone. As the strength of the inhibitory period increases, more neurons collectively fire after the strong inhibition is finished, significantly increasing the strength of the oscillation (Buzsáki, 2004). Moreover, many inhibitory neurons inhibit a more widespread area of cortex and it has been shown that specifically these long-ranging inhibitory neurons increase the strength of the oscillation (Buzsáki, Geisler, Henze, & Wang, 2004). The exact properties of inhibitory cells

and the way they interact with excitatory neurons in a neuronal ensemble determine the duration of the inhibitory period and thereby the frequency of an oscillation. As many different inhibitory cells have been identified and there combinations are almost endless oscillations can be adapted differentially.

Rhythms and temporal cueing: optimal phase

As described above, oscillations reflect a natural temporally varying property of the brain and it therefore seems efficient to code temporal information on these oscillations [see e.g. (Buhusi & Meck, 2005; Karmarkar & Buonomano, 2007; Pöppel, 1997; Schroeder & Lakatos, 2009; VanRullen & Koch, 2003), but see e.g. (Ivry & Schlerf, 2008) for another account or (Muller & Nobre, 2014) for an elaborative review]. Much work has focused on investigating these patterns during rhythmic stimulation (Besle et al., 2011; Cravo, et al., 2013; Henry & Obleser, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Thorne, De Vos, Viola, & Debener, 2011). When there is a temporal expectancy of stimulus occurrence it might be useful to have high sensitivity of detecting this stimulus. This could theoretically be achieved by aligning neuronal ensembles to the external input at the phase of the oscillation closest to the threshold for firing as there is a higher chance that an impulse from an outside stimulus will reach an action potential (figure 2C).

When stimuli are presented at a specific rhythm strong oscillatory patterns develop as evoked (excitatory) responses are induced for every impulse, effectively entraining the neuronal ensembles to the external stream of events. This changes the phase of the ongoing oscillatory patterns in the brain to the external stimulation as excitatory events in the cortex are aligned to strong input arriving in the cortex (Makeig, Debener, Onton, & Delorme, 2004). These patterns are induced differentially compared to the natural ongoing oscillations in the brain described in the previous section as they arise due to external input and not through the ongoing interaction between excitatory and inhibitory processes in the brain. However, they seem to preserve some of the same functional properties. For example, sorting trials based on the phase of ongoing oscillations shows that there is an optimal phase of detection

(Busch, Dubois, & VanRullen, 2009). This optimal phase also exists when aligning oscillatory phases to externally presented stimuli (Lakatos, et al., 2008). This shows the behavioral benefits afforded by temporal information, especially in rhythms. When the optimal phase co-occurs when stimuli are expected, these stimuli are processed more efficiently. The entrainment and subsequence benefit of presenting a stimulus at the optimal phase has been shown repeatedly (Cravo, et al., 2013; Henry & Obleser, 2012; Mathewson, et al., 2010).

The properties of the entrainment stimuli and task seem to determine which neuronal ensembles get entrained. For example, Lakatos and colleagues (2013) presented pure tone stimuli at a presentation rate of 1.5 Hz. Only neuronal ensembles that were selective to the specific frequency of the tone showed entrainment. Interestingly, other neuronal ensembles showed entrainment at the opposite phase of the presentation rate, effectively being inhibited at the time point at which stimuli occurred. This indicates that the processing of both the pitch and the time point of the target tone was enhanced, revealing the selectivity such entrainment can induce. The phase and strength of the entrained oscillation can also be influence by attentional resources (Besle, et al., 2011; Lakatos, et al., 2008). When not only tones, but also visual stimuli are presented in anti-phase of a 1.5 Hz stimulus stream, the alignment of ongoing oscillations changes depends on whether attention is directed to the visual or auditory stimuli (Lakatos, et al., 2008). Moreover, the strength of this entrainment is influenced by the likelihood of a target stimulus occurring (Stefanics et al., 2010).

While for rhythms it is intuitive that oscillations can align to the temporal input they provide, for temporal cueing this relation is less straightforward. However, already for a single impulse, ongoing oscillations seem to align to one specific phase point (Makeig, et al., 2004; Mercier et al., 2015). Timing of subsequently presented stimuli relative to the new induced phase will determine whether the stimulus will arrive at a high or low excitable phase (figure 2C) and the strength of the oscillation determines the behavioral benefits afforded at the optimal phase (Fiebelkorn et al., 2013). If the phase of the oscillation is vital for detection, one could predict that the exact phase to which the cue aligns the oscillations might changes dependent on the time point at which a target is expected, to ensure that the optimal phase would overlap with

the expected time of the target. In a behavioral study Fiebelkorn et al. (2011) show that the point of maximum behavioral benefit in a foreperiod paradigm indeed depends on the time point that participants expect a stimulus. However, brain data linking phase effects to behavioral benefits afforded by temporal cueing seems to be limited (see e.g. Thorne et al., 2011; Lakatos et al., 2009).

Other foreperiod period paradigms have shown an overall decrease in alpha power (9-13 Hz) at expected time points (Rohenkohl & Nobre, 2011). This power decrease is associated with the role of alpha oscillation in the brain to inhibit the activity of specific areas (Jensen, Bonnefond, & VanRullen, 2012; Pfurtscheller, Stancak Jr, & Neuper, 1996). But no differential phase reset has been shown (but also see Chapter 5). On the one hand, it might be that there is only one specific period to which an impulse can reset oscillations. On the other hand, the temporal difference used for most foreperiod paradigms is often over a second which might be a too long interval to see these effects.

Parsing of information

The enhancement of processing at a specific point on the oscillation is not only beneficial to optimize perception at that specific time point, it also provides a means to separate information from each other, thereby parsing the incoming information in separate chunks (VanRullen & Koch, 2003). This could be useful as the storage of continuous data might be inefficient and the separation of information in continuous data could be very difficult. It therefore seems useful to sample the environment during bursts of high excitable periods of oscillations instead of storing a continuous stream. This is illustrated by the continuous wagon wheel illusion: in movies we typically perceive the motion direction of a fast spinning car wheel as moving backwards. This occurs as the sampling frequency of the sequential images in the movie is too slow to capture the fast dynamics. In daily life we sometimes also perceive this backward moving wheel. VanRullen, Reddy, and Koch (2006) have shown that this illusion is related to the strength of ongoing oscillations around 13 Hz. They propose that information might be subsampled at 13 Hz, therefore evoking the illusion especially when power is high. As such this

subsampling serves as a temporal reference frame to code information in separate chunks.

Another good illustration of the sampling of information seems to be in the tracking of speech. Speech is composed of different levels of information. At the highest level are there the sentences that are composed of words which are subsequently composed of syllables and phonemes (smallest units of speech sound pronounceable). To understand a word first the phonemes and syllables have to be extracted and understood. Therefore, continuous speech needs to be parsed in useful information. Syllables last around 250 ms and tracking of this information therefore has to occur at a 4 Hz rate. Brain responses to continuous speech do occur around this theta range (Giraud & Poeppel, 2012; Zion Golumbic et al., 2013). More importantly, the response is reduced when the speech is played backward (Peña & Melloni, 2012). However, it is at this moment difficult to dissociate the theta response as an evoked response to meaningful stimuli, or a response induced as a prediction mechanism to track the relevant moments of the rhythmic speech. It is however clear that this range is important for the intelligibility of speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

Phase coding

One of the strongest ongoing oscillatory patterns in the brain is the hippocampal theta oscillation around 4-9 Hz [see e.g. (Vinogradova, 1995)]. Hippocampal theta oscillations are coupled to higher frequency gamma oscillations such that the phase of the theta oscillations determines the amplitude of the gamma oscillations (Bragin et al., 1995; Soltesz & Deschenes, 1993). When a rat has to move through a maze different neuronal populations coding for different locations in space (place cells) fire at a specific phase of the theta oscillation (O'Keefe & Recce, 1993). The order of the firing is in sequence with the order at which the rat will move through the maze, such that place cells coding the first upcoming location will fire directly after the inhibitory period of the theta oscillation. When a rat does not know the room, this pattern is absent. Therefore, this pattern of firing has been associated with memory traces representing a stored memory pattern of sequential upcoming locations [for a review see (Buzsáki & Moser, 2013)]. As soon as the rat

starts moving along the planned path the firing pattern changes. Since the rat moves, place cells will start firing earlier and earlier on the oscillation cycle as the location they are coding for is closer to the path of the rat. This process is called phase precession. In this way the rat codes the upcoming points in space through a phase code.

The idea of this type of phase coding seems attractive and more widespread over the brain. For example, Kayser, Ince, and Panzeri (2012) have shown that adding information about theta phase significantly improves the discriminability of spike trains coding for different representations recorded from primary auditory and primary visual cortical areas. Multiple other studies have also shown the added value of phase information for categorization (Rey, Fried, & Quiroga, 2014; Turesson, Logothetis, & Hoffman, 2012; Lopour et al., 2013). The idea of temporal segregation of representations is useful as neurons coding for one representation could be located in distant areas and only through temporal coherence form a code (Fries, 2005; Singer, 2009). Moreover, neurons can be dynamically engaged depending on the context, thereby enforcing a flexible coding scheme.

It is an open question what the organization principles of phase coding might be as it does not follow directly why specific stimuli should be encoded on specific phases. For the place cells the first upcoming closest location is encoded 'earliest' on the theta phase. This early phase point corresponds to the earliest point in the theta phase at which an action potential might occur and at which the gamma oscillation has only a moderate amplitude. In this coding the most salient representation (the one that is the closest in time) has a strong enough impulse to be encoded on a relatively low excitable level on the theta oscillation. This coding scheme has recently also been proposed for cortical oscillations (Jensen, Gips, Bergmann, & Bonnefond, 2014). However, for some representations it might not be so clear why one representation would be more salient as the other and this might require another type of coding.

Phase coding of time

The concept of phase precession and place cells seems closely related to temporal processing. As discussed above, temporal information can be viewed as the reference to a sequence of events and the information

about time is only defined relative to this sequence. The coding of location on hippocampal theta phase is exactly this: the temporal order in which the rat will pass locations in the future. This type of coding does not have a direct relationship to exact temporal parameters, as the phase difference between the sequential locations does not have to correspond to the exact temporal difference at which the rat passes the subsequent locations (as this is dependent on the velocity of the rat). However, in theory it would be possible to have a direct mapping between temporal intervals and phase differences. Yet, there is not much evidence showing that the temporal relation between two stimuli is encoded in phase differences. In an elegant study Kosem, Gramfort, & van Wassenhove (2014) have shown in an audiovisual recalibration experiment that after recalibration the phase relation between slow delta oscillations in the auditory cortex is shifted toward the perceived simultaneity difference. There was a strong correlation between the shifted time point of perceived simultaneity and the phase different. This study indicates that perceived timing matches neuronal oscillatory timing.

If phase reset consistently occurs after an initial cue [see e.g. (Makeig, et al., 2004)], any target occurring after the cue will always be presented at a specific point on the oscillatory phase. This could provide an intuitive way to code temporal information on phase and also ensure that different representations are clearly separated from each other (figure 2D). Up to date, this hypothesis needs to be verified (but see Chapter 7 and 8).

Outline of Thesis

The work described in the current thesis focusses on how temporal information is used to guide behavior. On the one hand, in a dynamic environment temporal statistics are learned fast and implicitly to attend to specific moments in time. On the other hand, consistent temporal statistics between stimulus pairs are encoded in the brain to guide perception in the future. This thesis focusses on both types of perceptual influences.

Part I of this thesis investigates temporal statistics that are acquired within one experimental session and uses behavioral as well as brain data

to investigate how temporal information is used to optimize our perception. In chapter 2 we investigate whether adding two types of temporal cues, rhythmicity and temporal cueing, can improve detection performance more than the benefits afforded by an individual temporal cue on its own. The benefit acquired via rhythmicity is further investigated in chapter 3, in which we show how low intensity sounds become audible by resonating brain response to sub-threshold, not yet audible stimuli. This resonance property seems to depend on the expectations of participants that a rhythmic stream will continue, as is shown in chapter 4. This chapter demonstrates that temporal information is not only extracted from the immediately preceding temporal cues, but also from broader contextual information. The active role of the brain in attending to time is further explored in chapter 5 in which we show that low frequency oscillatory patterns are used to attend to stimuli that have a more uncertain temporal occurrence.

Part II of this thesis focusses how the consistent temporal relationship between the onset of mouth movements and speech sounds influences behavioral and brain mechanisms for perceiving speech. In chapter 6 we show that there exists a consistent temporal relationship between the onset of mouth movements and specific syllable types and that this information aids identification of these syllables. Chapter 7 demonstrates that this consistent relationship expresses itself in the neuronal coding of syllables: oscillatory phase biases syllable identification, suggesting that different syllables are preferentially processed at one specific phase. This was demonstrated using both EEG patterns as well as an entrainment paradigm. We verify in chapter 8 with functional magnetic resonance imaging (fMRI) that phase indeed seems to be part of the representation of these syllables. In this chapter we show increased information about syllable identity when syllables are presented at their “preferred” phase.

References

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience, 31*(9), 3176-3185.

Bragin, A., Jandó, G., Nádasdy, Z., Hetke, J., Wise, K., & Buzsáki, G. (1995). Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *The Journal of Neuroscience, 15*(1), 47-60.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience, 6*(10), 755-765.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience, 29*(24), 7869-7876.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience, 7*(5), 446-451.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*(5679), 1926-1929.

Buzsáki, G., Geisler, C., Henze, D. A., & Wang, X.-J. (2004). Interneuron diversity series: circuit complexity and axon wiring economy of cortical interneurons. *Trends in Neurosciences, 27*(4), 186-193.

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience, 16*(2), 130-138.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436.

Correa, A. (2010). Enhancing behavioural performance by visual temporal orienting. Attention and time, 357-370.

Correa, A., Lupiáñez, J., Milliken, B., & Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Perception & Psychophysics, 66*(2), 264-278.

Correa, A., Lupianez, J., & Tudela, P. (2006). The attentional mechanism of temporal orienting: Determinants and attributes. *Experimental Brain Research, 169*(1), 58-68.

Correa, A., & Nobre, A. C. (2008). Neural modulation by regularity and passage of time. *Journal of Neurophysiology, 100*(3), 1649-1655.

Coull, J., & Nobre, A. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology, 18*(2), 137-144.

Coull, J., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience, 18*(18), 7426-7435.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics, 72*(8), 2274-2288.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*(4), 719-725.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *The Journal of Neuroscience, 31*(27), 9971-9981.

Fiebelkorn, I. C., Snyder, A. C., Mercier, M. R., Butler, J. S., Molholm, S., & Foxe, J. J. (2013). Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage, 69*, 126-137.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*(7), 773-778.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517.

Hamid, O. H. (2014). *The role of temporal statistics in the transfer of experience in context-dependent reinforcement learning*. Paper presented at the Hybrid Intelligent Systems (HIS), 2014 14th International Conference on.

Hamid, O. H., Wendemuth, A., & Braun, J. (2010). Temporal context and conditional associative learning. *BMC neuroscience, 11*(1), 45.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, 109*(49), 20095-20100.

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science, 15*(5), 688-696.

Hudspeth, A. J. (1989). How the ear's works work. *Nature, 341*(6241), 397-404.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences, 12*(7), 273-280.

James, W. (1886). The perception of time. *The Journal of speculative philosophy*, 374-407.

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences, 16*(4), 200-206.

Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences, 37*(7), 357-369.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313-319.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron, 53*(3), 427-438.

Kayser, C., Ince, R. A., & Panzeri, S. (2012). Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS computational biology, 8*(10), e1002717.

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*(4), 750-761.

Lasley, D. J., & Cohn, T. (1981). Detection of a luminance increment: effect of temporal uncertainty. *JOSA, 71*(7), 845-850.

Lewkowicz, D. J., & Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development, 85*(2), 685-694.

Los, S. A., Knol, D. L., & Boers, R. M. (2001). The foreperiod effect revisited: Conditioning as a basis for nonspecific preparation. *Acta psychologica, 106*(1), 121-145.

Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1612.

Luce, R. D. (1986). *Response times:* Oxford University Press.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-141.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences, 8*(5), 204-210.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America, 100*, 1777.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition, 115*(1), 186-191.

Mercier, M. R., Molholm, S., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Foxe, J. J. (2015). Neuro-Oscillatory Phase Alignment Drives Speeded Multisensory Response Times: An Electro-Corticographic Investigation. *The Journal of Neuroscience, 35*(22), 8546-8557.

Muller, T., & Nobre, A. C. (2014). Perceiving the passage of time: neural possibilities. *Annals of the New York Academy of Sciences, 1326*(1), 60-71.

Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing research, 258*(1-2), 72-79.

Newton, I., Motte, A., & Cajori, F. (1987). *Mathematical principles of natural philosophy:* W. Benton: Encyclopaedia Britannica.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*(1), 133.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*(3), 317-330.

Peña, M., & Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience, 24*(5), 1149-1164.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238.

Pfurtscheller, G., Stancak Jr, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. *International Journal of Psychophysiology, 24*(1), 39-46.

Pöppel, E. (1997). A hierarchical model of temporal perception. Trends in cognitive sciences, 1(2), 56-61.

Powers III, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. T*he Journal of Neuroscience, 29*(39), 12265-12274.

Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research, 63*(2), 106-128.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1085.

Rey, H. G., Fried, I., & Quiroga, R. Q. (2014). Timing of single-neuron and local field potential responses in the human medial temporal lobe. *Current Biology, 24*(3), 299-304.

Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *The Journal of Neuroscience, 31*(40), 14076-14084.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience, 32*(24), 8424-8428.

Roseboom, W., Kawabe, T., & Nishida, S. Y. (2013). Audio-visual temporal recalibration can be constrained by content cues regardless of spatial overlap. *Frontiers in Psychology, 4.*

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schwartz, J.-L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Compututational Biology, 10*(7), e1003743

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Singer, W. (2009). Distributed processing and temporal codes in neuronal networks. *Cognitive neurodynamics, 3*(3), 189-196.

Soltesz, I., & Deschenes, M. (1993). Low-and high-frequency membrane potential oscillations during theta activity in CA1 and CA3 pyramidal neurons of the rat hippocampus under ketamine-xylazine anesthesia. *Journal of Neurophysiology, 70*(1), 97-116.

Stefanics, G., Hangya, B., HernÃ¡di, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of Neuroscience, 30*(41), 13578-13585.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience, 31*(10), 3853-3861.

Triviño, M., Arnedo, M., Lupiáñez, J., Chirivella, J., & Correa, Á. (2011). Rhythms can overcome temporal orienting deficit after right frontal damage. *Neuropsychologia, 49*(14), 3917-3930.

Turesson, H. K., Logothetis, N. K., & Hoffman, K. L. (2012). Category-selective phase coding in the superior temporal sulcus. *Proceedings of the National Academy of Sciences, 109*(47), 19438-19443.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? Trends in cognitive sciences, 7(5), 207-213.

VanRullen, R., Reddy, L., & Koch, C. (2006). The continuous wagon wheel illusion is associated with changes in electroencephalogram power at ~13 Hz. *The Journal of Neuroscience, 26*(2), 502-507.

Van Vreeswijk, C., Abbott, L., & Ermentrout, G. B. (1994). When inhibition not excitation synchronizes neural firing. *Journal of computational neuroscience, 1*(4), 313-321.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain research, 1111*(1), 134-142.

Vinogradova, O. (1995). Expression, control, and probable functional significance of the neuronal theta-rhythm. *Progress in neurobiology, 45*(6), 523-583.

Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., & Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Experimental Brain Research, 186*(4), 525-537.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871-884.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32-35.

Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition, 118*(1), 75-83.

Wallace, M., Wilkinson, L., & Stein, B. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246-1266.

Woodrow, H. (1914). The measurement of attention. *The Psychological Monographs*, 17(5), i.

Zahn, T. P., & Rosenthal, D. (1966). Simple reaction time as a function of the relative frequency of the preparatory interval. *Journal of Experimental Psychology, 72*(1), 15.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198-210.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron, 77*(5), 980-991.

CHAPTER 2

RHYTHMICITY AND CROSS-MODAL TEMPORAL CUES FACILITATE DETECTION

Corresponding Manuscript:
Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion Golumbic, E. M. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia, 63*, 43-50.

Abstract

Temporal structure in the environment often has predictive value for anticipating the occurrence of forthcoming events. In this study we investigated the influence of two types of predictive temporal information on the perception of near-threshold auditory stimuli: 1) intrinsic temporal rhythmicity within an auditory stimulus stream and 2) temporally-predictive visual cues. We hypothesized that combining predictive temporal information within- and across-modality should decrease the threshold at which sounds are detected, beyond the advantage provided by each information source alone. Two experiments were conducted in which participants had to detect tones in noise. Tones were presented in either rhythmic or random sequences and were preceded by a temporally predictive visual signal in half of the trials. We show that detection intensities are lower for rhythmic (vs. random) and audiovisual (vs. auditory-only) presentation, independent from response bias, and that this effect is even greater for rhythmic audiovisual presentation. These results suggest that both types of temporal information are used to optimally process sounds that occur at expected points in time (resulting in enhanced detection), and that multiple temporal cues are combined to improve temporal estimates. Our findings underscore the flexibility and proactivity of the perceptual system which uses within- and across-modality temporal cues to anticipate upcoming events and process them optimally.

Introduction

Increasingly, the brain is thought of as intrinsically proactive, not merely relying on bottom-up sensory information to interpret perceptual information. Instead, even low-level sensory cortices are thought to be constantly creating and updating internal models of the external world, to anticipate and predict upcoming events (Bar, 2011; Friston, 2011; Nobre, Correa, & Coull, 2007; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Schubotz, 2007; Summerfield & Egner, 2009; Summerfield et al., 2006). In addition to predicting the content of upcoming stimuli - e.g. features or location – recent research indicates that anticipating the timing of upcoming sounds significantly improves perceptual judgement. Specifically, at least two types of temporal expectations are shown to improve behavioural performance: Rhythmic regularity within a stimulus sequence decreases reaction times and improves accuracies of responses to supra-threshold stimuli when target stimuli occur at an anticipated moment, compared to stimuli occurring randomly or at unanticipated times (Ellis & Jones, 2010; Jones, Moynihan, MacKenzie, & Puente, 2002; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010; Niemi & Näätänen, 1981), as well as improving stimulus sensitivity (Rohenkohl, Cravo, Wyart, & Nobre, 2012). In addition, temporal cueing within- and across modalities has been used extensively to show that a constant time-interval between a cue and target can improve the speed of target detection (Correa, Lupiáñez, Milliken, & Tudela, 2004; Coull & Nobre, 1998; Lange & Röder, 2006) and recognition (Griffin, Miniussi, & Nobre, 2001) by means of temporal preparation (Los & van den Burg, 2013). In particular, visual cues appear to be a natural temporal cue for audition (Thorne & Debener, 2008; Van Wassenhove, Grant, & Poeppel, 2005, 2007). A prominent example is speech, since observed lip movements and facial gestures are temporally correlated with, and precede, the auditory input (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Ten Oever, Sack, Wheat, Bien, & Van Atteveldt, 2013; Van Wassenhove, et al., 2005, 2007). Moreover, lip movements and facial gestures have intrinsic rhythmic regularities (Giraud & Poeppel, 2012; Greenberg, Carvey, Hitchcock, & Chang, 2003; Luo, Liu, & Poeppel, 2010; Zion Golumbic, Poeppel, & Schroeder, 2012). Thus, in natural situations, such as speech,

we are faced with intermixed temporal information to predict upcoming events, provided by cross-modal as well as rhythmic temporal cues.

The behavioral advantages afforded by these two types of temporal expectations – stimulus rhythmicity and cross-modal temporal cueing – imply that attentional resources can be dynamically allocated to points in time when input is expected (Jones, Johnston, & Puente, 2006; Jones, et al., 2002; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Large & Jones, 1999; Nobre, et al., 2007; Nobre & Coull, 2010). However, it is not clear whether multiple types of cues are used jointly to improve temporal prediction and optimally allocate attention. Since many naturalistic stimuli, such as speech, music and biological motion combine both cross-modal temporal cues and intrinsically rhythmic properties (Zion Golumbic, et al., 2012), investigating the joint contribution of temporal cues from these two sources bears substantial ecological relevance.

Here, we used two complementary auditory detection paradigms to investigate the influence of temporal cues on threshold intensities, since the above-described 'attention in time' framework predicts that reliable temporal prediction can enhance perceptual sensitivity to subtle stimuli. We manipulated both the temporal structure within the sound stream as well as the presence of cross modal (visual) cues, and investigated the influence of each cue on detection intensities, as well as the combination of both cues. Our hypothesis was that both types of temporal predictions – rhythmicity and cross-modal cueing - would lower sound detection intensities. Rhythmic prediction during the auditory only conditions might not have a strong effect on detection thresholds since, by definition, sounds are "below threshold" before participants indicate that they have heard them. Adding visual input could significantly improve the rhythm percept, thus enriching the temporal prediction. Therefore, we expect an interaction effect in which the combination of cross-modal and rhythmic temporal cues would provide the lowest detection thresholds (Trommershauser, Kording, & Landy, 2011).

Materials and Methods

Participants

Twelve volunteers participated in Experiment 1 (age 20-40; average age: 23.5, 5 male) and twenty volunteers participated in Experiment 2 (age 21-33; average age 25.4, 7 male). All had normal or corrected to normal vision. Informed consent was obtained before the study, which was approved by the New York University Committee on Activities Involving Human Subjects (NYU UCA/HS; Experiment 1) and by the Local Ethical Committee at the Department of Psychology and Neuroscience at the Maastricht University (Experiment 2). Participants were randomly selected and were unaware of the purpose of the study during the experiment. For taking part in the experiment participants received monetary compensation.

Stimulus material

Auditory stimuli were sinusoidal 1 kHz beeps of 50ms duration (including a linear rise and fall time of 5ms) embedded in continuous white noise (53 dB) and presented diotically via headphones (Sennheiser HD 380 Professional, Sennheiser Electronic Corporation, Wedemark, Germany in Experiment 1, Sennheiser HDM25-1 in Experiment 2). The visual stimuli were Gaussian white circles of 50ms duration (generated using the Gaussian generator of the Visual Stimulus Generation Toolkit implemented in the software Presentation used for stimulus delivery, with parameters: mu = -10 and sigma = 60; Neurobehavioral Systems, Inc., Albany, NY), presented foveally on a gray background (rgb: 115,115,115). The visual angle of the Gaussian was 3.1 degrees (corresponding to the width of the 95% contrast interval relative to the center intensity). Both experiments were run in dimly lit sound shielded rooms and participants were seated approximately 57 cm from the screen.

Experimental procedure

In order to investigate the influence of temporal cues on auditory detection we ran two experiments, using complementary approaches for evaluating detection thresholds.

Experiment 1: In the first experiment we employed the "method of limits" approach to evaluate perceptual thresholds (Gescheider, 1997), using an 'increasing' paradigm followed by a 'decreasing' paradigm. In the 'increasing' paradigm participants heard a stream of auditory beeps embedded in continuous white noise (figure 1).

The signal to noise ratio (SNR) of the tone targets was initially below threshold, and the intensity of the tones increased monotonically over the trial. Participants were asked to indicate via button press when the target signals were first detected. In the first four trials, the starting SNR was 0.25 % (none of the participants were able to detect the stimulus with this SNR). SNR was defined as the maximal amplitude in the presented sound divided by the maximal amplitude of the white noise. In subsequent trials, the starting intensity was set to be 7.5% SNR lower than the lowest intensity previously-detected, and this level was

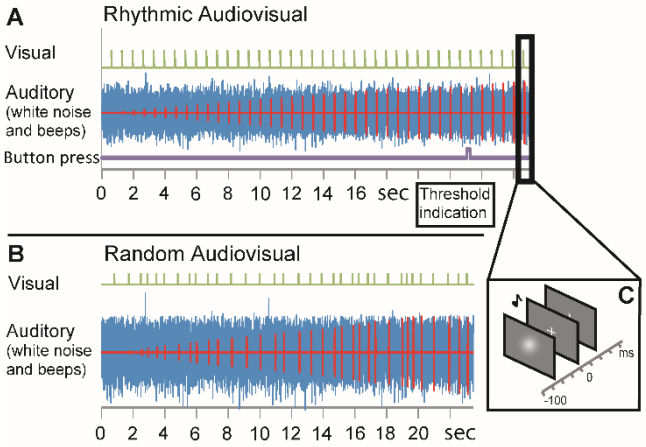


Figure. 1. Illustration of a trial in the rhythmic audiovisual condition (A) and a trial in the random audiovisual condition (B), both in the 'increasing' paradigm. In the auditory channel, beeps (red) were embedded in white noise (blue), with their intensity increasing monotonically over the trial. In the audiovisual conditions, a white Gaussian circle was presented 65 ms prior to each beep (C). The button press (purple) indicates the moment that the participant indicates hearing the sound for the first time.

monitored throughout the experiment to ensure a minimum of 5% SNR difference with the lowest detected intensity judgment. Over the trial, sound intensities increased incrementally in steps of either 0.5 or 1% SNR. The two different incremental steps were randomized to ensure that the sequence of sounds and length of the trials were not identical across trials. After participants indicated detection of auditory stimuli, 4-6 additional beeps were presented at the same intensity level. The 'decreasing' paradigm paralleled the 'increasing' paradigm, but the sounds started well above detection threshold and decreased in intensity over the trial. Participants had to indicate when they could no longer hear the sounds. Here too, the first four trials were used to determine the individual starting intensities per trial (starting intensity of the first four trials was 17.5% SNR), and ensured that the starting intensity was at least 5% above the highest intensity of the detection judgment.

We manipulated the temporal structure of each trial by changing the inter-stimulus interval (ISI) between the tones. In half of the trials there was a constant ISI of 666 ms (Rhythmic condition), whereas in the other half the ISI was randomized among one of 21 evenly spaced time points between 300 and 1000 ms, maintaining an average ISI of 666 ms (Random condition). Additionally, in half of the trials the Gaussian white circle preceded every auditory stimulus, with a fixed audio-visual stimulus onset asynchrony (SOA) of 65 ms (AudioVisual condition). We choose this interval since it has previously been shown to give optimal cross-modal effects for audiovisual tasks (Thorne & Debener, 2008). Thus, in total there were four conditions: Random Auditory (RaAu), Rhythmic Auditory (RhAu), Random AudioVisual (RaAV), and Rhythmic AudioVisual (RhAV). Designing the paradigm in this way served the purpose of implementing a distinct rhythmic or random temporal structure to a continuous stream of stimuli, which is closer to natural listening conditions. It also mimics natural situations in which visual information is salient, but auditory stimuli vary in intensity over time, for example when listening to a person in a noisy environment. In all conditions, participants were explicitly instructed to maintain fixation on a gray cross in the middle of the screen when no visual input was presented. Trials were randomized across conditions (20 trials per condition) and the experiment was divided in four blocks of

approximately seven minutes each. After every block participants were encouraged to take a break.

Experiment 2: One drawback of Experiment 1 is that perceptual thresholds calculated using the method of limits approach confounds perceptual sensitivity with response-bias (Green & Swets, 1966). To control for the possibility that the results of Experiment 1 were driven purely by response-bias, we ran a second experiment, using the same four conditions, in which detection thresholds were determined using a staircase procedure in a two alternative forced choice task (2AFC; Green & Swets, 1966). In this task participants were presented sequentially with two 3-second-duration intervals of white noise at 53 dB. In one of the intervals (randomly chosen on every trial) five sounds were embedded in the noise, and the subjects were instructed to indicate in which interval (first or second) they heard the sounds The four conditions were the same as in Experiment 1 (RaAu, RhAu, RaAV, and RhAV). For the two visual conditions, both intervals contained visual stimuli. Since it is difficult to create a temporally-random sequence of stimuli within a finite interval of 3 seconds, a constant set of 5 ISIs was chosen (350, 500, 814, and 1000 ms) to maximize the temporal variability within each trial. The order of these ISI was randomized in each trial. In addition, in all conditions the onset of the first sound was jittered between 200, 300, 400 and 500 ms after the white noise onset, to reduce expectation effects.

To obtain a measure of the detection threshold we implemented four independent weighted staircase procedures (Kaernbach, 1991) in which the order of the conditions was randomized. In these procedures, every correct response led to a decrease in sound intensity in the next trial of the same condition and every incorrect response led to an increase in sound intensity. Since correct responses can be achieved via 1) actually hearing the stimulus or 2) guessing, the decrease in sound intensity was three times smaller than the increase in sound intensity, which corresponds to a detection threshold of 75% at staircase convergence. Volume increases were approximately 7.5% SNR in the beginning, after the second reversal 3% SNR, and after the fourth and later reversals 0.75% SNR. A reversal was defined as a change from correct to incorrect responses or vice versa for one specific condition. After 12 reversals the staircase of that specific condition was terminated. Starting intensity was 19% SNR, which was for all participants above detection threshold. If for

three of four conditions the staircase was finished, additional trials of the other conditions were randomly added to remove predictability about condition type. Participants were encouraged to take a break after every 30 trials.

If not stated otherwise procedures were the same as in Experiment 1.

Data analysis

Experiment 1: First, we constructed psychometric functions for detection thresholds for each condition. To construct these functions, we calculated the mean detection rate at each intensity level (in bins 1% SNR wide), separately for each condition. A cumulative Gaussian was fitted to the individual data with the psychometric fitting toolbox modelfree v1.1 (Zchaluk & Foster, 2009). To eliminate effects of hysteresis (see e.g. Fender & Julesz, 1967; Palmer, 1999; Ratlifl et al., 1986) it is common to average the 50% detection values over the two paradigms (increasing and decreasing). These calculated averages per condition were used as the dependent variable in a two-way repeated measures ANOVA with factors Rhythmicity (Rhythmic versus Random) and Modality (Auditory versus AudioVisual).

Experiment 2: An exponential decay was fitted for the four conditions separately for all the SNR values presented during the whole experiment (Treutwein, 1995), using the lsqnonlin function implemented in MATLAB. The function was as follows:

$SNR(x)=19e^{\lambda x}+C$

where λ corresponds to the decay constant, C to the convergence value, and x to the trial number. The starting quantity was fixed at 19 (identical to the starting SNR in Experiment 2). To ensure that the lsqnonlin estimation did not result from a local minimum we repeated the procedure 30 times and took the final estimate as the fit with the most variance explained. The final convergence values¬¬ of the exponential decay were used as the dependent variables in a two-way repeated measures ANOVA with the factors Rhythmicity (Rhythmic versus Random) and Modality (Auditory versus AudioVisual).

Results

Experiment 1

For all participants, the mean percentage detection rate distribution had a shape typical of detection paradigms and could be reliably fit with a cumulative Gaussian function (figure 2; average explained variance 98.6%; see e.g. Florentine, Buus, & Geng, 1999; Green, 1995; Nachmias, 1981). The analysis using average 50% detection levels of the fitted

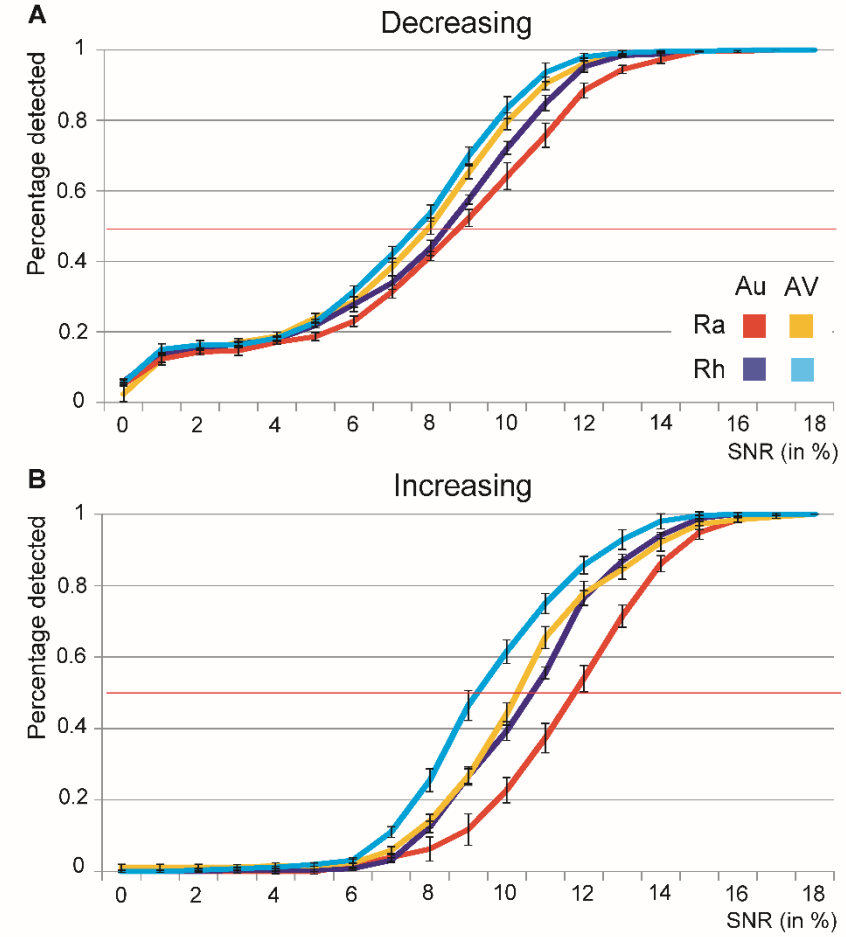


Figure. 2. Averaged detection percentages. Lines represent the average detection rate per intensity bin for the decreasing (A) and increasing (B) paradigm. All error bars represent the within-subjects standard error of the mean as described by (Morey, 2008).

psychometric functions (figure 3) revealed a main effect for Modality [Fig 3b; $F(1,11) = 40.41$, $p < 0.001$, partial $\eta 2=0.786$], that indicated that Audiovisual stimuli yielded lower thresholds than Auditory stimuli. Also a main effect for Rhythmicity was found [$F(1,11) = 62.62$, $p < 0.001$, partial $\eta 2=0.851$], that showed lower threshold for Rhythmic stimuli compared to Random stimuli. The interaction effect was not significant [$F(1,11) = 1.30$, $p = 0.279$].

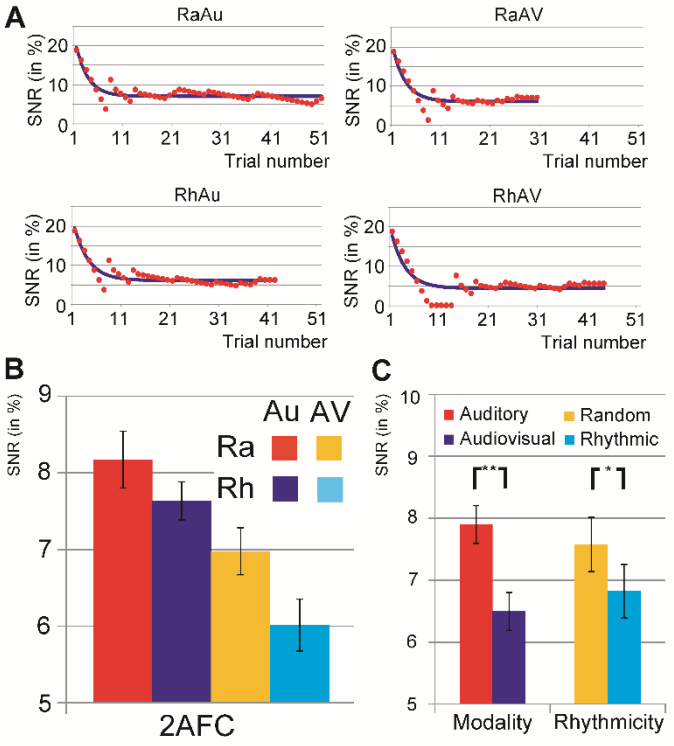


Figure 3. (A) Example of fitted decay functions for all conditions for a representative participant. Red dots indicate the actual SNR values and blue lines indicates the fitted decay. (B) Average SNR for the convergence values of the fitted exponential decay shown for all conditions separately. (C) The results for the two main effects of Rhythmicity and Modality. Error bars indicate the within-subjects standard error of the mean described by Morey (2008). Asterisks and double asterisks indicate significance at p-values of 0.05 and 0.01, respectively.

Experiment 2

The results of the 2AFC task are displayed in figure 4. The fitted exponential decay explained on average 60.2% of the variance and converged to the threshold (Fig 4a). A main effect for Modality was found [$F(1,19) = 38.68$, $p < 0.001$, partial $\eta 2=0.671$] in which Audiovisual stimuli yielded lower thresholds than Auditory stimuli. Additionally, a main effect for Rhythmicity was found [$F(1,19) = 6.22$, $p = 0.022$, partial $\eta 2=0.247$], in which Rhythmic stimuli yielded lower threshold than Random stimuli. The interaction effect was not significant [$F(1,19) = 0.45$, $p = 0.510$].

Discussion

The aim of the current study was to investigate how temporally-predictive visual cues and within-modality temporal regularities might change the detection of near-threshold auditory stimuli. As anticipated, we found that both types of predictive information improve auditory perception, such that sounds at lower intensity levels are judged as audible if they are preceded by visual input and/or are part of a rhythmic

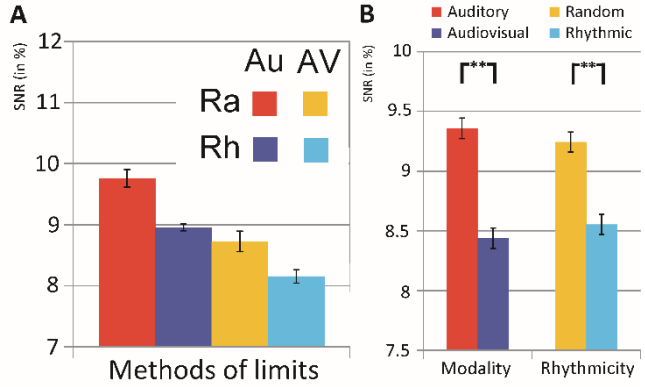


Figure 4. (A) Averaged SNR for 50% detection rate derived from the two psychometric curves are shown for all conditions separately. (B) Rhythmicity and modality main effects were significant in both the increasing and decreasing paradigms. Error bars indicate the within-subjects standard error of the mean described by Morey (2008). Double asterisks indicate significance at p-values of 0.01.

sequence. Critically, we further show that the combination of two types of predictive information lowered detection intensities even further (Trommershauser, et al., 2011), indicating that different sources of temporal information can be combined to optimize perception. We replicated the same pattern of results using two independent methods for quantifying perceptual thresholds, and show that these effects are not explained merely by participants' internal determinants such as response bias. Rather, our results suggest that temporal predictability can change perceptual sensitivity, and provides strong evidence for the 'attention in time' hypothesis (Barnes & Jones, 2000; Large & Jones, 1999; Schroeder, et al., 2008). This is consistent with electrophysiological results showing increased neuronal excitability at moments in time that stimuli are expected (Cravo, Rohenkohl, Wyart, & Nobre, 2011; Besle et al. 2011; Cravo, Rohenkohl, Wyart, & Nobre, 2013; Lakatos, et al., 2008; 2009; 2013).

Audiovisual Effects

Visual cues lowered the auditory detection intensity in both experiments. Since there was a constant temporal delay between visual and auditory stimuli, we suggest that the temporal predictability between the stimuli can be used to temporally prepare for the auditory stimulus, as has been shown in previous studies (Lange & Röder, 2006; Los & Van der Burg, 2013). Indeed, in cued reaction time tasks, the largest decrease in reaction times is typically found when there is a constant delay between the cue and the target, and this advantage is reduced as the delay becomes more variable (Niemi & Näätänen, 1981). Moreover, there is a crucial temporal window during which audiovisual stimuli are integrated (Lindström, Paavilainen, Kujala, & Tervaniemi, 2012; Van Atteveldt, Formisano, Blomert, & Goebel, 2007; Van Wassenhove, et al., 2007; Zampini, Shore, & Spence, 2003) and although the width of the window varies, the point of maximal integration is consistently when visual stimuli precede auditory stimuli (Thorne & Debener, 2008; Van Wassenhove, et al., 2007). Similarly, electrophysiological recordings show an enhancement of the neural response to auditory tones when they are preceded by a somatosensory or visual stimulus (Lakatos, Chen, O'Connell, Mills, &

Schroeder, 2007; Kayser & Logothetis, 2009; Lakatos et al., 2009; Thorne, De Vos, Viola, & Debener, 2011; Wallace, Wilkinson, & Stein, 1996), with the largest AV effect found at an audiovisual SOA of ~ 65 ms. This SOA has also been found to have the largest behavioral AV facilitation effect (Thorne & Debener, 2008) and therefore we choose this SOA in the current study. Although in other studies 300 ms has been found as optimal reaction time facilitation effect (Niemi & Näätänen, 1981), we did not choose this SOA since then the visual stimulus would be exactly in anti-phase of the auditory rhythm. Additionally, when audiovisual information is presented repeatedly with a fixed SOA temporal 'recalibration' occurs such that the audiovisual stimuli are more often perceived as synchronous at that SOA (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004). Although we did not specifically test for this, the fixed SOA used here probably induced this recalibration effect, which may have increased the integration of the audiovisual stimulus pair.

We cannot rule out the possibility that additional factors known to promote multisensory processing, such as spatial proximity (Plank, Rosengarth, Song, Ellermeier, & Greenlee, 2012; Wallace, et al., 1996) or content congruency (Beauchamp, Lee, Argall, & Martin, 2004; Van Atteveldt, Formisano, Goebel, & Blomert, 2004) also contribute to the observed effects. For example, visual input was also found to increase detectability for low intensity auditory stimuli presented simultaneously (Lovelace, Stein, & Wallace, 2003). Nonetheless, it seems plausible that the temporal relationship of the audiovisual stimulus pair used here, with the visual stimulus leading with a consistent, effective SOA contributed to reduction of the detection intensities by promoting temporal preparation.

The additive effects of Rhythmicity and Audiovisual cues

In the audiovisual conditions, Rhythmicity further reduced auditory detection compared to the Random condition. This result suggests that even though the visual input could perfectly predict the timing of auditory stimuli (since there was a constant lag of 65 ms), having temporal regularity within the sequence provides a significant additional

benefit for perception. One reason for this finding could be that, since people's temporal estimates are not entirely accurate (Eisler, 1976), particularly in the visual domain (Welch & Warren, 1980), having two sources of predictive information in different time scales (666 ms between two sequential visual cues and 65 ms from the visual cue to the sound) sharpens temporal predictions. This is consistent with previous studies showing that complementary cues about a stimulus can be expressed by integrating the reliability of the perceptual estimates of individual cues, with the combination yielding higher reliability than each single cue alone (e.g. Bulthoff, 1996; Landy, Maloney, Johnston, & Young, 1995; Oruç, Maloney, & Landy, 2003). Moreover, our data show an almost perfect additive effect of rhythm and visual cues which suggest that the two temporal prediction processes might convey two separate and possibly independent mechanisms (Sternberg, 2001, Woodman, Kang, Thompson, & Schall, 2008). One possibility for how these two types of information may work in concert to additively lower detection thresholds is a 'winner take all' approach, in which the occurrence of upcoming events is predicted by parallel mechanisms separately utilizing either the rhythmic structure or the known cross-modal SOA, and behavioral detection occurs when a stimulus is detected through one of those mechanisms. Alternatively, it could be that the type of benefit afforded by rhythmic-visual cues is synergistic in essence and is qualitatively different from that provided by non-rhythmic visual cues (Correa & Nobre, 2008; Nobre, et al., 2007; Schroeder & Lakatos, 2009). According to this view, visual cues occurring at random times invoke a 'vigilance mode' of operation (Schroeder & Lakatos, 2009), since the participant cannot anticipate when the cue will occur, and once it has occurred only has 65 ms to orient attention towards a potential auditory stimulus, requiring the swift allocation of computational resources. However, if the visual cues themselves are presented rhythmically, participants can enter a 'rhythmic mode' in which the timing of all stimuli – both visual and auditory – is completely predictable. It has been suggested that such a 'rhythmic mode' is a more automatic and implicit process, requiring less metabolic demand, whereas a 'vigilance mode' requires explicit and controlled processing (Capizzi, Sanabria, & Correa, 2012; Correa, 2010; Schroeder, et al., 2010; Van Atteveldt et al., 2011). Additional research is required to further characterize how these two modes of attention in time

interact and work together to affect perceptual processing, particularly since the interpretation of additive effects on perceptual thresholds as reflecting contribution of sequential or parallel processes is not straightforward (Dubois, Poeppel, & Pelli, 2013; Sternberg 2001; Miller, van der Ham, Sanders, 1995).

The effect of Rhythmicity in the absence of Visual cues

Although not as effectively as the visual input, adding rhythmicity to the auditory stream also decreased detection thresholds. What is striking is that this effect occurred even when no visual stimulus was presented and auditory stimuli were not yet detected. This exciting finding implies that the rhythmic pattern of tones is detected at lower intensities than each individual tone. It suggests a central role for temporal integration and detection of temporal regularities in near-threshold perception. This pattern is in line with previous findings that auditory detection thresholds are reduced for low intensity auditory stimuli when they are presented for a longer duration, supposedly brought about by the aggregation of subthreshold information (Florentine, Fastl, & Buus, 1988; Lütkenhöner, 2011; Yrttiaho, Tiitinen, Alku, Miettinen, & May, 2010) (similar effects are found in the visual system, see: Anstis, Geier, & Hudak, 2012; Daikhin & Ahissar, 2011; Minelli, Marzi, & Girelli, 2007). Since it has been shown that 3-4 stimuli are sufficient for a neuronal population to identify a rhythmic structure (Lakatos, et al., 2008; Thorne, et al., 2011), it seems plausible that the (even subliminal) processing of a few rhythmic stimuli provides enough information about the temporal structure of the auditory stream to decrease perceptual thresholds. One intriguing question is how precisely isochronous a stimulus trains needs to be to still provide a perceptual benefit. Neuronal entrainment has been previously demonstrated for tone sequences with a temporal jitter of up to 20% (SOAs distributed between 666 ±150 ms ;Lakatos, et al. 2008). This suggests that the system can tolerate some degree of jitter and still maintain temporal predictions that can facilitate perception, which would be beneficial from an ecological perspective since many real-life stimuli such as speech and music have temporal regularities but are not perfectly isochronous. The robustness of temporal prediction to jitter, and its

influence on both perception and on neural processing needs to be systematically studied in future work.

Conclusion

We show that both temporal regularity within a stimulus stream and cross-modal temporal cueing decrease auditory detection thresholds. Moreover, both types of temporal information are used in combination to prepare our system for incoming stimuli and may play complementary roles in focusing 'attention in time'. These findings are a testament to the flexibility and proactivity of the perceptual system (Schroeder, et al., 2010; Zion Golumbic, et al., 2012; Van Atteveldt, Murrary, Thut, & Schroeder, 2014), in that thresholds for reporting auditory detection are not necessarily fixed but rather are strongly influenced by contextual factors, like those tested here. Our findings have implications for understanding the role of temporal prediction in processing more complex and natural stimuli, such as speech, which contain both intrinsic regularities (Giraud & Poeppel, 2012; Greenberg, et al., 2003; Luo & Poeppel, 2007) as well as temporally predictive cross-modal cues such a facial and head movements (Schroeder, et al., 2008; Chandrasekaran, et al., 2009; Grant & Seitz, 2000; Munhall & Vatikiotis-Bateson, 2004; Schwartz, Berthommier, & Savariaux, 2004; Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013), both of which are likely to influence the fundamental operations of auditory cortex (Lakatos et al., 2013; Schroeder, et al., 2008). Although additional research is needed to understand the usage of multiple contextual factors during perception, we show that contextual information can be combined from different sources to allocate our attention in time, thereby sensitizing and optimizing perception.

Acknowledgements: This study was supported by grants from the Dutch Organization for Scientific Research [NWO; grant numbers 406-11-068 (to StO) and 451-07-020 (to NvA)], the National Institutes of Health [NIH; grant numbers R01-DC0566 (to DP), R01-DC011490 (to CES), MH103814 (to CES), and F32-MH093061 (to EZG)], and European

Community's Seventh Framework Programme [FP/2007-2013; grant number 22187 (to NvA)].

References

Anstis, S., Geier, J., & Hudak, M. (2012). Afterimages from unseen stimuli. *i-Perception, 3*(8), 499-502.

Bar, M. (2011). *Predictions in the brain: Using our past to generate a future:* OUP USA.

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive psychology, 41*(3), 254-311.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience, 31*(9), 3176-3185.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809-823.

Bulthoff, H. H. (1996). *Bayesian decision theory and psychophysics.*

Capizzi, M., Sanabria, D., & Correa, Á. (2012). Dissociating controlled from automatic processing in temporal preparation. *Cognition, 123*(2), 293-302.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436.

Correa, A. (2010). Enhancing behavioural performance by visual temporal orienting. *Attention and time*, 357-370.

Correa, Á., Lupiáñez, J., Milliken, B., & Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Attention, Perception, & Psychophysics, 66*(2), 264-278.

Correa, Á., & Nobre, A. C. (2008). Neural modulation by regularity and passage of time. *Journal of Neurophysiology, 100*(3), 1649-1655.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience, 18*(18), 7426-7435.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *Journal of Neurophysiology, 106*(6), 2964-2972.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

Daikhin, L., & Ahissar, M. (2011). Responses to deviants are modulated by subthreshold variability of the standard. *Psychophysiology, 49*(1), 31-42.

Dubois, M., Poeppel, D., & Pelli, D. G. (2013). Seeing and Hearing a Word: Combining Eye and Ear Is More Efficient than Combining the Parts of a Word. *PloS one, 8*(5), e64803.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin, 83*(6), 1154.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics, 72*(8), 2274-2288.

Fender, D., & Julesz, B. (1967). Extension of Panum's fusional area in binocularly stabilized vision. *JOSA, 57*(6), 819-826.

Florentine, M., Buus, S., & Geng, W. (1999). Psychometric functions for gap detection in a yes–no procedure. *The Journal of the Acoustical Society of America, 106*, 3512.

Florentine, M., Fastl, H., & Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America, 84*, 195.

Friston, K. (2011). Prediction, perception and agency. *International Journal of Psychophysiology, 83*(2), 248-252.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*(7), 773-778.

Gescheider, G. (1997). Chapter 3: The Classical Psychophysical Methods. In L. E. Associates (Ed.), *Psychophysics: the fundamentals* (3rd ed.): Psychology Press

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*, 1197.

Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *The Journal of the Acoustical Society of America, 97*, 3749.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1974): Wiley New York.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics, 31*(3), 465-485.

Griffin, I. C., Miniussi, C., & Nobre, A. C. (2001). Orienting attention in time. *Frontiers in Bioscience, 6,* 660-671.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive psychology, 53*(1), 59-96.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313-319.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception, & Psychophysics, 49*(3), 227-229.

Kayser, C., & Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience, 3*.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*(4), 750-761.

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64*(3), 419-430.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision research, 35*(3), 389-412.

Lange, K., & Röder, B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. *Journal of Cognitive Neuroscience, 18*(5), 715-729.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119.

Lindström, R., Paavilainen, P., Kujala, T., & Tervaniemi, M. (2012). Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. *Frontiers in Psychology, 3*.

Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 1612-1624.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*(2), 447-453.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS biology, 8*(8), e1000445.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001-1010.

Lütkenhöner, B. (2011). Auditory signal detection appears to depend on temporal integration of subthreshold activity in auditory cortex. *Brain research, 1385*, 206-216.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition, 115*(1), 186-191.

Miller, J., van der Ham, F., & Sanders, A. F. (1995). Overlapping stage models and the additive factor method. *Acta Psychologica, 90*, 11-28.

Minelli, A., Marzi, C. A., & Girelli, M. (2007). Lateralized readiness potential elicited by undetected visual stimuli. *Experimental Brain Research, 179*(4), 683-690.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 9*(45.4), 49.46.

Munhall, K., & Vatikiotis-Bateson, E. (2004). Spatial and Temporal Constraints on Audiovisual Speech Perception. In S. C. Calvert G.A., Stein B. (Ed.), *The Handbook of Multisensory Processing* (pp. 177-188). Cambridge, Massachusetts: The MIT Press.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision research, 21*(2), 215-223.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*(1), 133.

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology, 17*(4), 465-470.

Nobre, K., & Coull, J. T. (2010). *Attention and time:* Oxford University Press.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision research, 43*(23), 2451-2468.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*: The MIT press.

Plank, T., Rosengarth, K., Song, W., Ellermeier, W., & Greenlee, M. W. (2012). Neural correlates of audio-visual object recognition: Effects of implicit spatial congruency. *Human Brain Mapping, 33*(4), 797-811.

Ratlifl, F., Comsweet, T., Ross, J., Holt, J., Johnstone, J., Ratliff, F., et al. (1986). Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature, 324*, 20.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience, 32*(24), 8424-8428.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20*(2), 172-176.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences, 11*(5), 211-218.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition, 93*(2), B69-B78.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106*, 147-246.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science, 314*(5803), 1311-1314.

Ten Oever, S., Sack, A., Wheat, K. L., Bien, N., & Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology, 4*.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience, 31*(10), 3853-3861.

Thorne, J. D., & Debener, S. (2008). Irrelevant visual stimuli improve auditory task performance. *Neuroreport, 19*(5), 553.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision research, 35*(17), 2503-2522.

Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*: Oxford University Press.

Van Atteveldt, N., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex, 17*(4), 962-974.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*(2), 271-282.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6),1240-1253

Van Atteveldt, N., Musacchia, G., Sehatpour, P., Zion Golumbic, E. M., Lakatos, P., Gaspar, P., et al. (2011). A combined EEG-fMRI investigation of "rhythmic" versus "vigilant" processing II: functional neuro-anatomy and slow dynamics, in:. *International Conference on Cognitive Neuroscience (ICON XI), Mallorca, Spain*.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32-35.

Wallace, M., Wilkinson, L., & Stein, B. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246-1266.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638.

Woodman, G. F., Kang, M.-S., Thompson, K., & Schall, J. D. (2008). The effect of visual search efficiency on response preparation. *Psychological Science, 19*, 128-136.

Yrttiaho, S., Tiitinen, H., Alku, P., Miettinen, I., & May, P. J. (2010). Temporal integration of vowel periodicity in the auditory cortex. *The Journal of the Acoustical Society of America, 128*(1), 224-234.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198-210.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*(6), 1414-1425.

Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual Input Enhances Selective Speech Envelope Tracking in Auditory Cortex at a "Cocktail Party". *The Journal of Neuroscience, 33*(4), 1417-1426.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*(3), 151-161

CHAPTER 3

EVIDENCE FOR ENTRAINMENT TO SUB-THRESHOLD RHYTHMIC AUDITORY STIMULI

Corresponding Manuscript:
Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion Golumbic, E. M. (work in progress). Evidence for entrainment to sub-threshold rhythmic auditory stimuli.

Abstract

Many environmental stimuli contain temporal regularities, a feature which allows predicting the occurrence of forthcoming input. It has been proposed that ambient low-frequency neuronal oscillations phase-lock (entrain) to rhythmic stimuli, thus aligning high excitability phases with events within the stream, effectively enhancing neuronal responses to them (Besle et al., 2011; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Schroeder & Lakatos, 2009) and subsequently enhancing feature sensitivity (Arnal, Doelling, & Poeppel, 2014; Cravo, Rohenkohl, Wyart, & Nobre, 2013). As most studies use highly-salient rhythms, it is unclear whether the effect arises purely from the rhythmic evoked responses or alternatively could also arise through subtle subthreshold modulations in absence of clear evoked responses. To evaluate these possibilities we investigated the changes in neural dynamics as sub-threshold rhythmic stimuli gradually become audible. Here, using magnetoencephalographic (MEG) recordings, we report significant delta phase-locking to rhythmic sounds prior to report of their detection. Importantly, this subthreshold entrainment could be dissociated from auditory evoked responses, as these only appeared when sounds were loud enough to be detectable. The current findings support the proposition that entrainment of low-frequency oscillations serves a mechanistic role in enhancing perceptual sensitivity to rhythmic stimuli. This framework also has broad implications for understanding the neural mechanisms involved in generating temporal predictions and their relevance for perception, attention and consciousness (Large & Jones, 1999; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Zion Golumbic, Poeppel, & Schroeder, 2012).

Results

Entrainment of neural electrical activity to temporal regularities in the environment enhances neuronal responses by aligning depolarized phases of ongoing oscillations to the arrival time of expected events (Besle, et al., 2011; Lakatos, et al., 2008; Schroeder & Lakatos, 2009). Since these phases correspond to periods of high neuronal excitability, weaker inputs can elicit action potentials and thus facilitate downstream information processing (Buzsáki & Draguhn, 2004). It has been shown repeatedly that perceptual sensitivity varies systematically as a function of low-frequency phase (Arnal, et al., 2014; Cravo, et al., 2013; Fiebelkorn et al., 2013), and that rhythmic temporal regularities within a stimulus carry behavioral benefits (Arnal, et al., 2014; Cravo, et al., 2013). In most studies entraining stimuli are presented at supra-threshold levels. This has the drawback that the resulting rhythmic patterns of electrical activity are dominated by large evoked responses, which obscure the smaller underlying fluctuations in the membrane potential. The latter would show the true added value of rhythmic stimulation for stimulus detection: already during subthreshold stimulation oscillatory entrainment patterns develop and these align the optimal excitability state of the neuronal ensemble to the events in the stimulus stream, increasing the chance of detection. In a previous study we showed that rhythm can indeed have a beneficial effect on detection thresholds even when the rhythm itself is below perceptual detection threshold (Ten Oever, Schroeder, Poeppel, Van Atteveldt, & Zion Golumbic, 2014). This indicates that temporal information is extracted and exploited even before a stream is consciously perceptible. The current study aimed to investigate the neural mechanism that renders these low intensity sounds audible.

We presented participants with auditory tones (1 kHz beeps of 50 ms) presented in a rhythmic (inter-stimulus interval 667 ms, 1.5 Hz) or random (inter-stimulus interval varying between 300-1000 ms, on average 1.5 Hz) sequence. The tones were embedded in noise and initially below threshold. Over the course of the trial the intensity increased gradually in different steps to ensure that the time point of detection was not predictable (in steps of 0.25, 0.5 or 0.75% of signal to noise (SNR) increases) and participants indicated when they started to hear the sounds

(figure 1A). Trials varied in length, but were on average relatively long; often being longer than 30 seconds.

Analysis focused on the three channels with the largest auditory response in each hemisphere from each participant, as obtained from an independent auditory localizer (figure 1B; see experimental procedures for details). For the main experiment, we first aligned and epoched the data around the time of sound onset between -1 to 1 sec. We classified all epochs as either pre- or post-threshold. Pre-threshold epochs were classified as follows: for each participant we first determined the "minimal-detection intensity", which was the lowest intensity over both conditions in which the participants detected the sound. Only epochs in which all the sounds were of intensities below this minimal-detection intensity were classified as "pre-threshold." This served as a conservative

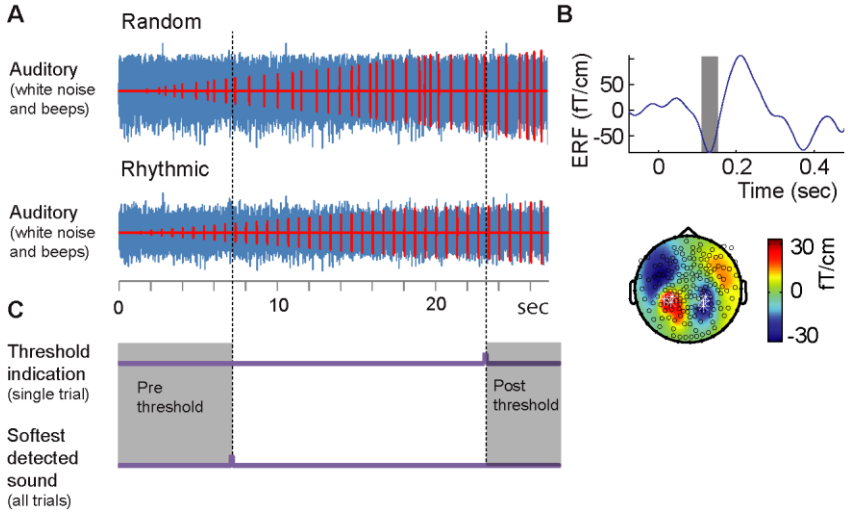


Figure 1. Illustration of the trials and auditory localizer. A) Beeps (red) were embedded in white noise (blue), with their intensity increasing monotonically over the trial. The button press (purple) indicates the moment that the participant indicates hearing the sound for the first time. B) Example for the determination of peak amplitude, latency, and electrodes of the M100n for one participant in an independent auditory localizer. The ERF of the average of the six highlighted channels is displayed (multiplying the left channels with -1; this will be the convention in the following figures). Topography reflects the average of the grey shaded area in the time course. C) Pre and post threshold epoch selection. All sounds below the softest detected sound over all trials are classified as pre threshold. All sounds after threshold indication of that specific trial are classified as post threshold.

threshold estimation, as "pre-threshold" epochs only included sounds at intensities that were never detected. Post-threshold epochs included all the epochs in which the sound was at or above the detection threshold in that specific trial (figure 1C).

Figure 2A shows the event related fields (ERF) for pre- and post-threshold intervals filtered at 1-2 Hz and at 1-20 Hz. Overall it seems that a 1.5 Hz modulation is present in the rhythmic condition, even in the pre-threshold window. The inter-trial coherence (ITC) also shows clear peaks at 1.5 Hz and its harmonics for the rhythmic, but not for the random condition, in both the pre and post detection threshold intervals (figure 2B). A repeated measures ANOVA for ITC at 1.5 Hz with the factors Condition (random and rhythmic), Detection Threshold (pre and post), and Hemisphere (left and right) indicated that the rhythmic condition had a significantly stronger ITC than the random condition

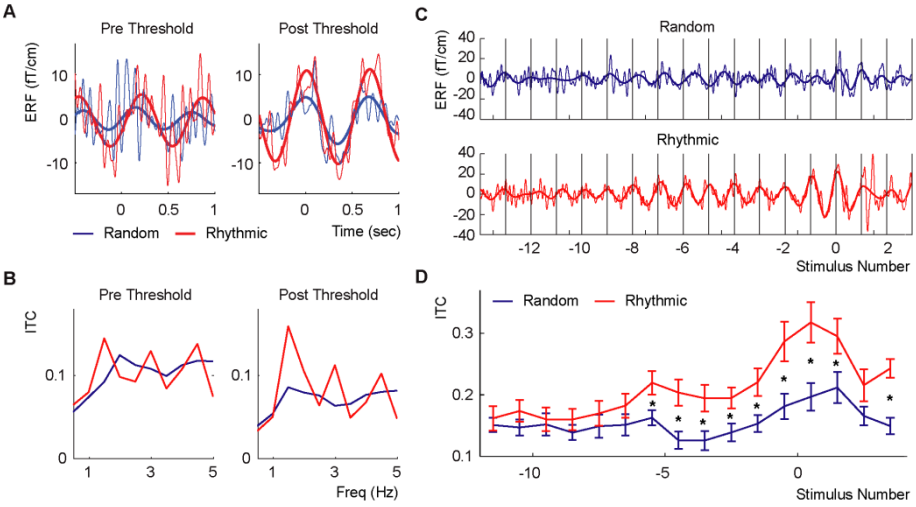


Figure 2. Inter trial coherence (ITC) analysis. A) Average ITC over participants of all the pre and post detection threshold epochs for both the random (blue) and rhythmic (red) condition. Clear peaks at 1.5 Hz and its harmonics are visible in the rhythmic condition. B) Event related fields (ERF) for the pre and post detection epochs for both condition for data filtered between 1-20 Hz and between 1-2 Hz. C) ERF traces of the full trial, locked to the last stimulus prior to detection (t=0) for data filtered between 1-20 Hz and 1-2 Hz. Vertical black lines indicate the times of sound presentation in the rhythmic condition. D) ITC over the course of the trial. Error bars indicate the standard error of the mean. Asterisks indicate significant differences between the random and rhythmic condition.

[$F(1,14) = 28.90$, $p < 0.001$]. No other effects were significant. In figure 2B the stronger ITC for the rhythmic condition is evident, even in the pre-threshold interval.

To better understand the development of entrainment over the time-course of the trials we sorted the sounds relative to their detection (stimulus number 0 indicating the sound that was first detected). From figure 2C it is evident that the phase of the averaged MEG signal at sound onset is quite stable already early in the time course. To test the consistency of these phases within each trial we re-labeled the epochs according to their position relative to detected sound in that specific trial. To get a reliable estimate of the ITC we used a moving window approach in which the epochs of 2 adjacent sound stimuli were included for the estimation of the ITC (figure 2D). This was necessary as we were limited in trial numbers (due to the length of each trial often being longer than 30 seconds). We again performed a repeated measures ANOVA for the ITC at 1.5 Hz with the factors Condition (random and rhythmic), Hemisphere (left and right), and Stimulus number (16 levels ranging from -11.5 to 3.5. Half numbers represent that the moving window includes two epochs of adjacent sounds). We found an interaction between condition and stimulus number [$F(15,224) = 2.248$, $p = 0.022$]. Additionally, we found a main effect for stimulus number [$F(15,224) = 7.195$, $p < 0.001$] and condition [$F(1,14) = 21.909$, $p < 0.001$]. We performed a simple effects analysis comparing the random and rhythmic condition at each stimulus number (tests were corrected for multiple comparisons using false discovery rate). From stimulus number -5.5 on the ITC of the rhythmic condition was higher than the random condition, and this effect lasted for the rest of the trial (excluding stimulus number 2.5). This analysis indicates that inter-trial coherence increases well before stimulus detection, minimally 4 stimuli before sound detection (considering the length of the epochs and the moving window approach).

In a next step, we wanted to verify more directly that the effect of interest indeed reflects sub-threshold entrainment of neuronal oscillations, and not a series of evoked responses to sub-threshold stimuli. To this end we measured two different measures that could be indicative of evoked responses: 1) the overall power and 2) the N100m response. We calculated the power using the same trial selection as for the ITC

analyses, but extracting the square of the absolute value of the complex Fourier spectra for each trial and averaging the power. For statistical analysis we normalized the power by calculating the relative change subtracting the mean of all conditions of the individual power values and dividing it by this mean. In figure 3A the non-normalized overall power is shown for both the pre- and post-threshold interval. In both spectra a clear 1/f distribution is visible, typical of any EEG/MEG response (Miller, Sorensen, Ojemann, & Den Nijs, 2009; Pritchard, 1992). The ANOVA for difference in 1.5 Hz power [with the factors Condition (random and rhythmic), Detection Threshold (pre and post), and Hemisphere (left and right)], showed a significantly higher 1.5 Hz power for the post-threshold trials compared to the pre-threshold trials [see inset at figure 3A; $F(1,14) = 6.029$, $p = 0.028$]. Moreover, a three-way detection*hemisphere*condition interaction was visible [$F(1,14) = 25.950$, $p < 0.001$] as well as a hemisphere*condition interaction [$F(1,14) = 5.192$, $p = 0.039$]. The three-way interaction seemed to be driven by a stronger power increase in the left compared to right hemisphere only in the random condition. This effect was absent for the rhythmic condition [detection effect in the random left hemisphere: $F(1,14) = 3.912$, $p = 0.003$; hemisphere*detection interaction in the random condition: $F(1,14) = 7.286$, $p = 0.017$; hemisphere*detection interaction in the rhythmic condition: $F(1,14) = 0.237$, $p = 0.634$]. Direct contrasts between the two

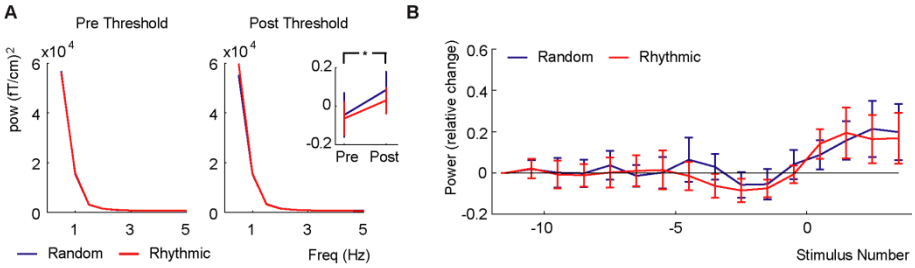


Figure 3. Power analysis. A) Average power over participants of all the pre and post detection threshold epochs for both the random (blue) and rhythmic (red) condition. The MEG shows a typical 1/f distribution, but no peaks at 1.5 Hz are present. The inset on the right indicates the relative change in 1.5 Hz power (relative to the average of all conditions) which is significant (as indicated by the asterisk) B) 1.5 Hz power over the course of the trial. The relative change in power is presented for both conditions (compared to stimulus number -11.5). Error bars indicate the standard error of the mean.

conditions were not significant [main effect condition F(1,14) = 0.327, p = 0.576]. The development of the 1.5 Hz power also showed that prior to sound detection no clear power increases are visible, but they seemed to develop after detection (figure 3B; relative change to stimulus number -11.5). The ANOVA for the development of 1.5 Hz power showed a significant effect of stimulus number [F(14,196) = 3.035, p = 0.033]. To verify the seemingly present increase in power after sound detection we compared the power values at each stimulus number with zero. However, none of these contrasts survived multiple comparisons.

In a second step to estimate the evoked responses we measured the amplitude within the time-window associated with the N100m responses to each auditory stimulus in the rhythmic condition from -11.5 stimuli prior to indication of detection until +3.5 after detection performing the same moving window approach as before. This N100m

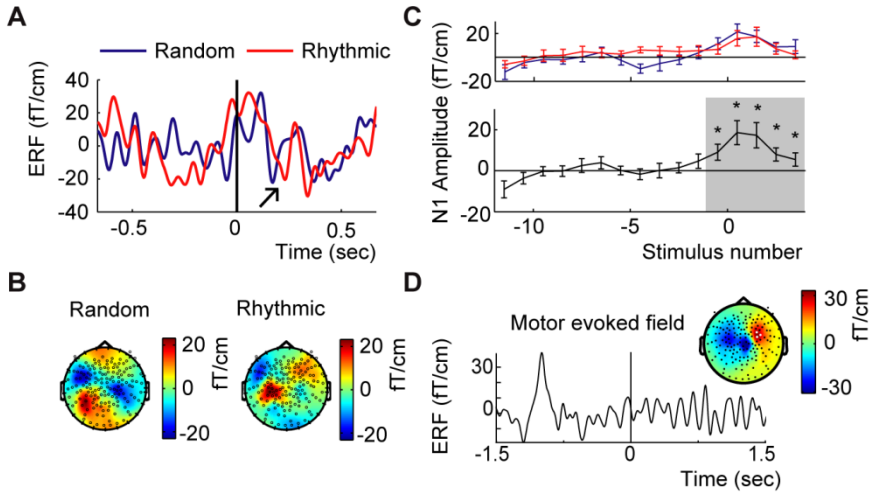


Figure 4. ERF and N100m effects. A) ERF average over the different hemispheres only using the epochs of the first detected sound (filtered between 0.5 and 20 Hz). The arrow indicates the location of the N100m. B) The topographies of the N100m using the individual peak timing per participant. C) Development of the N100m over the course of the trial for the random (blue), rhythmic (red), and their average (black). Error bars indicate the standard error of the mean. Asteriks indicate where the N100m is significantly stronger as the -11.5 stimulus number. D) Motor evoked field aligned to the response. Topography indicates the peak response. White marked channel in the topography of the motor evoked field are the plotted channels in the time course.

latency was individually determined via an independent localizer. Here, we extracted the latency of the peak amplitude in a window between 80 and 150 ms which had a matching N100m topography (figure 1B). This latency was used in the following analysis to extract the N100m amplitude calculating the average of all data points around a 50 ms wide interval. Initially, the analysis did not show any results; however it is known that stimuli with a low intensity have a significantly later N100m latency compared to high intensity stimuli (see e.g. (Elberling, Bak, Kofoed, Lebech, & Saermark, 1981; Lütkenhöner & Klein, 2007)), which were used in the localizer. Therefore, we shifted the individual latencies with fifty milliseconds (conform to the latency shift in the literature) and extracted the amplitude at this new latency (again using a 50 ms wide interval). A shifted N100m (around 180 ms) with low amplitude is also evident from figure 4A, which displays the grand average ERF of the first detected sound (figure 4B shows the topography). We entered the shifted N100m amplitudes into a Repeated Measures ANOVA with factors Hemisphere, Condition, and Stimulus number. This analysis revealed a significant effect of stimulus number [$F(15, 210) = 5.339$, $p < 0.001$] and an interaction between hemisphere and stimulus number [$F(15,210) = 2.059$, $p = 0.044$]. Evaluating this interaction effect further, we split up the data for each separate stimulus number and performed a t-test comparing the two hemispheres. This did not reveal any significant effects. To test the main effect of stimulus number, we performed t-tests comparing for each stimulus number the N100m amplitude with N100m amplitude of the first stimulus (-11.5). This analysis shows whether over the course of the trial an evoked response to the sounds would appear (assuming that no evoked response is present for stimuli 11 stimuli prior to detection). A significant effect for the stimulus numbers ranging from -0.5 to 3.5 was present, indicating that an evoked response only occurred after participants detected the sounds (figure 4C. It is important to note that the amplitudes of stimulus number 0.5 and 1.5 were likely influenced by visual evoked responses, as the fixation cross changed from grey to green after participants pressed the button). These effects could not be attributed to any motor evoked responses as the motor response was present at the more anterior right hemisphere (figure 4D) and all the channels used for the auditory analyses came from the posterior channels of the N100m response.

Discussion

We investigated the effects of stimulus rhythmicity on observers' detection thresholds by measuring neural responses to streams of sub-threshold auditory stimuli that gradually became audible, presented either with a rhythmic or random timing. We show significant delta entrainment for sub-threshold rhythmic sounds. This entrainment was dissociated from sensory evoked responses, which appeared only near the indication of conscious detection. These novel results show that despite being below conscious detection level, the rhythmic structure of sub-threshold stimuli is nonetheless sufficient for entraining delta oscillations. This may underlie our previously reported findings of perceptual benefits and lower detection-thresholds for rhythmic stimuli (Ten Oever, et al., 2014).

Sub-threshold effects

Our results are consistent with the "neural entrainment hypothesis" (Schroeder & Lakatos, 2009; Schroeder, et al., 2008) which postulates that phase alignment of delta oscillations to the rhythm of external sounds ensures that upcoming stimuli will fall on a high excitable phases of the oscillation, thereby increasing processing efficiency and reducing detection thresholds. However, entrainment to rhythms below perceptual threshold has to our knowledge not been reported before. This novel finding suggests that environmental rhythmic information is utilized by the brain even before we are aware of any stimulation.

The environment is full of different rhythmic structures that are important for human behavior, such as music, biological motion, and speech. Considering that the brain during rest seems to be composed of complex oscillatory patterns of rhythmically varying membrane potentials rather than purely random fluctuations (Berger, 1929; Buzsáki & Draguhn, 2004) its machinery seems especially sensitive to rhythmic inputs. Subthreshold inputs have been shown to align the phases of slow fluctuations of groups of neurons, even when the number of spikes does not increase (Buzsáki, 2004; Buzsáki & Draguhn, 2004; Pike et al., 2000). Here, we show that rhythmic input from the environment seems to act

upon this machinery thereby influencing the phase of these ongoing oscillations even when clear evoked responses are absent.

Absence of evoked responses

We measured evoked responses via a power analysis and the N100m response. Both measures showed increased responses when participants detected the sounds. These two separate measures complement each other and indicate that evoked responses do not seem to be driving any pre detection effects. One could argue that the absence of this effect is due to a lack of power to reach statistical significance in the measured evoked responses. We believe this unlikely as both measures did show reliable responses after sound detection. Moreover, both effects measure different aspects of the evoked response and show the same pattern. Additionally, the inter-trial coherence measure would also be negatively influence by changes in noise level as phase estimates are less reliable under these circumstances. Therefore, we conclude that our results reflect the alignment of ongoing oscillatory patterns to the presented sub-threshold rhythmic stream in absence of clear changes in evoked patterns.

Conclusion

Many natural stimuli contain temporal regularities, and the brain is tuned to process the natural statistics of the environment (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Simoncelli & Olshausen, 2001). Our study has implications for understanding the neural mechanisms involved in utilizing these temporal regularities. We show that even when rhythmically presented sounds are not consciously detected, they can proactively enhance subsequent processing. Specifically, we provide evidence supporting the 'entrainment hypothesis', which posits that alignment of neuronal excitable phases to the temporal structure of incoming stimuli enhances sensory processing. Our study indicates that this alignment occurs without any increase in evoked responses, strongly suggesting that it is the phase alignment, and not sensory evoked responses that drive the entrainment. This mechanism has broad

implications for understanding the neural basis of perception, attention, and consciousness (Large & Jones, 1999; Schroeder, et al., 2008; Zion Golumbic, et al., 2012), as it emphasizes the predisposition of the system to identify and utilize temporal statistics in the environment to form predictions and ultimately facilitate neural processing.

Experimental procedures

Participants

Sixteen participants completed the experiment (range: 23-37, mean age 27, 7 male). All had normal or correct to normal vision and gave written informed consent. Participants received monetary compensation. The study was approved by the New York University Committee on Activities Involving Human Subjects (NYU UCA/HS). One participant was removed of the data analyses since he did not follow the instruction of the behavioral task.

Stimulus material

Auditory stimuli were sinusoidal 1 kHz beeps, lasting 50 ms (with a linear rise and fall time of 5 ms) embedded in continuous white noise (53 dB). The software Presentation used for stimulus delivery; Neurobehavioral Systems, Inc., Albany, NY).

Procedure

Auditory localizer: Before the main experiment, participants in the MEG experiment underwent an auditory localizer procedure. This localizer consistent of a total of 200 auditory beeps (400 ms long) of which half had a high frequency (1000 Hz) and the other half a low frequency (250 Hz). Inter-stimulus interval (ISI) was varied between 1.2, 1.3, or 1.4 seconds. ISI and stimulus frequency was randomized. The task lasted approximately 5 minutes and participants had to fixate the screen.

Main experiment: Participants heard a stream of auditory beeps embedded in continuous white noise. The signal to noise ratio (SNR) of the beeps was initially below threshold, and the intensity of the beeps increased monotonically over the trial. Participants were asked to

indicate via button press when the auditory beeps were first detected. The starting SNR was 7%. Over the trial, SNR increased incrementally in steps of 0.25, 0.5 or 0.75%. The different incremental steps were randomized to ensure that the sequence of sounds and length of the trials were not identical across trial. Stimuli were presented until an SNR of 17.5%, independent of the participant's response. After the participant indicated to hear a sound the fixation cross changed color from grey to green and stayed like this for five consecutive sound before turning back to grey.

In the rhythmic condition there was a constant inter-stimulus interval (ISI) of 667 ms between the beeps, whereas for the random condition the ISI was randomized amid one of 21 evenly spaced time points between 300 and 1000 ms, maintaining an average ISI of 667 ms,. The current results were part of an experiment that also included an audiovisual condition where a Gaussian white circle preceded every auditory stimulus, however, the audiovisual trials were not included in the data analyses since in the current paper we were mainly interested in how rhythmic stimuli become audible from a noisy background.
Participants were explicitly instructed to maintain fixation on a grey cross in the middle of the screen. Trials were randomized across conditions (20 trials per condition) and the experiment was divided in four blocks of approximately eleven minutes each. After every block participants were encouraged to take a break. A trial was defined as the whole period from the onset of the white noise until the last sound was presented.

MEG recordings and data pre-processing

A 160-channel axial gradiometer (157 data, and 3 reference channels) whole-head MEG system (KIT, Kanazawa, Japan) was used for data acquisition. Head position was monitored via five electromagnetic coils attached to the participant's head located in respect to the nasion and both preauricular points using 3D digitizer software (Source Signal Imaging, Inc.) and digitizing hardware (Polhemus). Sampling rate was 1000 Hz with online filtering of DC-200Hz. Initial noise reduction was using the CALM algorithm (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001) implemented in the MEG160 software (KIT, Kanazawa, Japan). All other analyses were performed using the Fieldtrip toolbox

(Oostenveld, Fries, Maris, & Schoffelen, 2011) implemented in MATLAB (MathWorks). First, data was resampled to 256 Hz and bad channels were replaced by the average of the neighboring channels. Then, independent component analyses (ICA; using the logistic infomax ICA algorithm (Bell & Sejnowski, 1995), extracting 75 principle components) was performed to remove artifacts related eye blinks, eye movements, and heartbeat. Extreme trials were removed via visual inspection.

MEG Analyses

First we extracted from the independent auditory localizer which channels had the strongest N100m response on the left and the right hemisphere and their latency (collapsed over both frequencies). We took the three most positive channels on the left side and the three most positive channels on the right side for each individual; conform to the auditory topography of MEG (see figure 1).

ITC estimation: First we epoched the data around sound onset (-1 to 1 sec) and baseline corrected to the 200 ms prior to trial onset (similar for all upcoming analyses). Then we sorted the epochs either as pre- or post-threshold. Pre-threshold epochs included all the epochs that had a lower threshold than the minimum threshold value per condition. Also the epochs centered around the sounds directly before the value reached this minimum value were excluded. As the epoch window is longer as the inter-stimulus interval window this ensured that we only included intervals in which the stimulus intensity was never detected. Then we extracted the phase angles by performing a frequency analysis using hanning tapers over all individual epochs and calculated the ITC. Effects of 1.5 Hz entrainment were statistically tested using a repeated measurements ANOVA with the factors Condition (random or rhythmic), Hemisphere (left or right), and Detection Threshold (pre or post threshold). All filtering present in the figures was done prior to epoching to eliminate edge effects.

To evaluate the development of the ITC over the trial we re-labeled the epochs now reflecting their position relative in the trial in which zero indicates the first sound detected (-1 indicates the one sound prior to detection etc.). As we had a limited trial amount and therefore limited epochs (due to the length of the trials), we performed a moving window

approach to reliably estimate the ITC: two sequential epochs were used to estimate the ITC for all stimulus numbers ranging from -12 to 4. We again calculated the ITC and statistically tested the effect using a repeated measures ANOVA with factors Condition, Hemisphere, and Stimulus number (16 levels ranging from -11.5 to +3.5; the half number reflect the moving window approach). We used the Huyn-Feld method to correct for violations for spericity. Simple effects analyses were performed for significant interactions using false discovery rate to correct for multiple comparisons.

Power estimation: We repeated the analyses for the ITC estimation, except using the square of the absolute value of the complex Fourier spectra for the different ANOVA's. The power spectra for the pre- versus post-threshold analysis were normalized by subtracting the average over all conditions from the individual values and dividing by this average. Power spectra for the second analysis evaluating the development of the trial were normalized by subtracting the power of the -11.5 stimulus. Trials with extreme power values were removed.

Evoked response and N100m. The same epochs as the previous analyses were used evaluate the development of the N100m response. To exclude any effects caused by the delta entrainment, we filtered the data between 3 and 20 Hz. Then, we extracted the individual N100m responses by using the latency of the individually determined N100m latency. Fifty milliseconds was added to this latency as it is know that stimuli with a low intensity have a significantly later N100m latency compared to high intensity stimuli used in the localizer (see e.g. (Elberling, et al., 1981; Lütkenhöner & Klein, 2007)) and extracted the N100m amplitude around a 50 ms wide window around this latency. For statistical comparisons we first performed a 2*2*16 repeated measures ANOVA with factors Condition (random and rhythmic), Hemisphere (left and right), and Stimulus number (ranging from -11.5 to +3.5). We multiplied all the left hemispheric values with -1 since it is evident that via our channel selection the right hemisphere would have the opposite N100m sign as the left hemisphere. As a post-hoc analysis we performed pair-wise comparisons investigating whether the amplitude of the -11.5 was significantly different from all the other time points (corrected for multiple comparisons via false discovery rate).

To test whether we were not extracting motor responses we also epoched the data locked to the response and extracted the motor evoked field. This analysis indicated that the chosen channels did not overlap with the motor evoked response.

References

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity, 11*(1), 669-672.

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2014). Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cerebral Cortex, bhu103*.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation, 7*(6), 1129-1159.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. European *Archives of Psychiatry and Clinical Neuroscience, 87*(1), 527-570.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience, 31*(9), 3176-3185.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology, 116*(12), 2765-2774.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience, 7*(5), 446-451.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*(5679), 1926-1929.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

Elberling, C., Bak, C., Kofoed, B., Lebech, J., & Saermark, K. (1981). Auditory Magnetic Fields from the Human Cortex Influence of Stimulus Intensity. *Scandinavian audiology, 10*(3), 203-207.

Fiebelkorn, I. C., Snyder, A. C., Mercier, M. R., Butler, J. S., Molholm, S., & Foxe, J. J. (2013). Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage, 69,* 126-137.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119.

Lütkenhöner, B., & Klein, J.-S. (2007). Auditory evoked field at threshold. *Hearing research, 228*(1), 188-200.

Miller, K. J., Sorensen, L. B., Ojemann, J. G., & Den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. *PLoS Copmutational Biology, 5*(12), e1000909.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1.

Pike, F. G., Goddard, R. S., Suckling, J. M., Ganter, P., Kasthuri, N., & Paulsen, O. (2000). Distinct frequency preferences of different types of rat hippocampal neurones in response to oscillatory input currents. *The Journal of Physiology, 529*(1), 205-213.

Pritchard, W. S. (1992). The brain in fractal time: 1/f-like power spectrum scaling of the human electroencephalogram. *International Journal of Neuroscience, 66*(1-2), 119-129.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20*(2), 172-176.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience, 24*(1), 1193-1216.

Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion Golumbic, E. M. (2014). The influence of temporal regularities and cross-modal temporal cues on auditory detection. *Neuropsychologia, 63*, 43-50.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*(3), 151-161.

CHAPTER 4

SENSORY ENTRAINMENT EFFECTS ARE STRONGER WHEN USING VARYING ENTRAINMENT LENGTHS

Corresponding Manuscript:
Ten Oever, S., & Sack, A.T. (under review at Journal of Cognitive Neuroscience). Sensory entrainment effects increase with varying entrainment lengths.

Abstract

Sensory entrainment is becoming a popular method to investigate the role of oscillatory phase in perception. By rhythmically presenting stimuli at a specific rate, neuronal responses follow an oscillatory pattern at the presentation rate. Target stimuli that are presented at different delays after the entrainment can subsequently be associated with specific oscillatory phases. To reliably estimate phase effects it is essential to optimize entrainment parameters to induce entrainment as strong as possible. In the current study we investigate the effect of having varying entrainment lengths on entrainment strength. We presented a train of noise bursts at various presentation rates and entrainment lengths after which an ambiguous syllable was presented (this stimulus type has been shown to be phase dependent in a previous study) while recording EEG. Both behavioral and EEG data showed stronger entrainment effects for the shorter compared to the longer entrainment lengths. The enhanced entrainment effect for short trains likely reflects participants' expectation of entrainment continuation for these short trains. Our results provide a way to improve the sensitivity for behavioral phase dependent effects.

Introduction

In recent years the role of oscillation phase in perceptual processes has been systematically uncovered. It has been shown that the phase of oscillations is important for detection (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Fiebelkorn et al., 2013; Ten Oever, Van Atteveldt, & Sack, 2015) and improves reaction times to simple stimuli (Ellis & Jones, 2010; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010). Also the role of oscillatory phase in categorization seems evident (Kayser, Ince, & Panzeri, 2012; Lopour, Tavassoli, Fried, & Ringach, 2013; Ten Oever & Sack, in press; Watrous, Fell, Ekstrom, & Axmacher, 2015). The gained interest in the relevance of phase has triggered different ways to investigate oscillatory processes beyond methods that passively measure oscillations and post-hoc sort trials according to their respective phase. Instead, the causal role of these phases can be explored better by externally inducing these oscillations. One the one hand, this can be done by electrical stimulation with an alternating current (Herrmann, Rach, Neuling, & Strüber, 2013; Riecke, Formisano, Herrmann, & Sack, 2015). On the other hand, sensory entrainment has been extensively used to induce oscillations (Henry & Obleser, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008).

During sensory entrainment one sensory stimulus (e.g. a sound) is repeatedly presented at a specific presentation rate [or a stimulus feature is modulated at a specific rate, e.g. (Henry & Obleser, 2012)], thereby causing a cascade of evoked responses entraining an oscillation at this rate. When target stimuli are subsequently presented at different time points relative to the entrained frequency, a time course evolves, representing different oscillatory phases. In this manner, the role of phase for detection and categorization can be causally inferred.

A problem in sensory entrainment is that there should be no confounding of the entrainment stimuli and the target stimulus. When one systematically presents targets at different time points, one of the time points will overlap with the time point at which the entrainment stimulus is presented in the entrainment stream. Therefore it is a reasonable choice to stop entraining when one wants to present their target stimuli. However, this causes fading of the entrainment, reducing

the amplitude of the oscillation and thereby the strength of the effect (Lakatos et al., 2013).

The speed of the fading of the oscillations might be influenced by different factors. While entrainment seems primarily driven by evoked responses in the brain, there are accounts that attention and expectations also influence the entrainment (Lakatos, et al., 2008; Stefanics et al., 2010; Ten Oever, et al., 2015). For example, when entraining with an auditory and visual stream that are presented in anti-phase, the entrainment phase depends on whether one either attends to the visual or to the auditory stream (Besle et al., 2011; Lakatos, et al., 2008). Expectancy has been shown to influence entrainment by Stefanics and colleagues (2010). They presented a stream of auditory beeps in which some of the beeps indicated that there was a high chance that the next stimulus (which was always in phase with the stimulus train) would be a target. It was shown that entrainment is stronger when there is a higher chance that a following stimulus will be a target. In summary, it thus seems that the strength of entrainment depends on top-down mechanisms such as attention and expectancy, being strongest when participants pay attention to the stimulus train and expect a stimulus to occur.

So how could we use these top-down mechanisms to reduce the fading of the oscillations? Increasing attention to the stimulus train could be difficult, as the participants' task is always directed towards the target stimulus. Moreover, entrainment aids to attend to the predictable, rhythmic moments in time that are in phase with the entrainment stream (Jones & Boltz, 1989; Nobre, Correa, & Coull, 2007; Schroeder & Lakatos, 2009). Since the timing of the target is not in phase with the stimulus train, the entrainment will most likely also be reduced when the stimulus train finishes as the timing of target stimuli is variable. However, expectancy mechanisms might benefit entrainment when using different entrainment lengths that are presented in random order: while for the longest used entrainment length stimulus timing is relatively unpredictable (as only target stimuli can follow), participants have some temporal expectations for shorter entrainment lengths. Specifically, as participants are unaware of the length of the entrainer the likelihood of the following stimulus occurring in phase with the previous sounds is relatively high for short lengths [also see (Correa, Lupianez, & Tudela, 2006)]. It is therefore expected that entrainment paradigms would be

optimized for short lengths embedded with stimulus trains that also last longer.

In the current study we try to probe expectancy mechanisms by having three different entrainment lengths. We used part of the data from a paradigm that has been shown to produce entrainment effects [(Ten Oever & Sack, in press); also additional data was collected] and show that fitting performance was best for the shortest entrainment length used. Moreover, we verified with EEG that at the time points that target stimuli could occur inter-trial coherence was higher for the shortest compared to the longest entrainment length. These results show that sensory entrainment paradigms might be improved by using a multiple of entrainment lengths. This provides better methods to try to find entrainment effects and to discover the causal role of oscillatory phase in any cognitive and perceptual tasks.

Methods

Participants

In total 27 participants (8 male; age range: 18-52, mean age: 24.64) participated in the experiment. Three participants were excluded as they had a ceiling performance and 12 participants underwent EEG. All participants had normal or corrected-to normal vision and gave written informed consent prior to participation. Ethical approval was provided by the local ethical committee of the Faculty of Psychology and Neuroscience at Maastricht University. All participants were rewarded with a monetary compensation for participating.

Stimuli

Entrainment stimuli were bandpassed white noise burst (2.5-3.1 kHz) lasting for 50 ms. They were presented at a rate of either 1, 6.25 or 10 Hz. The length of the entrainment was varied. For 6.25 Hz and 10 Hz, the trains lasted either 2, 3, or 4 seconds. For 1 Hz the trains lasted either 4, 5, or 6 seconds to ensure that enough stimuli were presented to induce entrainment. After the entrainment finished a target stimulus was presented. This stimulus was an ambiguous stimulus that could either be

perceived as /da/ or /ga/ (a morph between /da/ and /ga/; for 7 of the participants of the 6.25 Hz EEG the stimulus was a complex tone, behavioral data not shown). This ambiguous stimulus was presented at different intervals after the last stimulus finished of which all stimulus onset asynchronies (SOAs) exactly covered two periods (the exact timing depending on the frequency; 6.25 Hz: from 0.1 to 0.58 sec in steps of 0.0267 sec. 10 Hz: from 0.1 to 0.28 in steps of 0.017 sec. 1 Hz: from 0.1 to 1.93 in steps of 0.167 sec).

Procedure

First, the most ambiguous stimulus of a da-ga morphed spectrum [see (Ten Oever & Sack, in press) for more details on the morphing] was determined. Morphed stimuli at each point of the spectrum (9 stimuli in total, each presented 15 times) were repeatedly presented and participants had to indicate whether they heard /da/ or /ga/. Trial onset was either 1.6, 1.8, or 2 seconds after the response to the previous trial. The mean proportion /da/ responses was calculated for each morph after which a cumulative Gaussian was fitted with the fitting toolbox modelfree v 1.1 (Zchaluk & Foster, 2009). The morph closest to 50% /da/ identification was used for the rest of the experiment. For the main experiment there were in total 108 trails per entrainment length (27 per SOA). All trials were presented in random order. Participants were required to fixate the screen at all times. Presentation software was used for stimulus delivery.

Not all participants went through all conditions. Specifically, for some of the participants the 6.25 Hz in the EEG data reflects another task, namely tone identification, while the behavioral task was the syllable identification (with different participants). The entrainment however was the same.

EEG acquisition and preprocessing

31 EEG channels were recorded using the easycap M22 set-up with additional channels TP9, TP10, C1, C3 and CPz. Eye movements were recorded from four electrodes above and below the left eye and lateral of both eyes. Online reference and ground were the tip of the nose and AFz

respectively. Data was online recorded with a bandpass filter of 0.1-200 Hz and a sampling rate of 500 Hz using BrainAmp EEG amplifiers (BrainProducts GmBh, Munich, Germany) and BrainVision Recorder (BrainProducts, GmBh, Munich, Germany). Impedance was kept below 15 kOhm (5kOhm for the ground and reference).

Eye blinks were removed using the function scrls_regression of the eeglab plugin AAR [(Gómez-Herrero et al., 2006); filter order: 3, forgetting factor: 0.999, sigma: 0.01, precision: 50]. Then, data was epoched from -3 – 3 seconds around entrainment offset. Trials with extreme variance were removed.

Data Analysis

Behavioral Analysis: The proportion of /da/ response was calculated for each SOA and entrainment length collapsed over all participants. These time courses were each fitted with a sinus using the function lsqnonlin in matlab (mathworks). The frequency of the fit was fixed to the entrainment frequency and we extracted the explained variance. Statistical testing was performed via bootstrapping (n = 1,000). The likelihood of getting such a high explained variance value was estimated for each entrainment length by randomly shuffling the SOA labels and estimating the relevance value again for each entrainment length and presentation rate separately.

In a second analysis we wanted to investigate more directly whether the explained variance reduced for increasing entrainment lengths. To get an estimate that both includes the strength of the decrease over the entrainment lengths and also the variance of diverting from a linear distribution over the three lengths we estimated the statistical value later used for bootstrapping as follows: 1) we subtracted the explained variance of entrainment length 3 from entrainment 1 (which would yield a positive value for decreased explained variance for increased lengths), and 2) subtracted the variance of entrainment length 2 from this difference (variance was calculated as the distance of entrainment length 2 from a line connecting entrainment lengths 1 and 3). This value would be positive if both the decrease in explained variance is high and the dispersion of the middle entrainment length is low. This value was calculated for the original labels and for the bootstrapped labels

(n = 1000) labels. In this analysis we did not bootstrap the SOA, but the entrainment lengths as we were interested in the effect that entrainment length had on the explained variance.

Presentation rates were only analysed when they showed a frequency modulation of syllable identification on the data collapsed over entrainment length (Ten Oever & Sack, in press), that is, presentation rates of 6.25 and 10 Hz (and thus not for 1 Hz and the tone identification). All bootstrapping was performed on the average instead of the individual participants due to a lack of power. This analysis has been performed in other studies as well (de Graaf et al., 2013; Fiebelkorn et al., 2011).

EEG analysis: For each entrainment length and each presentation rate we calculated the inter-trial coherence (ITC) by extracting the phase of the complex Fourier transform calculated via Morlet wavelets using 4 cycles. To estimate the effects of the entrainment we averaged the ITC for the relevant frequencies (1, 6.25, and 10 Hz for the three different presentation rates at an interval +/- 0.75 multiplied by the presentation rate) at all time points that target stimuli could occur for the respective conditions. Channels used for the analysis included the central channels where auditory EEG responses end up (CP2, CPz, CP1, C2, Cz, C1, FC2, FCz, FC1). As we were not interested in overall ITC differences over presentation rates we normalized the ITC values by calculating the z-values over the three entrainment lengths for each participant and presentation rate separately. We entered these z-values in a regression analysis including the predictors entrainment length and the interaction with the entrainment length and frequency (using two dummy variables for presentation rate 6.25 and 10 Hz and the interaction with entrainment length). We used a step-wise method to enter the variables (using SPSS software).

The same analysis as above was repeated for the time period just before entrainment offset (averaging over the same interval as above but before entrainment onset).

Results

Behavior

We fitted for both the 6.25 and 10 Hz presentation rates a sinus (figure 1A). For both presentation rates the explained variance was highest for the shortest entrainment length and lowest for the longest entrainment length (figure 1B). This was also reflected in the bootstrap statistics. P-values dropped for longer entrainment lengths and only for the shorter entrainment lengths showed a significant effect or a trend (6.25 Hz: $p = 0.04$, $p = 0.09$, and $p = 0.91$ for the three respective entrainment lengths. For 10 Hz this is $p = 0.07$, $p = 0.34$, and $p = 0.52$).

To further elaborate on this effect we subtracted the explained variance of entrainment length 3 from entrainment length 1 and subsequently subtracted the variance of the explained variance of entrainment length 2 (figure 1C). This statistical significance was

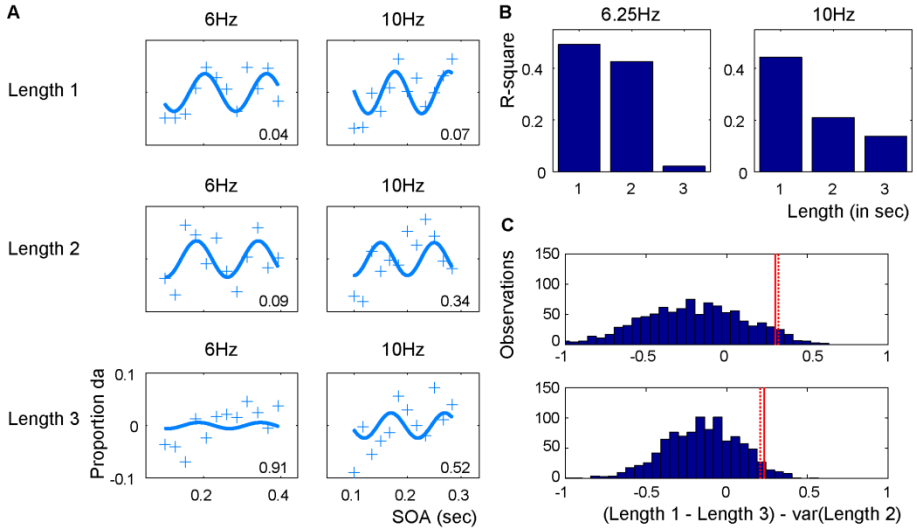


Figure 1. Behavioral results. A) Time course of proportion /da/ responses for each entrainment length and presentation rate the original data (crosses) and fitted data (solid line) is shown with mean subtracted. Longer entrainment length typically have a worse fit compared to short lengths. B) The explained variance for the fit for each entrainment length. C) The histograms of the bootstrapped data for the comparison of the strength of the entrainment length effect. The red solid lines indicate the value of the original data. The red dotted lines indicate the 95 percentile of the bootstrapped data.

determined via bootstrapping the entrainment length labels. Results showed a significant effect for 10 Hz (p=0.04) and a trend for 6.25 Hz (p=0.06) indicating that indeed it seemed that the explained variance parametrically reduced dependent on entrainment length.

EEG

Figure 2A displays the ITC for all presentation rates and entrainment lengths around entrainment offset. Especially in the 1 Hz condition only for the shortest entrainment length there is strong 1 Hz ITC after the entrainment finished. This effect is also present for the other two entrainment lengths, but less strongly. Also strong ITC are visible at other frequency bands, especially in the 10 Hz condition. However, the time point directly after the entrainment is highly influenced by the evoked response to the target stimulus. This is especially strong in the 10 Hz condition as the time interval of target presentation is very narrow.

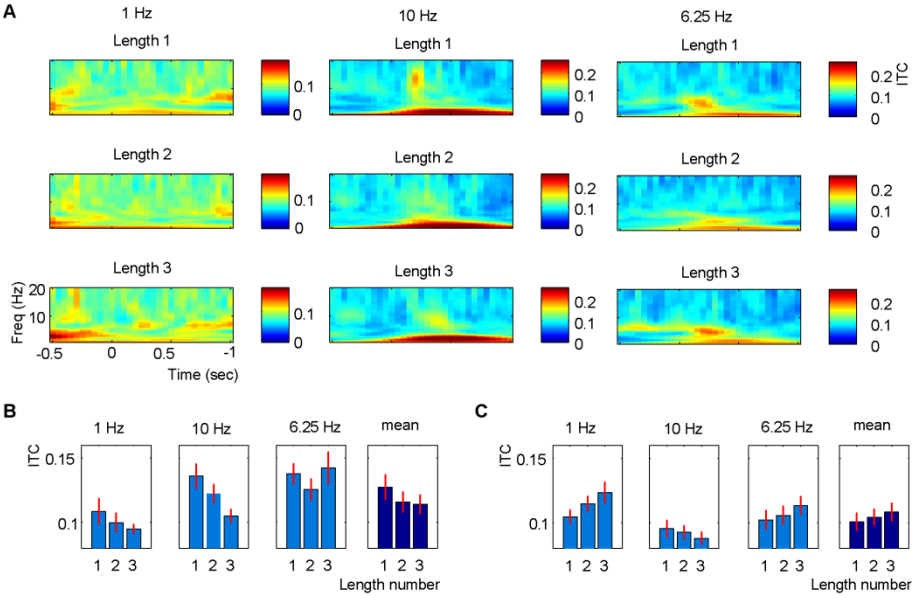


Figure 2. EEG results. A) The ITC values for each presentation rate and entrainment frequency. Time point zeros indicates entrainment offset. B) The extracted ITC values at the interval at which target stimuli could be presented. C) The extracted ITC values at the interval before target stimuli could be presented. Error bars indicate the standard error of the mean.

We extracted the ITC at the relevant frequency points for each of the three presentation rates at the interval at which target stimuli could be presented and entered the z-normalized scores in a multiple regression (including interaction terms). The final regression model only included the predictor entrainment length [$F(1,106) = 9.921$, $p = 0.002$; r-square = 0.293]. This indicates that for longer entrainment lengths ITC values were lower (figure 2B). Although in figure 2B it seems that the 6.25 Hz presentation rate showed a different ITC pattern over entrainment lengths, this was not statically demonstrated by a significant interaction between entrainment length and 6.25 Hz presentation rate.

Figure 2C shows the ITC values for the time interval prior to entrainment offset. The overall pattern seems to indicate a slight increase in ITC for longer entrainment lengths. The conducted regression did not include any predictors in the model indicating that there was no linear pattern for the entrainment length. However, when just including entrainment length there was a trend suggesting stronger ITCs for longer entrainment lengths [$F(1,106) = 3.835$, $p = 0.053$; r-square = 0.187].

Discussion

We investigated the effect of entrainment length on the strength of sensory entrainment with EEG as well as behavioral responses. We presented auditory noise bursts at presentation rates of 1, 6.25, and 10Hz after which ambiguous syllables were presented after three different entrainment lengths. It was predicted that for shorter entrainment lengths effects would be stronger as participants are unaware of the length of the stimulus train and it is therefore likely that entrainment would continue for shorter lengths. We found that behavioral entrainment as well as inter-trial coherence was stronger for shorter entrainment lengths. The current results show how sensory entrainment protocols could be improved. This modulation in entrainment protocols could thereby enhance the sensitivity of detecting behavioral phase effects.

Improving sensory entrainment protocols

There is growing interest to investigate the causal role of oscillatory phase for behavior. This requires optimizing sensory entrainment protocols. The current results show one way to improve entrainment protocols by adding variable entrainment lengths. In this way entrainment effects will be the strongest for the shortest entrainment length used as participants expect the entrainment to continue. Alternatively, expectation mechanisms could be probed by having a continuous stream of entrainment in which target stimuli are presented at random time points which are unknown to the participant [see e.g. (Mathewson, et al., 2010)].

Besides the effect of temporal expectations that can be counteracted with having varying entrainment lengths, also other parameters can influence the final behavioral outcome. For example, one also has to decide the presentation rate of stimulation, which is often inferred from EEG or electrophysiology research, as well as the specific stimulus properties of the entrainer. Would one systematically modulate a specific feature that induces the oscillations, such as for example a frequency modulated tone (Henry & Obleser, 2012), or would one rather repeatedly present the same stimulus at a specific rate (Jones, Johnston, & Puente, 2006; Lakatos, et al., 2008)? Furthermore, the stimulus choice should predominantly depend on the region one wants to entrain. Primary auditory regions should be stimulated with narrow band noise, while higher order regions are more sensitive to broadband noise (Kaas & Hackett, 1998). Moreover, many auditory regions are frequency selective and it has been shown that regions non-selective for the entrainer frequency might show the opposite entrainment compared to regions selective for this frequency (Lakatos, et al., 2013). Correct choice of entrainment parameters is vital to ensure that the correct areas are entrained at the correct time points with minimal amount of fading of the evoked oscillation. In this way, sensory entrainment protocols could become a more powerful tool to investigate the role of oscillatory phase of perception.

Proactive expectations of stimulus occurrence

There is a growing consensus that our brain is not a passive monitor of sensory input but rather proactively acts towards future input by predicting the input and selectively changing its response (Friston, 2005; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010). The current results also reflect a mechanism of expectations: for short entrainment lengths participants expect the entrainment train to continue, thereby producing a stronger entrainment. This is consistent with findings of stronger entrainment for increased target expectation (Stefanics, et al., 2010). An alternative explanation is that for longer entrainment trains there is more response suppression (Bourbon, Will, Gary, & Papanicolaou, 1987; Näätänen, Paavilainen, Rinne, & Alho, 2007; Ritter, Vaughan, & Costa, 1968), selectively reducing the response to subsequent stimuli. We believe this explanation is not very likely as for most mismatch paradigms maximal suppression is already present for a small amount of repetition of stimuli (Näätänen, et al., 2007). Moreover, there was no evidence of reduced entrainment for longer trains just before the entrainment finished. On the contrary, there was an indication that ITC increases with longer entrainment. We therefore hold that the most likely explanation is that entrainment length influences the temporal expectations of participants of entrainment continuation.

Conclusion

As sensory entrainment is becoming a more popular paradigm to investigate effects of oscillatory phase it is vital to optimize all entrainment parameters to ensure detecting behavioral phase effects. Our study provides one way to enhance entrainment by inducing varying entrainment lengths. Besides the improved methodology we also provide evidence for the proactive nature of the brain (Bar, 2011; Schroeder, et al., 2010; Summerfield & Egner, 2009), which rather than only directly responding to the presented stimuli, responds differentially when stimuli are expected. Our study provides contributions revealing how attention is selectively driven to specific points in time (Nobre, et al., 2007; Ten Oever, Schroeder, Poeppel, Van Atteveldt, & Zion Golumbic, 2014) and how this mechanism can be used to optimize entrainment in experimental settings.

References

Bar, M. (2011). *Predictions in the brain: Using our past to generate a future:* OUP USA.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience, 31*(9), 3176-3185.

Bourbon, W. T., Will, K. W., Gary, H. E., & Papanicolaou, A. C. (1987). Habituation of auditory event-related potentials: a comparison of self-initiated and automated stimulus trains. *Electroencephalography and clinical neurophysiology, 66*(2), 160-166.

Correa, A., Lupianez, J., & Tudela, P. (2006). The attentional mechanism of temporal orienting: Determinants and attributes. *Experimental Brain Research, 169*(1), 58-68.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

de Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., & Thut, G. (2013). Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. *PloS one, 8*(3), e60035.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics, 72*(8), 2274-2288.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *The Journal of Neuroscience, 31*(27), 9971-9981.

Fiebelkorn, I. C., Snyder, A. C., Mercier, M. R., Butler, J. S., Molholm, S., & Foxe, J. J. (2013). Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage, 69,* 126-137.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815-836.

Gómez-Herrero, G., De Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Van Huffel, S., et al. (2006). *Automatic removal of ocular artifacts in the EEG without an EOG reference channel.* Paper presented at the Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, 109*(49), 20095-20100.

Herrmann, C. S., Rach, S., Neuling, T., & Strüber, D. (2013). Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Frontiers in human neuroscience, 7.*

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive psychology, 53*(1), 59-96.

Kaas, J. H., & Hackett, T. A. (1998). Subdivisions of AuditoryCortex and Levels of Processing in Primates. *Audiology and Neurotology, 3*(2-3), 73-85.

Kayser, C., Ince, R. A., & Panzeri, S. (2012). Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS computational biology, 8*(10), e1002717.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*(4), 750-761.

Lopour, B. A., Tavassoli, A., Fried, I., & Ringach, D. L. (2013). Coding of Information in the phase of local field potentials within human medial temporal lobe. *Neuron, 79*(3), 594-606.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition, 115*(1), 186-191.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*(12), 2544-2590.

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology, 17*(4), 465-470.

Riecke, L., Formisano, E., Herrmann, C. S., & Sack, A. T. (2015). 4-Hz Transcranial Alternating Current Stimulation Phase Modulates Hearing. *Brain stimulation, 8*(4), 777-783.

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: a study of short terms changes in average evoked responses. *Electroencephalography and clinical neurophysiology, 25*(6), 550-556.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20*(2), 172-176.

Stefanics, G., Hangya, B., HernÃ¡di, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of Neuroscience, 30*(41), 13578-13585.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Ten Oever, S., & Sack, A. T. (in press). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*

Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion Golumbic, E. M. (2014). The influence of temporal regularities and cross-modal temporal cues on auditory detection. *Neuropsychologia, 63,* 43-50.

Ten Oever, S., Van Atteveldt, N., & Sack, A. T. (2015). Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. Journal of *Cognitive Neuroscience, 27*(9), 1811-1822.

Watrous, A. J., Fell, J., Ekstrom, A. D., & Axmacher, N. (2015). More than spikes: common oscillatory mechanisms for content specific neural representations during perception and memory. *Current Opinion in Neurobiology, 31,* 33-39.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*(6), 1414-1425.

CHAPTER 5

INCREASED STIMULUS EXPECTANCY TRIGGERS LOW-FREQUENCY PHASE RESET DURING RESTRICTED VIGILANCE

Corresponding Manuscript:
Ten Oever, S., Van Atteveldt, N., & Sack, A. T. (2015). Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. *Journal of Cognitive Neuroscience, 27*(9), 1811-1822.

Abstract

Temporal cues can be used to selectively attend to relevant information during abundant sensory stimulation. However, such cues differ vastly in the accuracy of their temporal estimates, ranging from very predictable to very unpredictable. When cues are strongly predictable, attention may facilitate selective processing by aligning relevant incoming information to high neuronal excitability phases of ongoing low-frequency oscillations. However, top-down effects on ongoing oscillations when temporal cues have some predictability, but also contain temporal uncertainties, are unknown. Here, we experimentally created such a situation of mixed predictability and uncertainty: a target could occur within a limited time window after cue, but was always unpredictable in exact timing. Crucially to assess top-down effects in such a mixed situation, we manipulated target probability. High target likelihood, compared to low likelihood, enhanced delta oscillations more strongly as measured by evoked power and inter-trial coherence. Moreover, delta phase modulated detection rates for probable targets. The delta frequency range corresponds with half-a-period to the target occurrence window and therefore suggest that low-frequency phase-reset is engaged to produce a long window of high excitability when event timing is uncertain within a restricted temporal window.

Introduction

Ongoing neuronal oscillations reflect fluctuations of neuronal ensembles between low and high excitable phases, during which incoming stimuli are less or more efficiently processed, respectively (Buzsáki & Draguhn, 2004; Lakatos et al., 2005). When events are highly temporally predictive, for example in rhythms, high excitable phases can be aligned to the arrival time of incoming information, effectively changing the phase of ongoing oscillations (Kayser, Logothetis, & Panzeri, 2010; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008). The strength of this phase-reset increases when extra top-down resources are employed, for instance when there is an expectancy cue about the occurrence of relevant stimuli (Stefanics et al., 2010). In contrast, during absence of any temporal predictability - for example when a cat is waiting for a mouse to exit its mouse hole without any cues of when this will happen - Schroeder and Lakatos (2009) proposed a 'vigilance mode' of processing. In this mode, increased attention enhances the amplitude of high-frequency gamma oscillations to produce many, densely distributed, high excitable phases, and hereby optimizes the chance that an unpredictable stimulus will arrive during a high excitable phase.

Dichotomizing all possible temporal contexts in 'predictable' versus 'unpredictable' might not be sufficient since in many natural situations there is a mixture of both regularities and uncertainties in our temporal estimates, which has been tested in variable foreperiod studies (Los, Knol, & Boers, 2001; Niemi & Näätänen, 1981; Wright & Fitzgerald, 2004). Considering the cat-mouse-hole example, when the cat briefly hears the mouse toddling, its attention will be raised for a short while, because the mouse might come out very soon. In this example there are temporal cues (i.e. the sound of the mouse) which indicate that an event is expected to occur soon, but their temporal precision is not very accurate, leaving the exact event timing uncertain. With these kinds of intermediate temporal cues, it seems beneficial to allocate attention continuously for a constrained time, which we will define here as restricted vigilance. We hypothesize that low-frequency oscillations will be reset when attending for a restricted period of time, to ensure high-excitability during the entire window of stimulus occurrence while not using metabolically demanding gamma oscillations (Mukamel et al., 2005; Niessing et al.,

2005). These low-frequency oscillation would thereby efficiently cover the restricted vigilance window with the high-excitability half of the oscillation period.

To investigate top-down influences on phase-reset mechanisms during a situation of restricted vigilance we presented a blue or yellow circle which indicated a high (80%) or low (50%) probability of a target sound following. This probability manipulation allowed us to vary the top-down expectancy about the probability that a temporally unpredictable, but behaviorally relevant event will occur while keeping the restricted vigilance window constant. After these expectancy cues, low-intensity auditory stimuli were presented in the corresponding proportion of the trials, with stimulus onset asynchronies (SOA's) ranging from 0 to 450 ms, while recording EEG. Our findings reveal enhanced low-frequency oscillations (1-3 Hz) in evoked power as well as inter-trial coherence in the high-probability condition. The restricted vigilance window matches with half-a-period the revealed oscillatory frequency range, suggesting that the whole temporal attention window may be enclosed within the high excitability phase of the low-frequency oscillations, while the low excitable phases fall outside the window in which stimuli are expected. This low-frequency oscillation enhancement was significantly stronger in case of high as compared to low sound likelihood and moreover, only in the high probability condition delta phase modulated auditory detection. This influence of probability suggests that when event occurrence is likely delta phase is consistent over trials and this phase determines whether stimuli are perceived or not, while for unlikely events there is less phase consistency and subsequent delta phases do not determine the percept. These results reflect that phase-reset of low-frequency oscillations lead to a longer temporal attention window for likely relevant events that are presented in a restricted temporal window and underline the flexibility of phase-resetting as an important mechanism underlying selective processing.

Materials and Methods

Participants

Fourteen healthy volunteers took part in our study (mean age: 24, range 19-35, 7 male). All participants reported to have normal hearing and normal or corrected to normal vision. Before participating all gave informed consent. Ethical approval was given by the Ethical Committee of the Faculty of Psychology and Neuroscience at the University of Maastricht. Participants received a monetary compensation. One participant was excluded from the analyses; see section data analyses for details.

Stimuli and procedure

First, the individual detection threshold of participants was determined with the method of constant stimuli. Low intensity sounds (1 kHz beeps lasting 75 ms with on and off ramp of 5 ms) varying from 27 dB up to 42 dB in steps of 1.56 dB were presented in constant white noise (46 dB) and participants had to indicate whether they detected the sounds or not. A trial was 1 second long and the sound onset was randomized between 300-800 ms after trial onset. In total twenty stimuli were presented per intensity. Thereafter a cumulative Gaussian was fitted using the psychometric fitting toolbox modelfree v 1.1. (Zchaluk & Foster, 2009), implemented in MATLAB (mathworks), and the 70% detection threshold was calculated. This intensity was used in the main experiment. During the experiment detection rates between 30% and 85% were ensured by manually changing the intensity if the intensity was not in this range for a block.

After the threshold was determined the EEG cap was mounted and the main experiment started. In this experiment, a trial consisted of the presentation of a visual circle (visual angle 15 degrees, color blue (rgb: 0, 191, 255) or yellow (rgb: 238, 238, 0), lasting 75 ms), after which the low intensity sound was presented or not (Fig. 1). The task of the participant was to indicate whether they heard a sound or not on a four point scale (1 = I did not hear the sound, 2 = I think I did not hear the sound, 3 = I think I did hear the sound, 4 = I did hear the sound). The auditory stimulus

could be presented at the same time as the visual stimulus or up to 450 ms after the visual stimulus in steps of 50 ms (of which the order was randomized). Additionally, the sound could also be presented within the interval 1-1.5 second to ensure participants were still paying attention at the end of the trial. The detection question did not appear before the end of the trial (lasting 1.8 sec) and a random delay was inserted varying between 500 and 800 ms before the onset of the next trial. The probability of sounds presented depended on the visual color. For one color, a sound was presented in 50% of the trials, in 80% for the other color. The specific colors with a low vs. a high probability of sounds were counterbalanced over participants. Participants were informed about this difference by the instruction that for one color there was a high chance of the occurrence of a sound and for the other color there was a low chance of the occurrence of a sound. Hereby, we manipulated the participants' top-down expectancies of sound occurrence.

In total there were 1300 trials, resulting for the low probability in 25 trials per SOA (75 for SOA = 0) and 325 trials without a sound and for the high probability in 40 trials per SOA (120 for SOA = 0) and 130 trials without a sound. There were more trials of SOA 0 to ensure to participants were focusing at the beginning of the trial (see also (Fiebelkorn et al., 2011)). All trials were divided over 10 blocks lasting approximately 5 minutes each. Background color was grey (rgb: 100, 100, 100), and a black fixation cross was presented throughout the experiment. Participants were seated approximately 57 cm from the screen and Presentation software was used for stimulus delivery (Neurobehavioral

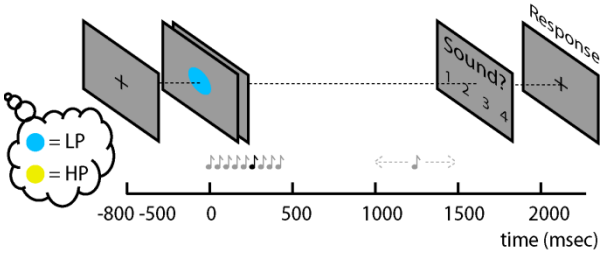


Figure 1. Example of a trial. After a variable delay a visual cue was presented. The color of the visual cue indicated high or low probability of auditory stimulus occurrence (LP or HP). One auditory stimulus (black) was presented per trial. All grey notes indicate when the auditory stimulus could occur. After 1.8 seconds the question trial appeared in which participants had to indicate whether they detected the stimulus or not.

Systems, Inc., Albany, NY). Participants were instructed to fixate the entire trial and to try to blink only after the question appeared on the screen.

EEG acquisition and preprocessing.

EEG data were recorded (DC-200 Hz, sampling rate 500 Hz) in a sound-attenuated and electrically shielded room with a 61 channel cap (Easycap, Montage No. 1), and two BrainAmp Standard EEG amplifiers (BrainProducts GmbH, Munich, Germany). The left mastoid was used as reference and Afz as ground. Three additional electrodes were placed to record eye movements (below the left eye, and at the lateral sides of both eyes). Impedance levels were kept below 15 kΩ. Data were analyzed using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011), the circular statistics toolbox (Berens, 2009), and custom MATLAB scripts.

Preprocessing steps were as follows. First, data were re-referenced to the average of the left and right mastoid. This reference was kept throughout all data analyses to keep the reference of the different analyses steps the same, but therefore might differ from commonly used average reference topographies. Second, data were notch filtered (50 Hz) to remove electrical noise. Additionally, epochs were created for all the trials (-1-3 sec relative to the visual stimulus), the mean of each single epoch was subtracted for all epochs, and data were resampled to 256 Hz. Independent component analysis was performed to remove blink and muscle artifacts (fast ICA with 50 PCA components). Remaining trials with high variance were removed by visual inspection.

Data analyses

Behavioral analyses: We calculated for each SOA and each conditions the sensitivity and bias (Green & Swets, 1966). Hits were defined as all the trials in which there was a sound and the participants pressed button 3 or 4. Misses were trials with a sound and participants choose option 1 or 2. If there was no sound presented, but participants pressed 3 or 4, there was a false alarm. Since any false alarm cannot be associated with one specific

SOA (as there is no sound), the sensitivity and bias estimates were calculated using the same false alarm rate for each SOA. A repeated measurement ANOVA with the factors SOA (eleven SOA points) and probability (low and high probability) was conducted with Greenhouse-Geisser correction for both the sensitivity and the bias. Post-hoc pairwise comparisons were conducted for significant effects, correcting for multiple comparisons via false discovery rate. One participant had a difference of 60.8% between the false alarm rates of the two conditions, indicating that the participant based his choices purely on the visual stimulus, and was therefore excluded.

EEG analyses

ERP analyses: Single epochs were band-pass filtered from 0.05-20 Hz (second order Butterworth filter) and corrected for the -200-0 ms baseline. Only epochs in which no sound was presented and in which participants indicated that there was no sound (button 1 or 2) were used for the EEG analyses (baseline correction and epoch selection is the same for all henceforth described analyses excluding the delta phase analysis). We used only these trials 1) because the participant would stop attending after hearing a possible sound and 2) we intended to focus our analyses on trials without any auditory evoked responses. For each individual, trials were averaged per condition and the two conditions were compared with each other using cluster analyses implemented in Fieldtrip (Maris & Oostenveld, 2007). On average for the low probability condition 248 trials were used (standard deviation is 59.3), and for the high probability condition 93 (standard deviation is 24.5). The same trials were used in the evoked and total power analyses. For the cluster analysis, first the paired samples t-values are calculated for all channels and time points (0.05-1.8 sec after visual stimulus onset and after the question onset), then clusters are defined based on these t-values and statistical significance is determined via Monte Carlo randomizations (the following parameters were used: cluster alpha of 0.05, dependent samples t-test alpha of 0.01, 10,000 randomizations, and the maximal sum of all the time and channel bins in one cluster as dependent variable. All reported p-values reflect a two-sided test).

Time frequency analyses – evoked power: Primarily, we were interested in the power spectrum directly evoked by the stimulus, that is, the stimulus phase locked power responses. This analysis would give us an idea which frequency bands could relate to a phase reset (Makeig, Debener, Onton, & Delorme, 2004). We therefore averaged the epochs for each individual for each condition and applied hanning tapers (time window linearly rising from 2 cycles at 1 Hz up to 10 cycles at 40 Hz. Frequencies below 1 Hz had a time window of 2 seconds) for frequency range 0.2-40 Hz and time range of 0-1 sec. Data were baseline corrected for the -0.2-0 seconds time interval. Also here a cluster analysis was performed. Since we were only interested in the frequency of the evoked response we averaged over 0.08-0.40 sec (since it had the highest amplitudes in the ERP results) and calculated significant frequency clusters with the same analyses as before.

Time frequency analyses – total power responses: To all individual epochs the same hanning tapers as for the phase locked time frequency analyses were applied. Thereafter, the power spectra were averaged within subjects. We analyzed using channel Fz (since it had the strongest effect for the evoked power), and took the same time and frequency range as in the phase locked time frequency analyses (again averaging over time). Data were baseline corrected for the -0.2-0 seconds time interval. We used Monte Carlo simulations (10,000 repetitions) for all the frequency points separately, using the t-values (of the paired t-tests) as dependent variable (Maris & Oostenveld, 2007). This method creates a simulated distribution of t-values by shuffling the labels of the two conditions and repeating the t-test calculation for these shuffled labels. This is done 10,000 times and the subsequently reported p-values reflect the proportion of shuffled labels that have a higher t-value as the original t-value. We report the average p and t-value of the significant frequency bins. We corrected for multiple comparison using the false discovery rate.

Inter-trial coherence: Inter-trial phase coherence characterizes how consistent the phases of different frequencies are over multiple trials, independent of power (Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996), and therefore provides an indication for pure phase resetting. We calculated for the single epochs Fourier spectra for Fz using hanning tapers for 0.2-40 Hz with the same parameters as for the time frequency analyses. Thereafter, we extracted the phases of the Fourier spectra of the

single epochs and calculated for each participant and each condition the inter-trial coherence (ITC). Since the ITC is inflated with fewer trials (i.e. trials with a high probability of a sound had fewer trials without a sound than trials with a low probability of a sound), we applied permutations to the low probability trials. Therefore, we first only calculated the ITC for the high probability trials. Then, we randomly selected an equal amount of trials for the low probability condition and calculated the ITC. This randomized trial selection procedure and ITC calculation was repeated 500 times and we took the mean of the repetitions as the ITC for the low probability trials. Then we used the same averaged time interval (0.08-0.4 sec) and frequency range (0.2-40 Hz) for statistics with the same methods as for the total power time frequency analysis.

Delta phase during misses and hits: To test whether indeed delta phase is important for detecting the stimuli we sorted all the trials containing sounds to hits and misses per probability condition. Thereafter we filtered all the data around delta (second order IIR Butterworth filter, using a causal bandpass filter with cut-off frequencies at 1-1.75 Hz, cut-off = -3 dB) and extracted per participant the mean angle and inter-trial coherence of the Hilbert transformation at sound onset. Across participants there were on average 104 hit trials and 132 miss trials for the low probability condition (standard deviation is 45.0 and 46.7 respectively), and 159 hit trials and 220 miss trials for the high probability condition (standard deviation is 57.4 and 65.7 respectively). A causal filter was used since we wanted to exclude any effects that could be due to differences in the evoked response between hits and misses (Zoefel & Heil, 2013). To estimate whether an interaction effect between probability condition and detection exists we first calculated the circular distance between hits and misses for both conditions separately for each participant. Then, we used the Zar's Hotelling test to investigate whether these distances have a different mean angle for different conditions (van den Brink, Wynn, & Nieuwenhuis, 2014; Zar, 1998). This test has the advantage that it takes into account the inter-trial coherence such that mean angles corresponding to a low inter-trial coherence are also considered to be less consistent in phase. Since for the distance measure two inter-trial coherences need to be considered (i.e. the one from the hits and the one from the misses), we choose to incorporate the minimum inter-trial coherence of the two. This seemed valid as the lowest ITC of

the hits and misses determines how consistent the overall phase difference will be. Thereafter, we used the same Zar's Hotelling test to test whether the mean angle for the hits and misses are different for the two conditions separately. All trials with SOA's ranging between 0 and 100 ms were excluded since effects of phase-reset are unlikely at such an early SOA considering the different transmission latencies between auditory and visual responses (see e.g. (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008)).

Results

Behavioral results

Sensitivity: The repeated measure ANOVA showed a main effect of SOA ($F(10,120) = 8.647$, $p < 0.001$; Fig 2A). Pairwise comparison showed that all the SOA points were significantly different from the 1000-1500 ms time point (Table 1). Furthermore it seems that the middle time points are detected best relative to earlier and later time points. This is confirmed with a significant quadratic contrast ($F(1,12) = 21.995$, $p = 0.001$). No higher order polynomials were significant. This contrast was also significant when excluding the last time point ($F(1,12) = 6.534$, $p = 0.025$). The main effects of Probability and the interaction of Probability*SOA were not significant ($F(1,12) = 1.700$, $p = 0.217$ and ($F(10,120) = 0.478$, $p = 0.776$, respectively).

Bias: The bias estimate showed the reversed pattern of the sensitivity (Fig. 2B). Again there was a main effect of SOA ($F(10,120) = 8.647$, $p < 0.001$) and a significant quadratic contrast ($F(1,12) = 21.995$, $p = 0.001$. The main effect of Probability and the interaction were not significant, although the main effect showed a slight trend of higher bias for the low probability condition ($F(1,12) = 3.180$, $p = 0.100$ and $F(10,120) = 0.478$, $p = 0.776$). As the false alarm rate does not change for different SOAs and both the bias and sensitivity are calculated as a linear transformation relative to the false alarm rate the pairwise comparisons for the main effect of SOA are the same for the bias as for sensitivity and therefore not reported.

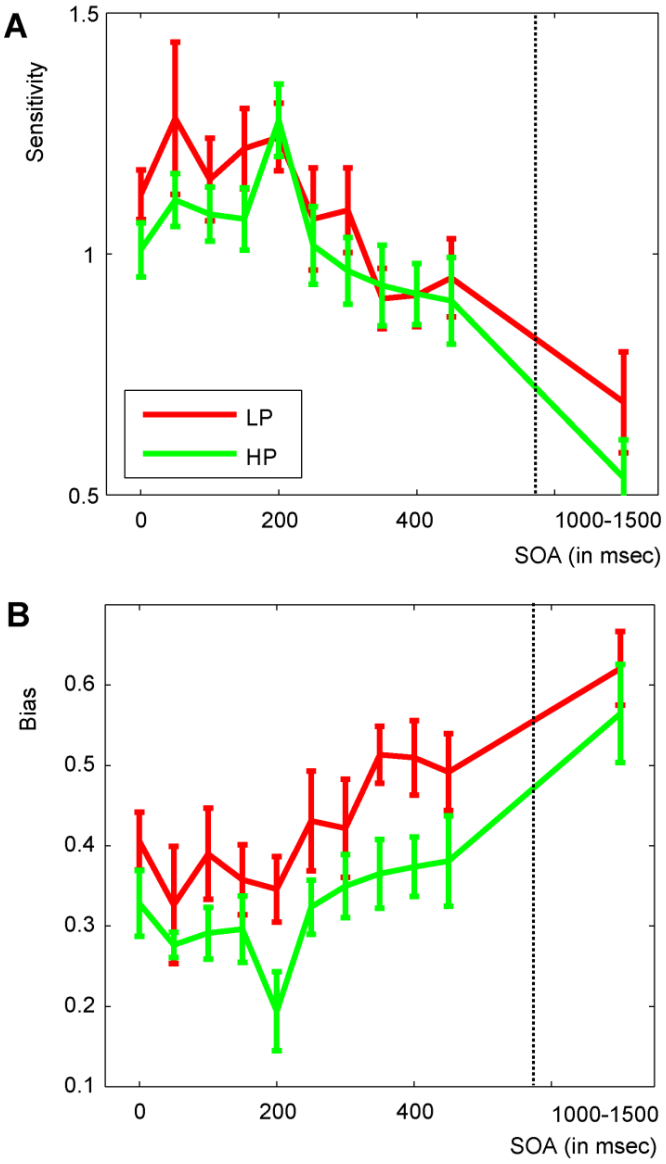


Figure 2. Behavioral results. A) Sensitivity and B) bias for the high (green) and low (red) probability condition over SOA. Error bars indicate the standard error of the mean. To calculate the SEM the individual data were normalized by subtracting the overall mean of the sensitivity/bias for the individual participants.

Although in this behavioral analyses no clear difference between the two probability conditions was present, the later EEG-behavioral analyses indicates a more subtle behavioral difference between these conditions.

EEG results

ERP results: The event-related potentials (ERP's) show two clear evoked responses, one in response to visual stimulus onset and one in response to question onset (Fig 3A). The cluster analyses (see methods for details) showed no significant effects of condition (high vs low probability) for the whole interval up to the question (0-1.8 sec) and after the question

Table 1. Crosstab of all the Pairwise Comparisons

	50		*100*		*150*		*200*		*250*	
0	-1.71	(.190)	-.74	(.554)	-1.83	(.171)	*-2.78*	*(.044)**	.21	(.894)
50			.75	(.554)	.62	(.628)	-.75	(.554)	1.51	(.240)
100					-.46	(.717)	-1.87	(.163)	.87	(.503)
150							-1.42	(.269)	1.21	(.327)
200									2.33	(.084)
250										
300										
350										
400										
450										

	300		*350*		*400*		*450*		*1000-1500*	
0	.50	(.706)	2.09	(.120)	*2.91*	*(.038)**	1.74	(.184)	*4.21*	*(.007)***
50	1.62	(.214)	2.65	(.053)	*3.47*	*(.021)**	2.11	(.120)	*4.95*	*(<.001)***
100	1.17	(.336)	2.57	(.060)	*4.03*	*(.010)***	2.34	(.084)	*5.04*	*(<.001)***
150	1.38	(.272)	2.68	(.052)	*4.09*	*(.010)***	*3.05*	*(.037)**	*5.41*	*(<.001)***
200	*2.93*	*(.038)**	*5.27*	*(<.001)***	*5.51*	*(<.001)***	*3.08*	*(.037)**	*6.87*	*(<.001)***
250	.18	(.894)	1.38	(.272)	1.77	(.181)	1.35	(.280)	*3.90*	*(.010)***
300			1.54	(.236)	2.03	(.128)	1.29	(.297)	*4.14*	*(.007)***
350					.09	(.930)	-.07	(.894)	*3.95*	*(.021)***
400							-.16	(.894)	*3.44*	*(.038)**
450									*2.96*	*(.038)**

All values in the table represent the results of the t-test for the two stimulus onset asynchronies (SOA's) corresponding to the SOA's of the row and column. The initial number is the t-value and the number between brackets the corrected p-value. Asterisk and double asterisks correspond to significance at the 0.05 and the 0.01 level, respectively.

(1.8-2.5 sec).

Time frequency analyses – evoked power: The evoked power analysis revealed a higher evoked response for low delta frequencies (Fig 3B; 1-3.25 Hz, cluster statistics = 462.19, p = 0.042) during the high probability condition that was frontal/centrally organized (Fig 3C). Also, a higher evoked alpha (9.5-17.25 Hz) response was found at occipital

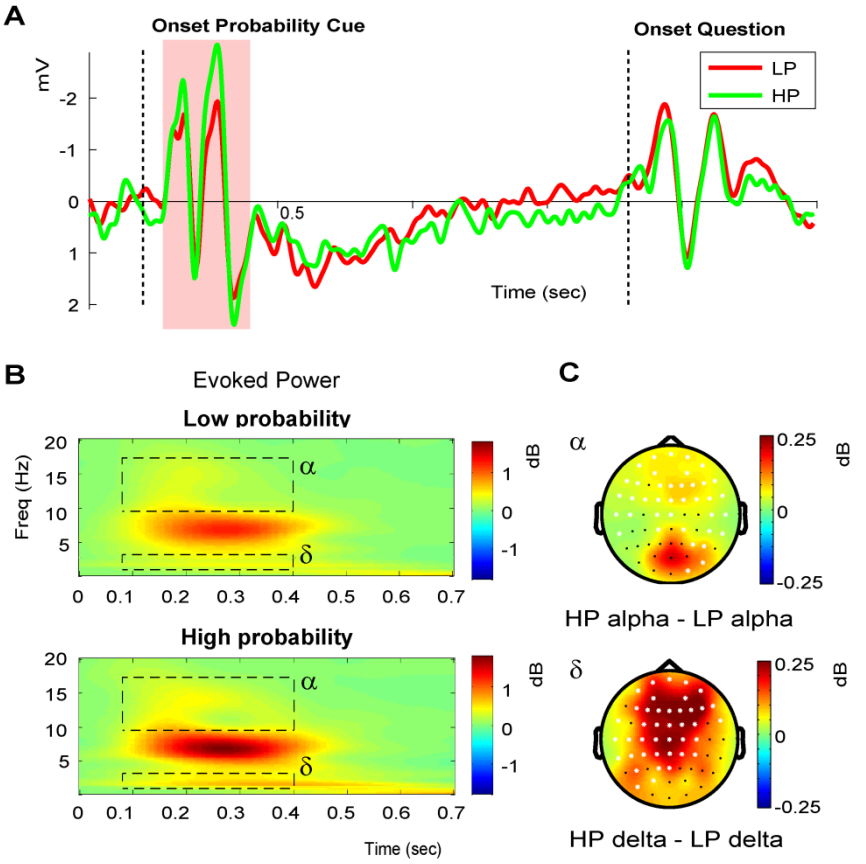


Figure 3. Event related potentials (ERP's) and evoked power for the trials without a sound. A) ERP of Fz for the whole trial time course. The red shaded area indicates the interval used for subsequent analyses. B) Phase locked time frequency spectra with two significant clusters ($p < 0.05$) estimated within the 0.08-0.4 time window, indicated by the two black rectangles. C) The corresponding topographic distributions with white asterisks indicating the significant channels. Topographies show the data of the high minus the low probability.

channels for the high probability condition. Although the highest difference values were occipitally located, a significant frontal cluster was found for this effect (cluster statistics = 1511.1, p = 0.004). This cluster likely relates to the visual evoked response. Although the figure might suggest a difference in the 5-10 Hz frequency range, no significant cluster was found.

Time frequency analyses – total power: The total power showed no significant effect at delta range. However, a significant difference at alpha range was found (Fig 4A; 11.5-15 Hz; average t-value over all significant alpha frequency bins (11) = 2.21, average p-value = 0.026), with stronger alpha power for the high probability condition.

Inter-trial phase coherence: The ITC plot shows a significant stronger delta ITC for the high probability conditions (Fig 4B; 1-1.75 Hz,

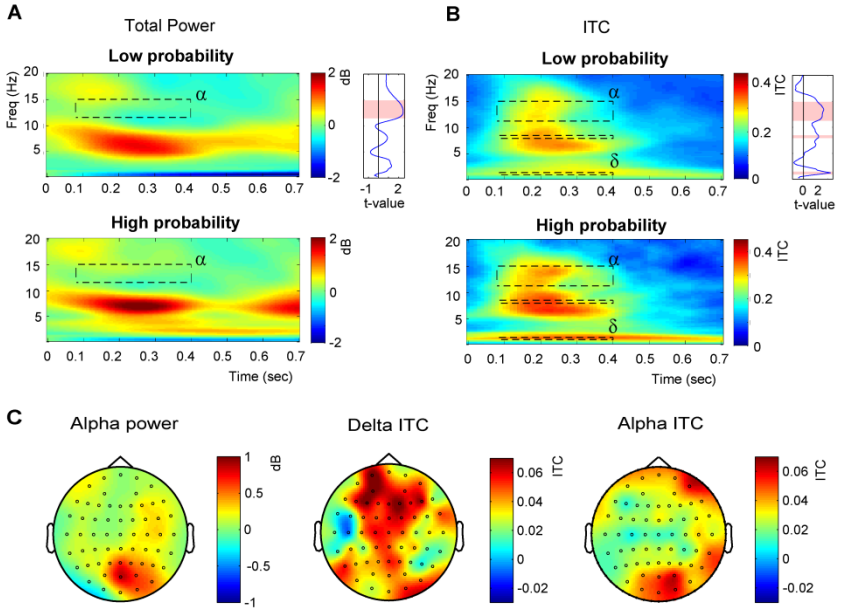


Figure 4. Power and inter trial coherence (ITC) estimates in Fz. A) Induced power for both conditions. Insert on the right shows t-values of the Llow vs. Hhigh probability comparison for the entire frequency range. B) ITC for both conditions. Insert on the right shows t-values of the low vs. high probability comparison for the entire frequency range. C) Difference topographic distributions of three different significant clusters (alpha power, delta ITC, and alpha ITC (cluster between 11.25 and 15 Hz)). Red shaded areas and black rectangle indicates significance ($p < 0.05$) within the 0.08-0.4 time window.

average t-value(11) = 2.89, average p-value = 0.01). Additionally, a strong response at alpha band is again present, significantly stronger for the high probability condition which was separated in two different bands in alpha range (from 8-8.5 Hz, average t-value(11) = 2.03, average p-value = 0.037, and from 11.25-15.00, average t-value(11) = 2.40, average p-value = 0.019).

Delta phase during misses and hits: The polar plots in figure 5 show the angle distribution separate for hits and misses and separate for the two conditions when SOA's from 150-450 ms were included. It seems that the mean direction between the hits and misses in the low probability condition does not show any difference, while in the high probability condition there is a phase difference. Indeed, the phase distances of the hits and misses were significantly different between the low and high probability condition ($F(2,11) = 4.793$, $p = 0.032$). Additionally, the two individual Zar's Hotelling tests showed that for the low probability condition there was no difference between mean phases

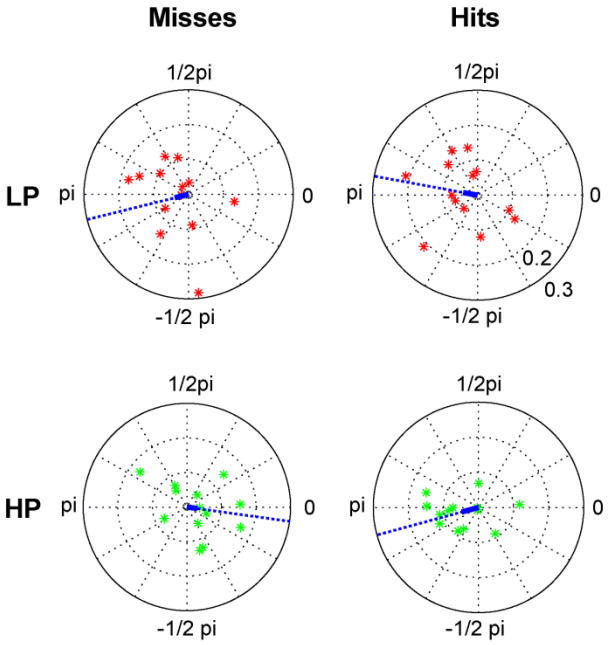


Figure 5. Delta phase effects. A) Circular histogram plots for all trials with a sound, sorted for hits and misses for both conditions (LP = low probability, HP = high probability). Blue arrows indicate the mean direction. For the LP the mean direction is the same for both hits and misses, while for the HP it differs.

($F(2,11) = 0.139$, $p = 0.872$), while for the high probability condition there was a significant difference between the mean phase of the hits and the misses ($F(2,11) = 5.39$, $p = 0.023$).

Discussion

Attention is thought to optimize selective information processing by orchestrating the synchronization between incoming temporally predictive information and high-excitability phases of ongoing low-frequency neuronal oscillations (Buzsáki & Draguhn, 2004; VanRullen & Koch, 2003). In many situations temporal cues are not highly predictive, containing some predictability as well as uncertainties. We hypothesized that in one of such situations, that is to say when attention has to be allocated continuously but within a restricted window in time ('restricted vigilance'), a longer, but not excessively long period of high neuronal excitability should provide optimal processing. An efficient way to accomplish this is phase-resetting of low-frequency oscillations; this provides high excitability over longer periods, but does not excessively use metabolically demanding gamma oscillations (Mukamel et al., 2005; Niessing et al., 2005). To investigate whether top-down expectancy of stimulus occurrence during a restricted period of vigilance results in a stronger reset of low-frequency oscillations, we manipulated the probability that a low-intensity, temporally non-predictive, auditory target would be presented in a given time window (450ms) following a visual cue. We indeed found stronger representations of low-frequencies (1-3 Hz) in the evoked power and phase coherence for the high probability condition. Moreover, delta phase determined hits and misses in the high probability condition, but not in the low probability condition. These findings indicate that during restricted vigilance, low-frequency phase-reset increases the window of enhanced excitability, with the strength of this mechanism being amplified with greater stimulus probability.

Attention window created through low-frequency phase-reset

Our current paradigm of restricted vigilance, in which attention has to be allocated for a constrained time window, seems to bridge two attention modes that have been suggested in the literature to operate for unpredictable versus predictable inputs (see also (Niemi & Näätänen, 1981)). Specifically, in the absence of any temporal predictability, Schroeder and Lakatos (2009) proposed a 'vigilance mode', in which primarily high frequency gamma oscillations are amplified. Boosting gamma oscillations may improve stimulus detection since they produce many, densely distributed, high-excitability phases, and this increases the chance that a temporally unpredictable input will arrive during a state of high neuronal excitability (Fries, 2005; Fries, Nikolic, & Singer, 2007). This mechanism is complementary to the 'rhythmic mode' which employs low-frequency oscillations to selectively process temporally predictable stimuli. During 'rhythmic mode' processing, the amplitude of gamma oscillations is phase coupled to lower frequencies, providing limited temporal windows during which processing is enhanced. Restricted vigilance requires parsing for a constrained time window and low frequency oscillations seem a plausible candidate, as these produce a high excitability phase for a longer period of time. Since in our paradigm the restricted vigilance window is 450 ms, an oscillation of approximately 1 Hz will fit this vigilance window with half-a-cycle (i.e. the high-excitability phase of the oscillation). We indeed find effects of stimulus probability (high vs. low probability) on phase consistency to be pronounced at low frequencies (around 1 Hz), that were related to the participants' behavior. This means that when relevant information can only occur within a limited time interval, processing is specifically enhanced by aligning high excitable phases to this interval via phase-reset of low-frequency oscillations. In future studies, it will be interesting to investigate whether for longer time intervals restricted vigilance will still operate and what will happen if the vigilance window occurs not right after the expectancy cue onset, but later in the trial.

The expectancy cue modulated the low-frequency effect by changing the strength of the phase reset. A similar effect has been shown in another EEG study (Stefanics, et al., 2010) in which rhythmic stimuli provided a predictable temporal structure. The pitch of the stimuli

indicated the chance of the next stimulus being a target. Similar as in the current study the strength of the phase reset depended on the likelihood of stimulus occurrence indicated by the expectancy cue. It thus seems that a rhythmic processing mode is strengthened by expectancy cues. It is an open question whether these effects can be purely attributed to changes in expectations or are modulated by changes in attention to the time window of the target stimulus occurrence (Summerfield & Egner, 2009). While expectancies are created by changing the probability of stimulus occurrence, attention guides perception via goal-directed amplification of responses. As in the current experimental setting stimulus probabilities of task-relevant features are manipulated, the effect of attention is difficult to dissociate from expectations (Summerfield & de Lange, 2014). While increases in stimulus expectancy often lead to decreases in neuronal responses (Näätänen, Paavilainen, Rinne, & Alho, 2007), increases in attention are associated with increased responses (Maunsell & Treue, 2006). Therefore, our results seem more compatible with the latter view, but studies separately controlling for attention and expectancy have to be conducted to verify this view.

Implicit versus explicit timing

In our paradigm, it seems that the temporal structure of the task is implicitly acquired and consequently influences perception (Coull & Nobre, 2008). This is in contrast with explicit timing during which an overt estimate of temporal information has to be made, for example when participants have to discriminate between the lengths of two intervals. During a task with implicit timing participants are required to make a motor or perceptual judgment while using the knowledge of when stimuli are more likely to occur (Niemi & Näätänen, 1981; Wright & Fitzgerald, 2004). Therefore the task is not temporal. There is evidence that the neuronal substrates of the two timing mechanisms are different (Coull & Nobre, 2008). Implicit timing uses mechanisms of the brain to temporally predict arriving targets (Schubotz, 2007). One such mechanism is the use of slow ongoing oscillations to align phases at a high excitable phase at the arrival time of a target (Lakatos et al., 2008). The current results indeed show that delta phase modulates perception when

events are more likely to occur. Additionally, there seems to be a difference whether implicit rhythmic temporal cues or single temporal cues (as in the current study) are guided via different mechanisms (Wilsch, Henry, Herrmann, Maess, & Obleser, 2015; Triviño, Arnedo, Lupiáñez, Chirivella, & Correa, 2011). Our study highlights that oscillatory phase reset mechanisms seem to play a role also with single temporally predictive cues, but it is still unclear whether the exact mechanism is the same as with a predictive input stream.

Alpha versus delta effects: evoked response versus phase-reset

In addition to the low-frequency effects in the delta range, the evoked response in the time-frequency analyses showed significantly stronger alpha band oscillations when events were more likely to occur. This effect likely reflects differences in the visual evoked response caused by increased likelihood. As mentioned above, the direction of the effect depends on the mechanism in place (i.e. attention or expectancy). Both increases (Hillyard, Hink, Schwent, & Picton, 1973; Näätänen, Gaillard, & Mäntysalo, 1978; Yamagishi et al., 2003) and decreases (Näätänen, et al., 2007) in evoked responses have been reported. The evoked delta band modulation however cannot easily be explained as a difference in the visual evoked response. Firstly, the topographic distribution does not include any occipital channels. Secondly, this low frequency band generally does not emerge for simple visual evoked responses. Moreover, we could show that the mechanism behind the modulation of evoked power in the delta band is different from the alpha band evoked power, by looking at the induced power and ITC: whereas the induced power as well as the ITC showed modulations in the alpha band, only the ITC showed significant changes for the delta band. It has been shown that when power changes are absent, ITC increases can be explained via a phase-reset mechanism that aligns the phases of ongoing oscillations without changing the amplitude (Makeig, et al., 2004). Therefore, we believe that alpha oscillations drive visual evoked responses, while delta oscillations have a modulatory role, and the collective delta effects found here reflect a phase-reset of which the frequency is likely influenced by the temporal predictions (i.e. 1 Hz frequencies are used to cover the

restricted vigilance window of 450 ms with half-a-cycle). The neuronal origin of this effect could reflect a change in auditory cortex excitability at the predicted arrival time of the auditory stimulus, as has been reported before (Lakatos, et al., 2008). However, the topography could also fit with an origin in the anterior cingulate cortex. This brain structure has been related to the monitoring and guidance of attentional selection (Buckley, et al., 2009; Womelsdorf, Ardid, Everling, & Valiante, 2014) and could therefore guide the temporal attention network to attend to relevant moments in time in the current task.

Delta phase determines percept

We found that especially in the high probability condition hits and misses depended systematically on delta phase since there was a significant phase difference between hits and misses for this condition. For the low probability conditions this was not the case and therefore suggests that the delta phase modulation is primarily present when expectations are high, thus reflecting a top-down mechanism. These results add to a growing set of findings showing that delta phase is important for auditory detection (Henry & Obleser, 2012; Lakatos, et al., 2008). As has been shown in a recent study, it is vital to ensure that no post-stimulus portions of the data are including in the phase estimation (Zoefel & Heil, 2013) which can be avoided by using causal filters.

Implications for multisensory research

The use of low-frequency phase-reset in multisensory settings is becoming increasingly evident (Schroeder, et al., 2008; Van Atteveldt, Murray, Thut, & Schroeder, 2014). Thus far, influences of low-frequency oscillations on multisensory perception have been associated with cross-modal phase resetting which results in subsequent periodic increases in detection thresholds or reaction times (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Naue et al., 2011; Romei, Gross, & Thut, 2012; Thorne, De Vos, Viola, & Debener, 2011; Thorne & Debener, 2013). However, these studies have a different emphasis than classical multisensory integration studies, which generally do not have a main

focus on unisensory detection or reaction times, but on integration and unity of multisensory information (Calvert, Campbell, & Brammer, 2000). For example, temporal unity is created when two stimuli are presented in temporal proximity (Spence & Squire, 2003; Van Wassenhove, Grant, & Poeppel, 2007). If reset frequency is modulated by the width of the restricted vigilance window, as our findings seem to suggest, integration of multisensory information might be confined in a similar framework. Specifically, an ambiguity in the requirement of temporal unity is that more complex stimuli, as well as stimuli that naturally belong together, seem to tolerate more variation in their temporal relation to still be integrated, compared to simple flashes and beeps (Van Wassenhove, Grant, & Poeppel, 2005; Vatakis & Spence, 2007; Zampini, Shore, & Spence, 2003). If integration occurs only when cross-modal stimuli fall within the same oscillation period (see e.g. (Fries, 2005; VanRullen & Koch, 2003), resetting of lower frequency oscillations for stimuli that naturally occur together (i.e. they predict each other) would result in wider temporal integration windows, since there is a longer temporal window in which the two stimuli will fall in the same period. Supporting the idea that oscillation period is vital for temporal integration is the finding that when two visual stimuli are presented with a specific delay, they are only judged as synchronous when they fall within the same visual alpha cycle (Gho & Varela, 1988; Valera, Toro, Roy John, & Schwartz, 1981). In brief, interplay between the likelihood that cross-modal stimuli can occur together and reset frequency might be related to the width of the temporal integration window. However, this prediction needs to be verified in future studies.

Conclusion

Selective information processing is crucial to our survival considering the constant presence of abundant sensory information. Therefore, it seems beneficial to exploit temporal cues in our environment as this enables proactive mechanisms for selective processing. When events are highly predictable, low-frequency phase-reset appears to guide selective processing (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Lakatos, et al., 2008; Schroeder & Lakatos, 2009). Here, we show that also in the absence

of a fully predictable temporal structure, low-frequency phase-reset is employed to attend to a time window in which events are more likely (also see (Fiebelkorn, et al., 2011)), revealing the full flexibility of this neural mechanism supporting selective processing. These results shed light on the adaptive nature of phase-reset to optimally sample the incoming information depending on top-down expectancies of stimulus occurrence and timing (Ten Oever, Schroeder, Poeppel, Van Atteveldt, & Zion Golumbic, 2014; Van Atteveldt, et al., 2014; Zion Golumbic, Poeppel, & Schroeder, 2012). Future research should confirm that phase-reset frequency is flexibly used to modify the temporal attention window, which could subsequently inform us about the functioning of other cognitive mechanisms, for example the variable temporal integration windows for multisensory inputs, or the flexible use of different time-scales during verbal communication.

Acknowledgments: We thank Kirsten Petras and Marie Marinelli for all their work during data collection. This work was supported by a grant from the Dutch Organization for Scientific Research (NWO; grant number 406-11-068).

References

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software, 31*(10), 1-21.

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science, 325*(5936), 52-58.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*(5679), 1926-1929.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649-657.

Coull, J., & Nobre, A. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology, 18*(2), 137-144.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *The Journal of Neuroscience, 31*(27), 9971-9981.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Fries, P., Nikolic, D., & Singer, W. (2007). The gamma cycle. *Trends in Neurosciences, 30(7*), 309-316.

Gho, M., & Varela, F. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *Journal de physiologie, 83*(2), 95.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1974): Wiley New York.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, 109*(49), 20095-20100.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(108), 177-180.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. *Current Biology, 20*(1), 19-24.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology, 94*(3), 1904-1911.

Los, S. A., Knol, D. L., & Boers, R. M. (2001). The foreperiod effect revisited: Conditioning as a basis for nonspecific preparation. *Acta psychologica, 106*(1), 121-145.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences, 8*(5), 204-210.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods, 164*(1), 177-190.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317-322.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science, 309*(5736), 951-954.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica, 42*(4), 313-329.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*(12), 2544-2590.

Naue, N., Rach, S., Strüber, D., Huster, R. J., Zaehle, T., Körner, U., et al. (2011). Auditory Event-Related Response in Visual Cortex Modulates Subsequent Visual Responses in Humans. *The Journal of Neuroscience, 31*(21), 7729-7736.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*(1), 133.

Niessing, J., Ebisch, B., Schmidt, K. E., Niessing, M., Singer, W., & Galuske, R. A. W. (2005). Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science, 309*(5736), 948-951.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*, 1.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience, 32*(24), 8424-8428.

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current Biology, 22*(9), 807-813.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Schubotz, R.I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in Cognitive Sciences, 11*, 211-218

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521.

Stefanics, G., Hangya, B., HernÃ¡di, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of Neuroscience, 30*(41), 13578-13585.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 34*(25), 8519-8528.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience, 16*(13), 4240-4249.

Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N., & Zion Golumbic, E. M. (2014). The influence of temporal regularities and cross-modal temporal cues on auditory detection. *Neuropsychologia, 63,*43-50.

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience, 31*(10), 3853-3861.

Thorne, J. D., & Debener, S. (2013). Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hearing research, 307*, 144-152.

Triviño, M., Arnedo, M., Lupiáñez, J., Chirivella, J., & Correa, Á. (2011). Rhythms can overcome temporal orienting deficit after right frontal damage. *Neuropsychologia, 49*(14), 3917-3930.

Valera, F. J., Toro, A., Roy John, E., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia, 19*(5), 675-686.

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6), 1240-1253.

van den Brink, R. L., Wynn, S. C., & Nieuwenhuis, S. (2014). Post-Error Slowing as a Consequence of Disturbed Low-Frequency Oscillatory Phase Entrainment. *The Journal of Neuroscience, 34*(33), 11096-11105.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7*(5), 207-213.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Attention, Perception, & Psychophysics, 69*(5), 744-756.

Womelsdorf, T., Ardid, S., Everling, S., & Valiante, T. A. (2014). Burst Firing Synchronizes Prefrontal and Anterior Cingulate Cortex during Attentional Control. *Current Biology, 24*(22), 2613-2621.

Wilsch, A., Henry, M.J., Herrmann, B., Maess, B., & Obleser, J. (2015). Slow-delta phase concentration marks improved temporal expectations based on the passage of time. *Psychophysiology, 52*(7), 910-918.

Wright, B. A., & Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Perception & psychophysics, 66*(3), 508-516.

Yamagishi, N., Callan, D. E., Goda, N., Anderson, S. J., Yoshida, Y., & Kawato, M. (2003). Attentional modulation of oscillatory activity in human visual cortex. *Neuroimage, 20*(1), 98-113.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198-210.

Zar, J. H. (1998). *Biostatistical Analysis* (4 ed.). Englewood Cliffs, New Jersey: Prentice Hall.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*(6), 1414-1425.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*(3), 151-161.

Zoefel, B., & Heil, P. (2013). Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Frontiers in Psychology, 4*.

CHAPTER 6

AUDIOVISUAL ONSET DIFFERENCES ARE USED TO DETERMINE SYLLABLE IDENTITY FOR AMBIGUOUS AUDIOVISUAL STIMULUS PAIRS

Corresponding Manuscript:
Ten Oever, S., Sack, A., Wheat, K. L., Bien, N., & Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology, 4*.

Abstract

Content and temporal cues have been shown to interact during audiovisual (AV) speech identification. Typically, the most reliable unimodal cue is used more strongly to identify specific speech features; however, visual cues are only used if the audiovisual stimuli are presented within a certain temporal integration window (TWI). This suggests that temporal cues denote whether unimodal stimuli belong together, that is, whether they should be integrated. It is not known whether temporal cues also provide information about the identity of a syllable. Since spoken syllables have naturally varying audiovisual onset asynchronies, we hypothesize that for suboptimal AV cues presented within the TWI, information about the natural AV onset differences can aid in speech identification. To test this, we presented low-intensity auditory syllables concurrently with visual speech signals, and varied the stimulus onset asynchronies (SOA) of the audiovisual pair, while participants were instructed to identify the auditory syllables. We revealed that specific speech features (e.g. voicing) were identified by relying primarily on one modality (e.g. auditory) Additionally, we showed a wide window in which visual information influenced auditory perception, that seemed even wider for congruent stimulus pairs. Finally, we found a specific response pattern across the SOA range for syllables that were not reliably identified by the unimodal cues, which we explained as the result of the use of natural onset differences between audiovisual speech signals. This indicates that temporal cues not only provide information about the temporal integration of audiovisual stimuli, but additionally convey information about the identity of audiovisual pairs. These results provide a detailed behavioral basis for further neuro-imaging and stimulation studies to unravel the neurofunctional mechanisms of the audio-visual-temporal interplay within speech perception.

Introduction

Although audition is our main informant during speech perception, visual cues have been shown to strongly influence identification and recognition of speech (Campbell, 2008). Visual cues are used to increase understanding, especially in noisy situations when auditory information alone is not sufficient (Bernstein, Auer, & Takayanagi, 2004; Grant, Wassenhove, & Poeppel, 2004; Sumby & Pollack, 1954). It is known that temporal, spatial and semantic cues in visual signals are used to improve auditory speech perception (Stevenson & James, 2009; Wallace, Wilkinson, & Stein, 1996). However, it is largely unknown how these different cues are combined to create our auditory percept. In the current research, we used semantically congruent or incongruent audiovisual syllables presented with varied stimulus onset asynchronies (SOAs) between the auditory and visual stimuli, to investigate the interaction between temporal and content factors during audiovisual speech perception (see e.g. Van Wassenhove, Grant, & Poeppel, 2007; Vatakis, Maragos, Rodomagoulakis, & Spence, 2012; Vatakis & Spence, 2006). Specifically, we were interested whether natural onset asynchronies inherent to audiovisual syllable pairs influence syllable identification.

Often, stop-consonant syllables (e.g. /ba/ and /da/) are used to examine syllable identification (see e.g. Arnal, Wyart, & Giraud, 2011; McGurk & MacDonald, 1976; Van Wassenhove, et al., 2007). Stop-consonants are consistent in the manner in which they are produced (the vocal tract is blocked to cease airflow), but vary in the type and amount of identity information conveyed by the visual and auditory channels. Specifically, whether or not the vocal tract is used to produce a consonant (i.e. the voicing of a sound, /ba/ vs. /pa/) is not visible, since the vocal tract is located in the throat. Therefore, the auditory signal is more reliable than the visual signal in determining the voicing of a speech signal (McGurk & MacDonald, 1976; Wiener & Miller, 1946). On the other hand, which part of the mouth we use for producing a syllable is mostly a visual signal. For example, uttering a syllable with our lips (like /ba/) versus our tongue (like da/) is more visible than audible. Visual speech thus conveys mostly information about the place of articulation (POA) of the sound, and adding acoustic noise to a spoken syllable makes the place of articulation (POA) particularly difficult to extract on basis of

auditory information (McGurk & MacDonald, 1976; Van Wassenhove, Grant, & Poeppel, 2005; Wiener & Miller, 1946). However, the amount of visual information about the POA varies for different syllables: bilabial syllables (pronounced with the lips) are better dissociated than coronal and dorsal syllables (pronounced with the front or body of the tongue, respectively). Thus, it seems that auditory and visual speech signals are complementary in identifying a syllable, since voicing information is best conveyed by auditory cues and POA information by visual cues (Campbell, 2008; Summerfield, 1987).

Auditory and visual stimuli can be linked based on their content information; the information about the identity (the 'what') of a stimulus. We will continue to use the term content information, although in other studies the term semantic information is also used (for a review see (Doehrmann & Naumer, 2008)). The amount of content information conveyed by a unimodal signal is variable, for different stimuli (as explained above) as well as for individuals perceiving the same stimuli, and the reliability of the information determines how strongly it influences our percept (Beauchamp, Lee, Argall, & Martin, 2004; Blau, Van Atteveldt, Formisano, Goebel, & Blomert, 2008; Driver, 1996; Van Wassenhove, et al., 2005). For example, the amount of content information present in visual speech signals is widely variable, as reflected in individual differences in lipreading skills (Auer Jr & Bernstein, 1997; MacLeod & Summerfield, 1987), and it has been shown that more profound lipreaders also use this information more (Auer Jr & Bernstein, 2007; Pandey, Kunov, & Abel, 1986). Additionally, visual speech signals that convey more content information (like bilabial vs. dorsal syllables, as explained above) bias the speech percept more strongly (McGurk & MacDonald, 1976; Van Wassenhove, et al., 2005). However, the influence of visual information on auditory perception often depends not only on the nature and quality of the visual signal, but also on the quality of the auditory signal, since visual input is especially useful for sound identification when background noise levels are high (Grant, et al., 2004; Sumby & Pollack, 1954). Thus, during audiovisual identification unimodal cues seem to be weighted based on their reliability, to create the audiovisual percept (Massaro, 1987, 1997). Additionally, the amount of weight allocated to each modality depends not only on the overall quality of the signal, but also on the reliability of the signal for the

specific feature that needs to be identified. For example, spatial perception is more accurate in the visual domain, therefore spatial localization of audiovisual stimuli mostly dependents on visual signals (Driver, 1996). One of the aims of our study was to provide further support for the notion that reliable modalities are weighted more heavily (Beauchamp, et al., 2004; Massaro, 1997). Specifically, we investigated whether systematic difference in the reliability of the voicing and POA features of the syllable (see above) biases which modality is weighted more heavily.

The main aim of our study was to investigate how the temporal relation between audiovisual pairs influences our percept. It is known that auditory and visual signals are only integrated when they are presented within a certain temporal window (Ernst & Bülthoff, 2004; Massaro, Cohen, & Smeele, 1996; R. Welch & Warren, 1986), this is the so-called temporal window of integration (TWI). The TWI is for example measurable with synchrony judgments, in which temporal synchrony of audiovisual signals is only perceived if audiovisual pairs are presented within a certain range of onset asynchronies (Meredith, Nemitz, & Stein, 1987; Spence & Squire, 2003). The TWI highlights that the temporal relationship of auditory and visual inputs is another important determinant for integration, in addition to information about the 'what' of a stimulus. The importance of this window has been replicated many times for perceptual as well as neuronal integration (Stein & Meredith, 1993; Van Atteveldt, Formisano, Blomert, & Goebel, 2007; Van Wassenhove, et al., 2007). Typical for the TWI is that the point of maximal integration occurs with visual stimuli leading (Zampini, Shore, & Spence, 2003). This seems to relate to the temporal information visual signals provide, namely a prediction of the 'when' of the auditory signal, since they naturally precede the sounds (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013). However, the difference between the onset of the visual and auditory signal varies across syllables (Chandrasekaran et al., 2009) and it is not known whether these natural onset differences can cue the identity of the speech sound. It has been shown that monkey auditory cortex and superior temporal cortex are sensitive to natural audiovisual onset differences in monkey vocals (Chandrasekaran & Ghazanfar, 2009; Ghazanfar, Maier, Hoffman, & Logothetis, 2005). In humans, it has been

shown that onset differences within the auditory modality are used to identify auditory syllables (Miller, 1977; Munhall & Vatikiotis-Bateson, 1998). For example, the distinction between a voiced or unvoiced syllable in the auditory signal is solely based on onset differences of specific frequency bands. However, it is not known whether audiovisual onset information is used to identify speech sounds. We hypothesize that inherent onset differences between auditory and visual articulatory cues can be used to identify spoken syllables. Specifically, we hypothesize that coronal (e.g. /da/) and dorsal (e.g. /ga/) stimuli (pronounced with the front or body of the tongue, respectively) might have audiovisual onset difference, in which dorsal stimuli produce longer onset differences due to a longer distance from the place of articulation to the external, audible sound.

Traditionally, only a single dimension in the auditory or visual signal is altered to investigate the influence of visual cues. However, more and more studies are showing interactions between different crossmodal cues. For example, Vatakis and Spence (2007) found that if the gender of a speaker is incongruent for auditory and visual speech, less temporal discrepancy is allowed for the stimuli to be perceived as synchronous. Stimuli in the McGurk effect (McGurk & MacDonald, 1976), in which an auditory [ba], presented with an incongruent visual /ga/ is perceived as a /da/, are also perceived as synchronous for a narrower temporal window, compared to congruent audiovisual syllables (Van Wassenhove, et al., 2007). Furthermore, in recent work we showed that auditory detection thresholds are lower if temporal predictive cues are available in both the auditory and visual domain (Ten Oever et al., submitted). In addition, interactions between semantic relatedness and spatial processing have been reported (Bien, Ten Oever, Goebel, & Sack, 2012; Driver, 1996; Parise & Spence, 2009), as well as interactions between temporal and spatial factors (Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012). However, it is still unknown how interactions between auditory and visual content as well as temporal cues influence speech identification.

In sum, for stop consonants, auditory cues provide content information with regard to voicing, whereas visual cues provide content information with regard to POA (with varying reliability, e.g. for bilabial vs. dorsal/coronal). Therefore, we were able to make use of these properties in order to investigate whether incongruent pairs of stimuli are

identified depending on the modality that has the most reliable information for the specific features; POA and voicing. Additionally, we used different SOAs to investigate the temporal profile of this effect. Specifically, we were interested in the temporal window in which visual information influences the auditory percept, and whether ambiguity in the identity of auditory syllables can be resolved using differences in natural audiovisual onsets in speech.

Materials and Methods

Participants

Eight healthy native Dutch volunteers (3 male, mean age 20.9, standard deviation 2.6) participated in the study. All participants reported to have normal hearing and normal or corrected to normal vision. Participants were unaware of the goal of the study before they completed the experiment. Informed consent was given before participating. Ethical approval was given by the Ethical Committee of the Faculty of Psychology at the University of Maastricht. Participants received €40 or student participation credits in compensation for their time.

Stimulus material

Six Dutch syllables, pronounced by a native Dutch female speaker, were used as auditory and visual stimuli (/pa/, /ba/, /ta/, /da/, /ka/, /ga/). For variability, we recorded three different versions of every syllable. Sounds were digitized at 44.1 kHz, with 16-bit amplitude resolution and were equalized for maximal intensity. Videos had a digitization rate of 30 frames per second and were 300*300 pixels. We used a method similar to method used in Van Wassenhove et al. (2005) to create the videos. Videos lasted 2367 ms, including a fade-in of a still face (8 frames), the still face (5 frames), the mouth movements (52 frames) and a fade out of a still face (5 frames). MATLAB (Mathworks) scripts were used to create these videos. Additionally, for every stimulus there was a still face video with the fade out and fade in frames. First, we tested three participants with

stimulus onset asynchronies (SOAs) between auditory and visual stimuli ranging from VA (visual lead) 300 ms up to AV (auditory lead) 300 in steps of 30 ms, since this range covers the temporal window of integration for syllables used before (see e.g. Van Wassenhove, et al., 2007; Vatakis & Spence, 2007). However, for the extreme VA and AV SOAs participants still seemed to use the visual information to determine their responses, therefore we chose to widen the SOA range (ranging from VA 540 to AV 540 ms in steps of 60 ms for the other participants). To align the incongruent auditory stimuli with the videos, the maximal intensity of the incongruent auditory stimulus was aligned with the congruent auditory stimulus.

Procedure

Each participant was tested in two separate experimental sessions, both lasting two hours. In the first session a staircase, a unimodal visual experiment, and the first part of the audiovisual experiment was conducted. The second session consisted of the remainder of the audiovisual experiment.

The staircase procedure consisted of a six-alternatives forced choice procedure in which participants were asked to identify the six different syllables without presentation of the videos. Syllables were randomly presented over a background of white noise. Depending on the accuracy of the response, the intensity of the white noise was increased or decreased for the next trial. A two-up, one-down procedure (Levitt, 1971) with a total of 20 reversals was employed, which equals approximately 70% identification threshold. The individually obtained white noise intensity was used in the following experiments as background noise for the individual participants.

In the unimodal visual experiment participants were requested to recognize the identity of the syllable based on the videos only. White noise was presented as background noise. First, a fixation cross was presented for 800 ms, followed by a syllable video. Finally, a question mark was presented with the six possible response options to which participants were requested to respond. After participants responded

there was a 200 ms break before the next trial started. In total, 360 stimuli were presented, 60 per syllable in four separate blocks.

The audiovisual experiment had a similar trial configuration to the unimodal visual experiment, but consisted of the presentation of audiovisual pairs. Only two visual stimuli were used here; /pa/ and /ga/. These specific syllables were selected because they differ from each other in terms of place of articulation: /pa/ is a bilabial syllable, pronounced in the front of the mouth, whereas /ga/ is coronal syllable, pronounced in the back of the mouth. Furthermore, it has been shown that identifying /pa/ is much easier than /ga/ (McGurk & MacDonald, 1976; Van Wassenhove, et al., 2005; Wiener & Miller, 1946), thus serving our aim to manipulate the amount of information provided by the visual stimulus. Participants were instructed to identify the auditory stimulus only (again choosing between the six possible response options), while ignoring the identity of the visual stimulus.

In total, 30 blocks were presented, distributed across the two sessions for all participants. Furthermore, per SOA there were 10 stimuli for every audiovisual combination for the five participants who saw the full range of SOAs, and 11 stimuli per SOA for the other three participants. Blocks lasted approximately 7 minutes each. Additionally, there were catch trials in which a visual or auditory unimodal stimulus (20 stimuli for each) was presented. During the auditory unimodal presentation randomly one of the still visual faces, which were also used during the fade-in of the moving faces, was presented. During the visual unimodal presentation white noise was presented at the same intensity as the audiovisual trials and participants had to indicate the identity of the visual stimulus. This ensured that participants were actually looking at the screen.

Participants were seated approximately 57 centimetres from the screen and were instructed to look at the fixation cross at all times if presented. Presentation software (Neurobehavioral Systems, Inc., Albany, NYs) was used for stimulus presentation. Visual stimuli were presented on a grey background (RGB: 100,100,100). After each block participants were encouraged to take a break and it was ensured that participants never engaged continuously in the task for more than half an hour.

Data analysis

With regard to the unimodal stimuli, we aimed to replicate previous findings stating that voicing is discriminated better in the auditory modality, whereas place of articulation (POA) is discriminated better in the visual modality (McGurk & MacDonald, 1976; Summerfield, 1987; Wiener & Miller, 1946). For the analysis concerning voicing, the percentage of voiced responses was calculated per voicing category. Thereafter, we averaged the response proportions and performed an arcsine-square-root transformation to overcome nonnormality caused by the restricted range of the proportion data (however in the figures proportions are kept for illustration purposes, since they are more intuitive). The calculated transformed response proportions per category were used as dependent variables in two repeated measurements ANOVAs, for the visual as well as for the auditory modality. For the visual unimodal analyses, the data from the unimodal visual experiment was used (although the data from the visual catch trials in the AV experiment gave comparable results), whereas for the auditory analyses the catch trials in the audiovisual experiment were analysed. To investigate whether participants could identify the voicing of the stimulus the factors Voicing of the stimulus (voiced versus unvoiced stimuli) and Voicing of the response were used. A similar analysis was performed to investigate whether POA could be identified in the auditory and visual modality. Here, the percentage of POA responses per POA category were calculated, arcsine-squared-root transformed, and the factors Place of Articulation of the stimulus (bilabial, coronal or dorsal) and Place of Articulation of the response were used in two repeated measurements ANOVAs for the visual and auditory modality. For significant interactions simple effect analyses per stimulus category were performed. If not otherwise reported, all multiple comparisons were Bonferroni corrected and effects of repeated measures were corrected for sphericity issues by Greenhouse-Geisser correcting the degrees of freedom.

For the AudioVisual analyses, we first performed the same analyses as for the unimodal stimuli, collapsed over the SOAs, separately for visual /pa/ and /ga/. Thereafter, linear mixed models were used to investigate the SOA effects. This approach was chosen to accommodate for the

missing data which arose because three participants were only presented with SOAs between VA 300 and AV 300 ms instead of VA540 to AV 540 ms. Per visual stimulus and per voicing level a mixed model was run with the factors Stimulus POA, Response (only responses that were on average per VC category above chance level were used for further analyses) and SOA. This factor was created by binning the differently used SOAs in nine bins with centre points: VA 50, 125, 275, and 475, 0, and AV 50, 125, 275, and 475. These bins were chosen to include all the SOAs used. Additionally, a random intercept was added to account for the individual variations in the baseline.

We hypothesized differential effects as a result of natural differences in onset asynchronies of mouth movements and congruent speech sounds, for example between dorsal (earlier movements) and coronal syllables (later movements). In order to investigate this hypothesis, we calculated the velocity of the mouth movements as follows. For each visual stimulus we zoomed in on the area around the mouth (see figure 1). Then, the mean of the absolute differences of the three RGB values per pixel for adjacent frames was calculated. Thereafter, to quantify the movement from one frame to the other, the variance of the mean absolute RGB differences over the pixels was calculated and this was repeated for all the frames. This resulted in a velocity envelope of the mouth movement (i.e., comparable to the derivative of the mouth movement – it indicates changes in the movement) in which a clear opening and closing of the mouth becomes visible (see figure 1). The result of this method is similar to the methods used by Chandrasekaran and colleagues (2009), such that the point of maximum velocity coincides

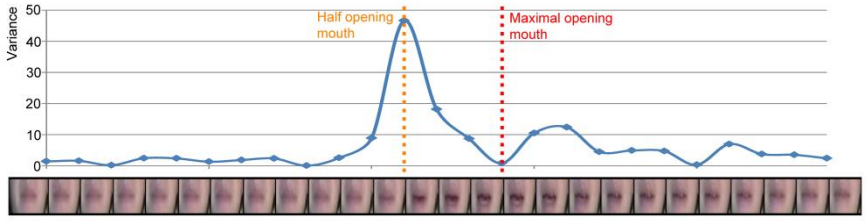


Figure 1. Example of the envelope of the velocity of the mouth movement of visual /pa/. Each dot represents the variance over all pixels of the mean RGB difference for two adjacent frames (the frame left and right of the dot). The orange dotted line represents the half opening of the mouth and the red dotted line represents the maximal opening of the mouth.

with a half open mouth and the minimum velocity coincides with a fully open mouth. To quantify the onset differences between the auditory and visual signals, the time point of maximal amplitude of the auditory signal was subtracted from the time point of maximal velocity of the visual signal. These values were later used in a linear mixed model (see Results section for details).

Results

Unimodal effects

We replicated previous results showing that voicing is most optimally discriminated in auditory syllables and POA most optimally in visual syllables (see figure 2 and table 1 and 3). Table 1 indicates that the response POA interacts with the stimulus POA only for the visual stimuli, which means that for a stimulus with a specific POA the POA categories have different response proportions during the visual experiment. Simple effects show that especially bilabial stimuli were identified correctly during the visual experiment (as indicated by significantly higher bilabial than dorsal and coronal responses). Dorsal and coronal visual stimuli were more often confused with each other. However, for the unimodal auditory stimuli, the interaction between response and stimulus POA did not reach significance, indicating that participants were not able to dissociate the POA of the auditory stimuli. Table 3 (top rows) shows significant simple effects of the voicing of the response per stimulus level for the auditory, but not the visual modality. This means that in the auditory modality, voicing was primarily categorized correctly.

Table 1. Results for the POA analyses of the unimodal stimuli

		POA Inter-action	Simple effects per stimulus level								
			Stimulus Bilabial (B)			Stimulus Coronal (C)			Stimulus Dorsal (D)		
			BvsC	BvsD	CvsD	BvsC	BvsD	CvsD	BvsC	BvsD	CvsD
Auditory	F/t	2.34	-	-	-	-	-	-	-	-	-
	p	0.12									
Visual	F/t	*178.4*	*23.2*	*26.8*	-0.92	*-9.89*	*-8.24*	2.70	*-9.6*	*-13.1*	-0.16
	p	*0.00***	*0.00***	*0.00***	1.00	*0.00***	*0.00***	0.09	*0.00***	*0.00***	1.00

The second column shows the interaction between stimulus and response place of articulation (POA interaction), and the other three columns show for stimuli with the different POAs the pairwise comparisons of the response proportions between the different POAs responses (B is bilabial, C is coronal, and D is dorsal). Auditory and visual rows indicate the results from the auditory only trials during the audiovisual experiment and the separate unimodal visual experiment, respectively. Results for post-hoc analyses are only shown if ANOVA tests are significant. Single and double asterisks indicate p-values below 0.05 and 0.01, respectively.

Multimodal effects collapsed over SOAs

During the audiovisual experiment, the voicing of the stimuli was identified correctly most of the time (as indicated by significant simple effects for the voicing analyses; see figure 3 and table 3), and resembles the results from the unimodal auditory analyses. The results for the POA, when visual /pa/ was presented, resulted in high response proportions (more than 0.8) for bilabial stimuli (see table 2), paralleling visual unimodal results. The POA response*stimulus interaction effect indicates that bilabial responses are specifically reported when the auditory stimuli is also bilabial, but in the simple effects the comparisons did not show significant differences (table 2, row 3). Similarly, the response distributions for dorsal stimuli in the unimodal visual experiment and the visual /ga/ during the audiovisual experiment seem to resemble each other, that is, in the audiovisual experiment participants also confused the coronal and dorsal POA.

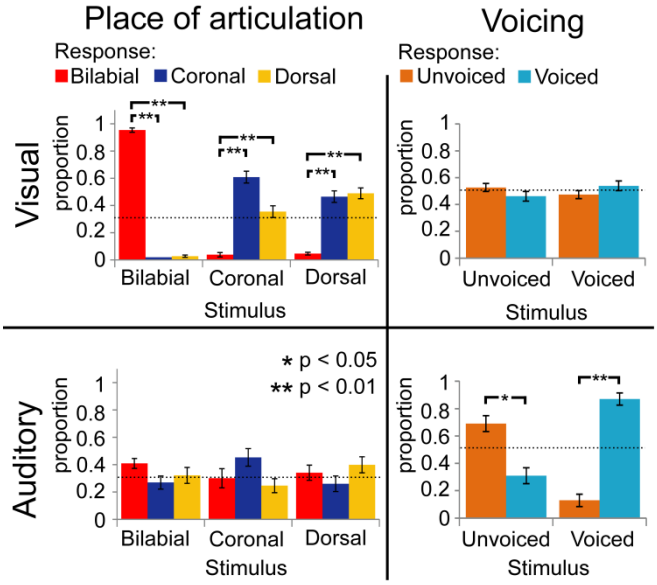


Figure 2. Results of unimodal analyses for auditory and visual signals separately. Horizontal axis represents the category of the stimulus and vertical axis represents the response proportions of the respective categories. Dashed lines indicate chance level performance. As shown, vision can dissociate place of articulation (POA) and audition can dissociate voicing (VC).

Table 2. Results for the POA analyses of the multimodal stimuli.

		POA Interaction	Simple effects for congruent response			POA Response; main effect	Pairwise Comparison of response level		
			BvsC	BvsD	CvsD		BvsC	BvsD	CvsD
AV	F/t	*6.30*	2.41	2.23	-1.89	*92.2*	*8.33*	*10.6*	1.15
visual /pa/	p	*0.02**	0.14	0.19	0.29	*0.00***	*0.00***	*0.00***	1.00
AV,	F/t	3.43	-	-	-	*39.78*	*-4.80*	*-7.94*	0.03
visual /ga/	p	0.07				*0.00*	*0.01***	*0.00***	1.00

The second column is similar as in table 1. The third column shows the simple effect for the visual congruent response option (for visual /pa/ the bilabial response), comparing whether for specific stimuli the congruent visual POA option has a higher proportion. The fourth column shows the main effect of the response of the POA. The last column shows the pairwise comparisons whether overall, one POA response is given more often than another (B is bilabial, C is coronal, and D is dorsal). Results for post-hoc analyses are only shown if ANOVA tests are significant. Single and double asterisks indicate p-values below 0.05 and 0.01, respectively.

Table 3. Results for Voicing for both unimodal and multimodal stimuli

		Voicing interaction	Response simple effects per stimulus level: Voiced vs Unvoiced	
			Stimulus Voiced	Stimulus Unvoiced
Auditory	F/t	*43.8*	*8.19*	*-2.83*
	p	*0.00***	*0.00***	*0.03**
Visual	F/t	*18.5*	1.66	-0.13
	p	*0.00**	0.14	0.90
AV	F/t	*112*	*8.71*	*-6.82*
Visual /pa/	p	*0.00***	*0.00***	*0.00***
AV	F/t	*87.2*	*11.42*	*-3.94*
Visual /ga/	p	*0.00***	*0.00***	*0.01***

The second column is the interaction of stimulus Voicing with response Voicing (Voicing interaction). The third and fourth columns are the simple effect analyses of the voicing of the response per stimulus level.Results for post-hoc analyses are only shown if ANOVA tests are significant. Single and double asterisks indicate p-values below 0.05 and 0.01, respectively.

The latter analysis shows that adding a visual stimulus changes the auditory percept for the different POA categories, such that with incongruent audiovisual POA, the correct POA response choice (i.e. the POA of the auditory stimulus) is nearly absent in the chosen responses. For example, although a dorsal auditory stimulus is presented (e.g. /ka/), if concurrently visual /pa/ is presented, the response options with dorsal POAs are only chosen approximately 10% of the times (see figure 3 and 4). Therefore, we decided that, for the analyses including the temporal factors, we would only use the response options that were given more than chance level per stimulus voicing and POA (POA: 0.33, voicing: 0.5). Mainly, because we were interested in the temporal pattern of the identification and a very low response rate could result in floor effects, biasing the statistical analyses. Thus for visual /pa/, auditory-unvoiced we only used response /pa/ (see figure 3; stimulus unvoiced and POA bilabial) and for visual /pa/, auditory-voiced we only used response /ba/

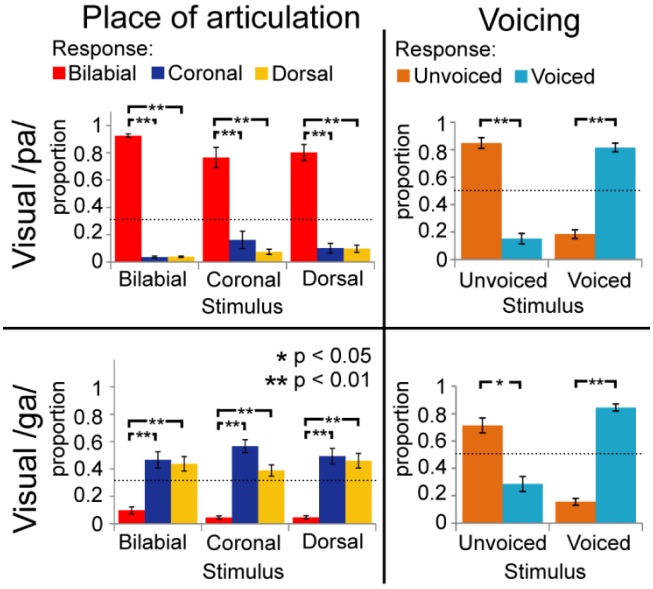


Figure 3. Results of multimodal analyses for visual /pa/ and /ga/ separately collapsed over stimulus onset asynchronies (SOAs). Horizontal axis represents the category of the stimulus and vertical axis represents the response proportions of the respective categories. Dashed lines indicate chance level of responding. Voicing (VC) is dissociable, but place of articulation (POA) responses depended on the unimodal visual response in Figure 2.

(stimulus voiced and POA bilabial). For visual /ga/, auditory-unvoiced response options /ta/ and /ka/ were used (stimulus unvoiced and POA coronal and dorsal respectively) and for visual /ga/, auditory-voiced response options /da/ and /ga/ were used (stimulus voiced and POA coronal and dorsal respectively).

Temporal effects during visual /pa/

Overall effects of SOA difference are shown in figure 4. The mixed model analyses for visual /pa/, auditory-unvoiced showed an main effect for POA and SOA (figure 5A; $F(2,180) = 34.04$, $p < 0.001$ and $F(8,180) = 10.88$, $p < 0.001$, respectively). Bilabial responses were reported significantly more than coronal and dorsal responses ($t(180) = 7.60$, $p < 0.004$ and $t(180) = 6.59$, $p < 0.001$, respectively). The main effect of SOA indicated that compared to an SOA of zero, for AV 475 and AV 275 lower /pa/ response proportion were given ($t(180) = -4.60$, $p< 0.001$ and $t(180) = -4.583$, $p < 0.001$, respectively). Thus, the proportion /pa/ responses were the least for incongruent bilabial presentation, and when auditory stimuli were leading more than 125 ms. Visual /pa/, auditory-voiced stimuli resulted in similar results: an main effect for POA and SOA (figure 5B; $F(2,180) = 13.59$, $p < 0.001$ and $F(8,180) = 4.83$, $p < 0.001$, respectively). Bilabial response proportions were higher than coronal and dorsal response proportions ($t(180) = -4.49$, $p < 0.001$ and $t(180) = -4.54$, $p < 0.001$, respectively). Here, for a smaller window /ba/ responses were given compared to visual /pa/ - unvoiced /pa/ responses, that is, the SOAs of AV 475, AV 275 and VA 475 were significantly different from an SOA of zero (AV 475: $t(180) = -4.027$, $p < 0.001$; AV 275: $t(180) = -3.639$, $p = 0.003$; and VA475: $t(180) = -3.584$, $p = 0.004$).

Temporal effects during visual /ga/

The multilevel analyses for the visual /ga/ unvoiced showed an interaction effect between response and SOA ($F(8,371) = 4.540$, $p < 0.001$). Results from the simple effects analyses in which the /ta/ and /ka/ responses per SOA level were compared indicated that for SOA VA 275 /ka/ was indicated more and for SOA AV 50,125, and 475 /ta/ was

indicated more (uncorrected values: -275 = -2.813, p = 0.008; 50: t(24) = 2.088, p = 0.041; 125: t(24) = 2.394, p = 0.022;. 475: t(24) =2.650, p = 0.014), but these effects did not survive correction for multiple comparisons. The interaction effect however, seems to be caused by more answered /ka/ with negative SOAs, and more answered /ta/ with positive SOAs (see figure 6A). For the visual /ga/, auditory-voiced the multilevel

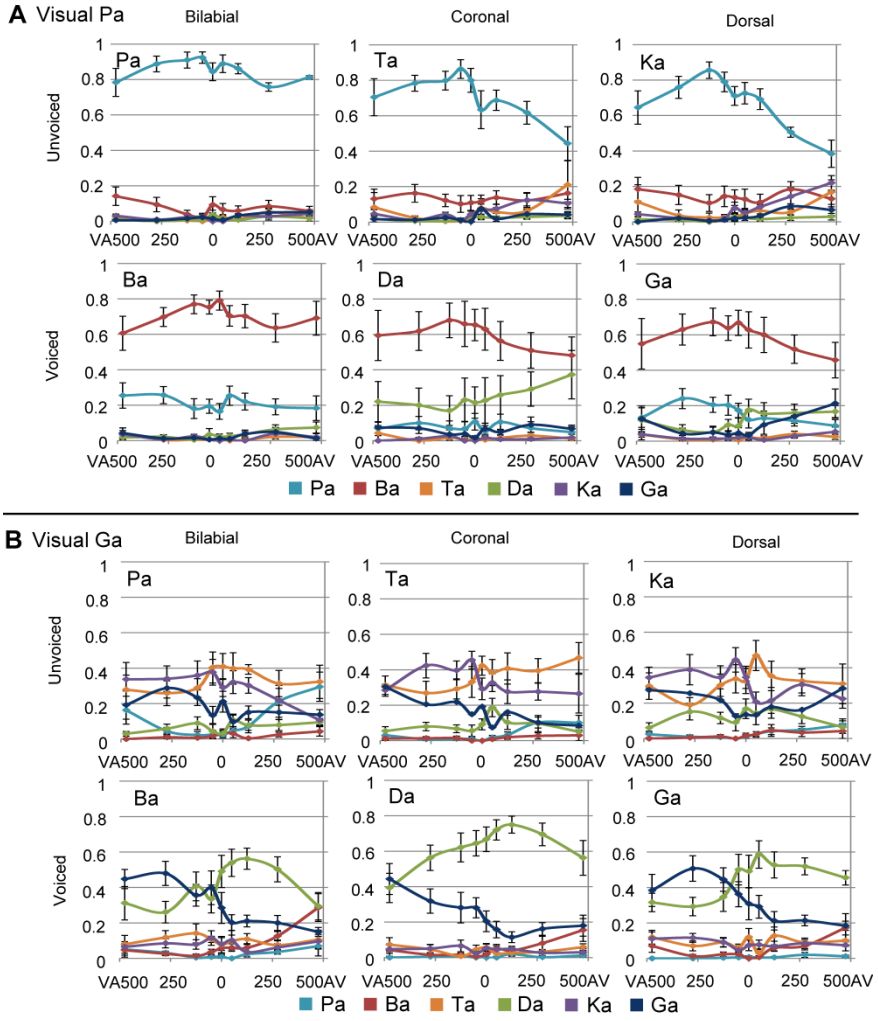


Figure 4: Overall results of the multimodal experiment for the different visual stimuli, auditory stimuli and stimulus onset asynchronies (SOAs) for visual /pa/ (A) and visual /ga/ (B). Negative SOAs indicate that the visual stimulus was shifted to an earlier point in time compared to the auditory stimulus.

analyses also showed an interaction of response and SOA (see figure 6B; $F(8,367) = 11.996$, $p < 0.001$). Additionally, it showed an interaction between stimulus POA and response ($F(8,367) = 26.480$, $p < 0.001$). One explanation for this last effect could be that our [da] stimulus was better identifiable unimodally than the other auditory stimuli (see figure 4), such that for stimulus POA coronal a higher proportion /da/ responses were given (since this was the right answer). This was similar during visual /pa/, auditory [da], which also showed a higher proportion /da/ compared to the correct responses during other incongruent combinations (figure 4A). For the response * SOA interaction we performed simple effects analyses per SOA level. For all AV SOAs and SOA 0 /da/ was reported significantly more than /ga/ (475: $t(24) = 4.667$, $p < 0.001$; 275: $t(24) = 7.624$, $p < 0.001$; 125: $t(24) = 9.089$, $p < 0.001$; 50: $t(24) = 6.615$, $p < 0.0001$; 0: $t(24) = 3.922$, $p = 0.004$).

'Crossing' identification for visual /ga/

Around the zero point, we observed a quick incline or decline in the response choice of participants for visual /ga/ (see figure 4B and 6), such that participants chose with positive SOAs more often coronal responses (/da/ or /ta/) and with negative SOAs more often dorsal responses (/ga/ or /ka/). The decline seems to be less strong for visual /ga/, auditory [da]. This is probably related to the better unimodal auditory identification of auditory [da]. However, also here the incline for /ga/ responses and

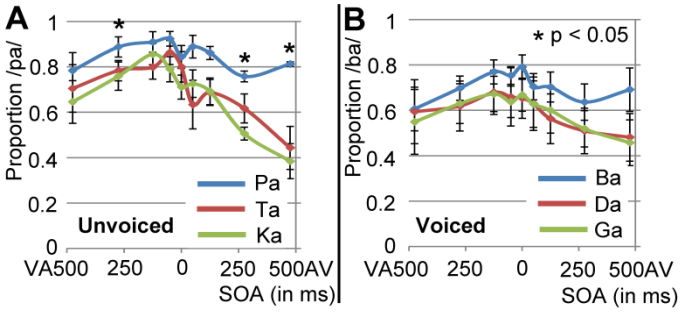


Figure 5. Results for visual /pa/ presentation for unvoiced stimuli (A) and voiced (B) stimuli. Only response proportions are shown for the response options that were given above chance level. These response options were /pa/ and /ba/, for unvoiced and voiced stimuli respectively.

decline for /da/ responses around zero is observable. The 'crossing' could relate to inherent differences in onsets between visual and auditory signals for coronal and dorsal stimuli. Indeed, a 2*3 ANOVA with factors POA and VC comparing onset differences between the maximal amplitude for visual velocity and auditory signal showed an effect of POAs (see figure 6C; $F(1,12) = 8.600$, $p = 0.005$). Pairwise comparisons showed that dorsal stimuli had significantly bigger AV onset differences than coronal or bilabial stimuli (dorsal-coronal: $t(5) = 2.757$, $p = 0.012$; dorsal-bilabial: $t(5) = 1.941$, $p = 0.033$; bilabial-coronal: $t(5) = 0.466$, $p = 1.000$). In our stimulus set we did not find a significant difference between voiced and unvoiced stimuli ($F(1,12) = 0.800$, $p = 0.389$), so we collapsed this for further analyses and figures.

To model whether these inherent differences in onset asynchronies could explain the observed crossing, a new mixed model analysis was conducted. Therefore, we changed the factor SOA into a quantitative

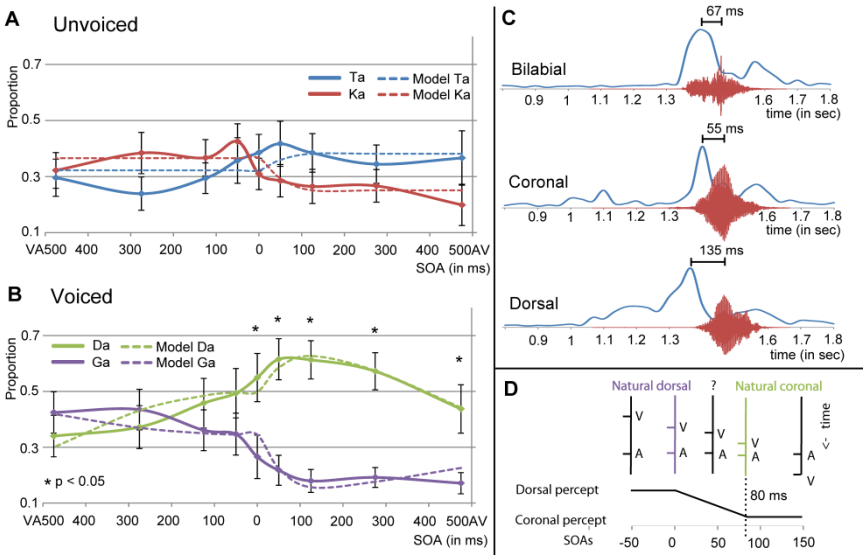


Figure 6. Results for visual /ga/ presention for unvoiced (A) and voiced (B) auditory stimuli. (C) shows the onset differences in visual velocity and auditory amplitude for the different place of articulations (POAs). (D) shows the predictor for the mixed model analyses using the natural dorsal and coronal onset asynchronies. The fit of the model together with the other significant predictors in the mixed model analyses is represented in (A) and (B) as dashed lines.

factor as described in figure 6D. The logic of the model is as follows: since both unimodal stimuli alone cannot conclusively define the identity of the stimulus (auditory unimodal can differentiate voicing, but visual unimodal can only exclude bilabial), two options are left. Our perceptual system might resolve this issue by using another cue, namely time differences between audiovisual syllable pairs. In our stimulus set, a SOA of zero is equal to the onset asynchronies of dorsal stimuli, because we aligned the stimuli based on the maximal amplitude of auditory [ga] (see figure 6C and 6D). The difference between dorsal and coronal onsets is on average 80 ms (average audiovisual asynchrony for dorsal is 135 ms and for coronal 55 ms). Therefore, the SOA for coronal stimuli in our stimulus set would be around + 80ms. With SOAs bigger than 80 ms the onset asynchronies match closer to coronal than to dorsal asynchronies. The opposite is true for audiovisual pairs with a long (experimental) visual lead: the onset asynchronies are close to dorsal asynchronies. In between these natural lags there is an ambiguity with regard to the identity of the stimulus. This factor therefore specifically tests our hypothesis that dependent on the audiovisual onset difference, participants would be biased in choosing the dorsal or coronal option, which provides new insight in the mechanism of how the percept is formed in case of ambiguous inputs. Additionally, we added a second order polynomial to the analyses to account for the downslope at the extremes.

The results of this mixed model showed an interaction between response and the created factor in both the unvoiced and voiced analyses (figure 6B; $F(1,385) = 22.446$, $p < 0.001$ and $F(1,379) = 58.166$, $p < 0.001$, respectively), indicating that indeed modelling the natural lag in audiovisual syllables explains the difference in the response choice for the different SOA. In both voicing levels dorsal responses had positive and coronal responses negative values for the parameter estimate (Unvoiced: parameter estimate -0.1410 and 0.0689 for /ta/ and /ka/ respectively and Voiced: parameter estimate -0.2212 and 0.1674 for /da/ and /ga/ respectively), verifying the hypothesized pattern of the effect in which negative SOAs should result in a dorsal percept. As in the previous analyses, POA showed an interaction with response for the visual /ga/ stimulus ($F(2,379) = 26.731$, $p < 0.001$). The second order factor was only of significance in the analyses with the voiced stimuli and showed an

interaction with response ($F(1,379) = 22.279$, $p < 0.001$), such that the parameter estimate was more negative for the /ga/ response.

Discussion

The current study investigated the influence of content and temporal cues on the identification of audiovisual syllables. We hypothesized that visual input influences the percept only within a constrained temporal window. Furthermore, we predicted that the more reliable unimodal content cues determine the percept more strongly. Finally, we hypothesized that information about natural audiovisual onset differences can be used to identify syllables. We revealed that during audiovisual speech perception visual input determines the POA and auditory input determines the voicing. Moreover, we confirmed the prediction of a wide window in which visual information influences auditory perception that was wider for congruent stimulus pairs. Interestingly, within this window, the syllable percept was not consistent, but differed depending on the specific SOA. This was particularly pronounced when the POA could not be reliably identified (i.e. between dorsal and coronal stimuli). We explained this temporal response profile using information about natural onset differences between the auditory and visual speech signals, which are indeed different for the dorsal and coronal syllables.

Multiple unimodal cues for audiovisual speech identification

Our data suggest that participants used the visual signal to identify the POA and the auditory signal to identify voicing during audiovisual presentation. We suggest that it is the reliability of the cue for the specific features of the syllable that determined the percept, since it has been shown before that the reliability of a cue can determine the percept (Andersen, Tiippana, & Sams, 2004; Massaro, 1997). This is also in line with our replication of the results that unimodally, visual stimuli are best dissociable by using POA and auditory stimuli are best dissociable by using voicing (Summerfield, 1987; Van Wassenhove, et al., 2005; Wiener & Miller, 1946). It appears that irrespective of the task, which was to identify the auditory stimulus, visual input influences perception.

Therefore, it seems that audiovisual speech is automatically integrated, since participants were not able to perform the task using only auditory cues as instructed. Integration in our study is shown by different identification responses for auditory and audiovisual presentation of the same spoken syllables. This perceptual effect is similar to the McGurk effect, in which identification of an auditory syllable is involuntarily influenced by an incongruent visual input (Gentilucci & Cattaneo, 2005; Soto-Faraco, Navarra, & Alsius, 2004). This indicates that during audiovisual speech perception, an integrated percept is created that uses the information of the visual as well as the auditory domain. In the current setting, since the auditory signal is non-optimal, this leads to a considerable bias in favour of the visual POA, for which the visual input is most reliable and thus dominant. In the McGurk effect, both signals are equally salient, resulting in a fused percept. So, when a unimodal signal is dominant during audiovisual integration, this predisposes perception.

Content predictions in audiovisual speech

In the current study we manipulated the predictability of the visual signal by using one visual syllable in which the POA can reliably be determined (/pa/) and another syllable in which the POA estimate is less reliable (/ga/). Previous research has shown that the information present in the visual signal is used to determine our percept, for example, Van Wassenhove (2005) showed facilitation of congruent speech dependent the amount of content information in the visual stimuli. Consistent with our results, Van Wassenhove and colleagues showed that, /pa/ stimuli which convey more content information about POA, influenced electro-encephalographic recordings more than a less informative syllable /ka/. In their study, an analyses-by-synthesis framework was proposed in which the auditory signal is evaluated, based on the predictive strength the visual signal has for the content of the auditory signal. This predictive strength should determine whether there is a McGurk effect (Van Wassenhove, et al., 2005) and should also correlate with prediction error when an incongruent auditory stimulus is presented (Arnal, et al., 2011). In a study using congruent audiovisual speech with auditory speech in white noise, Pandey, Kunov and Abel (1986) showed that more proficient

lip readers can still detect the auditory signal at higher noise levels, indicating that the predictive strength or the amount of information conveyed by the visual signal, influences the amount of benefit during auditory perception. Here, we also show that more predictable visual bilabial stimuli bias the percept more strongly, because visual /pa/ shaped the percept more profoundly than visual /ga/. This is in line with results from Vatakis and colleagues (2012) who found that the point of perceived synchrony needed more visual lead for stimuli pronounced more in the back of the mouth compared to bilabial stimuli. They argue that for more salient visual stimuli (i.e. bilabial stimuli) a smaller visual lead is required to reach synchrony perception. In our study, this is reflected in the amount of bias of the visual signal for the POA response choice. Since the auditory signal had a low signal to noise ratio, the visual signal biases the percept of POA completely, such that unimodal and audiovisual POA response proportions were the same.

Interplay between two distinct temporal cues in audiovisual speech perception

It is well-known that temporal cues are informative for audiovisual speech identification (Munhall & Vatikiotis-Bateson, 2004; Zion Golumbic, Poeppel, & Schroeder, 2012). Firstly, auditory and visual speech seems to temporally co-vary (Campbell, 2008). Especially in theta frequencies around 2-7 Hz, lip movement and the auditory envelope seem to correlate (Chandrasekaran, et al., 2009; Luo, Liu, & Poeppel, 2010; Müller & MacLeod, 1982). This feature has been considered a main source of binding and of the parsing of information (Campbell, 2008; Ghazanfar, Morrill, & Kayser, 2013; Poeppel, 2003) and removing this frequency reduces auditory intelligibility (Ghitza, 2012; Vitkovitch & Barber, 1994). Secondly, visual signals generally precede auditory signals, providing temporal predictability of the arrival of the auditory signal (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). Finally, audiovisual speech perception has generally been shown to have a broad integration window (Dixon & Spitz, 1980; Grant & Greenberg, 2001), which has led to the conclusion that audiovisual speech perception has loose temporal associations (Munhall & Vatikiotis-Bateson, 2004). Our results also

indicate that visual input influences the auditory percept for a wide range of SOAs. For example, we show that with auditory [ba] and visual /ga/, the visual signal influences the percept for a time window in which the visual signal is shifted 500 ms earlier in time, relative to the auditory signal, up to when the visual signal was shifted 300 ms later in time, relative to the auditory signal (SOAs ranging from VA 500 up to AV 300 ms). Only at the most positive SOA (AV 500) is visual information not used and the correct answer [ba] is present in the given responses.

Although we find integration during a wide window, the results do not support a very loose temporal association, since we also found evidence for the use of natural temporal audiovisual onset differences in identifying the syllable. However, this information was only used when unimodal cues did not provide enough information. Therefore, we propose the following mechanism for the interplay of articulatory cues (POA and voicing), temporal integration cues, and temporal onset cues (see figure 7): First, the visual and auditory components of a syllable activate syllable representations based on their "preferred" cue and reliability. However, these activations have some decay, such that at some point in time after the visual stimulus was presented, visual information does not influence the percept anymore (the temporal window of integration, TWI). Within this window more reliable cues will cause more activation of specific representations (i.e., visual cues will activate representations of syllables with corresponding POAs and auditory cues will activate representations of syllables with corresponding voicing). In a winner-takes-all framework, which is the case in an identification task, only one representation can win and that will be the representation with the strongest input. However, in addition to the visual and auditory articulatory cues, the activation of syllable representation is also based on the encoded natural onset differences. That is, for dorsal stimuli (e.g., /da/), maximal activation will occur later than for coronal stimuli (e.g., /ga/). When an ambiguous auditory stimulus arrives, it will activate multiple representations (the three voiced representations in the figure). The representation that is most active at that point in time, depending on the audiovisual onset difference, will win the competition. In the figure, visual /ga/ input cannot dissociate the coronal (/da/ and /ta/) from the dorsal (/ga/ and /ka/) POA, and auditory information cannot dissociate the POA at all. Therefore, if the auditory stimulus arrives early

(resembling natural coronal audiovisual onset differences), the most active representation will win the competition, in this example /da/. For later presentation, /ga/ will be more activated, and when the decay is completed there is no bias from the visual cue (since no representations are active), and one of the three voiced stimuli has to be chosen. This way, audiovisual onset differences only influence identification when ambiguous auditory stimuli are presented within the TWI, and only if the visual POA cues are not decisive.

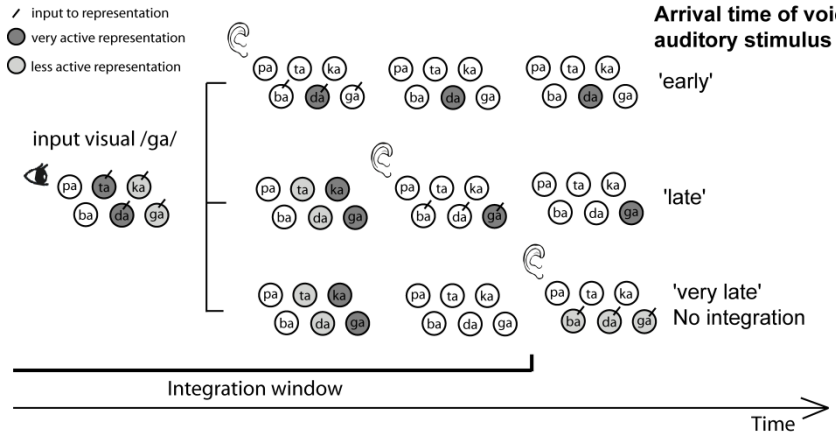


Figure 7. Proposed mechanism explaining the interplay between place of articulation (POA), voicing, temporal integration and temporal onset cues. The figure shows what happens during the presentation of visual /ga/ and an ambiguous voiced stimulus. For the visual syllable /ga/, POA cues present in the visual signal activate (indicated by darker circles) coronal (/ta/ and /da/) and dorsal (/ka/ and /ga/) representations in a time-dependent manner: the activation decays over time (indicating the TWI), and depending on the natural audiovisual onset differences, maximal activation occurs at different time points for the two POAs (later for dorsal than for coronal).Therefore, the time when auditory information activates representations of syllables (represented along the vertical axis) is important for winning the decision making process. When auditory syllables arrive early, and therefore resemble more closely natural audiovisual onset differences for /da/, /da/ is more active than /ga/, and has the highest chance to win the decision making process. In this example, the visual cues can not distinguish between the different coronal and dorsal possibilities, and the auditory cues cannot distinguish the POA at all, so the arrival of the auditory information (early vs. late) facilitates this decision; early onset will activate the coronal /da/ and late onset will activate the dorsal /ga/ syllable.

Temporal window of integration is influenced by audiovisual congruency

The temporal window of audiovisual integration (TWI) is generally measured by evaluating whether participants can indicate if audiovisual events are presented simultaneously or not (Vroomen & Keetels, 2010), assuming that when participants can reliably dissociate the two, the audiovisual event is perceived as two separate events and not bound together. However, little research has been done to assess whether audiovisual SOA differences also influence unimodal perception, which was one of the aims of the current study. Applying the same logic as that used for simultaneity judgments, events that are bound should influence unimodal perception more than when they are perceived separately. We here show that especially during congruent audiovisual voicing (visual /pa/, auditory unvoiced), the response proportions of /pa/ are higher (figure 5). Also, visual influence seems to have a wider temporal window of integration for the congruent pairing of visual /pa/ with auditory /pa/, as the visually determined /pa/ response proportion appears higher for a wider temporal window (although the statistical test did not show this). One explanation for these congruency effects is the 'unity assumption' stating that when two stimuli naturally belong together they are bound more strongly and therefore are more difficult to dissociate over a wider temporal window (Welch & Warren, 1980). However, it could be that with extreme SOAs, visual information is not used and participants rely only on the auditory signal, that is, in the case of congruent audiovisual /pa/ pairing they would also report /pa/ with auditory presentation only. Nonetheless, the unimodal auditory experiment showed that the POA for unvoiced stimuli could not be dissociated, neither could it for /pa/. Thus, the use of auditory information alone should not result in a higher proportion of /pa/ responses. For the incongruent pairs, identification with the most positive SOA seems similar to unimodal unvoiced auditory perception, hence participants did not seem to use visual information, indicating that for this SOA integration did not take place. Similar results have been found by Vatakis and Spence (2007), who showed that judging simultaneity is more difficult when the gender of the speaker is congruent with the speech sound. Although there are also conflicting results, for speech the unity assumption seems plausible (Vroomen & Keetels, 2010).

One difference between simultaneity judgments and stimulus identification across SOAs seems to be that the point of maximal integration is more biased towards visual leading when explicitly asking about identity (Van Wassenhove, et al., 2007; Zampini, et al., 2003). Therefore, varying SOAs and measuring unimodal perception might provide a different approach to measure whether integration occurs over a broader range of SOAs. This approach does not investigate whether two stimuli are perceived as simultaneously, but serves the goal to investigate the temporal patterns in which a unimodal stimulus influences the perception of another unimodal stimulus, for example the content of a stimulus. This judgment might be more natural, since in daily life, identifying stimuli is a more common act than explicitly judging their coincidence.

Possible neuronal mechanisms

Based on previous literature, the brain area most consistently involved in audiovisual integration is the posterior superior temporal sulcus (Calvert & Lewis, 2004). It has been found active during visual and audiovisual speech perception (Callan et al., 2004; Calvert et al., 1997), seems to be sensitive for congruent versus incongruent speech signals (Calvert, Campbell, & Brammer, 2000; Van Atteveldt, Formisano, Goebel, & Blomert, 2004; Van Atteveldt, Blau, Blomert, & Goebel, 2010), and responds to audiovisual onset differences (Chandrasekaran & Ghazanfar, 2009; Van Atteveldt, et al., 2007). In the temporal domain it seems that different temporal features (co-variations between mouth velocity and speech envelope and visual-auditory speech onset differences) have to be combined to shape our percept. Chandrasekaran and Ghazanfar (2009) showed that different frequency bands are differently sensitive for faces and voices in superior temporal cortex. Although theta oscillations have been shown to be influenced by input from other senses (Kayser, Petkov, & Logothetis, 2008; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007), they have not been shown to have specific effects dependent on the voice-face onset differences and might therefore mostly be used to parse the auditory signals, enhance auditory processing, and might even relate to the audiovisual TWI (Poeppel, 2003; Schroeder, et al., 2008). However,

higher frequency oscillations have been shown to vary dependent on voice-face onset differences, and might be involved in encoding the identity of a syllable, thus explaining the current results. This is consistent with the notion that the auditory speech system depends on theta as well as gamma frequencies (Poeppel, 2003), and this latter time-scale might also be important in coding differences in natural audiovisual onset differences, and its influence on perception. These temporal constraints however would have to be investigated, for example by using combined behavioural and electrophysiological measures, or using transcranial magnetic stimulation at varying time points.

Conclusion

Our findings show that within the integration window, visual information biases the auditory percept, specifically regarding the features in which the auditory signal is ambiguous (i.e. POA). Additionally, these findings indicate that natural temporal onset differences between auditory and visual input have a noteworthy influence on auditory perception. Although visual input has an influence over a wide temporal window during our experiment, we show that this initial binding of information does not conclusively determine our percept. Instead, it serves as a prerequisite for other interaction processes to occur that eventually form our perceptual decision. The final percept is determined by the interplay between unimodal auditory and visual cues, along with natural audiovisual onset differences across syllables. These results shed light on the compositional nature of audiovisual speech, in which visual, auditory, and temporal onset cues are used to create a percept. This interplay of cues needs to be studied further to unravel the building blocks and neuronal basis of audiovisual speech perception.

Acknowledgements: This study was supported by a grant from the Dutch Organization for Scientific Research (NWO; grant number 406-11-068)

References

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research, 21*(3), 301-308.

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience, 14*(6), 797-801.

Auer Jr, E. T., & Bernstein, L. E. (1997). Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *The Journal of the Acoustical Society of America, 102,* 3704.

Auer Jr, E. T., & Bernstein, L. E. (2007). Enhanced visual speech perception in individuals with early-onset hearing impairment. *Journal of Speech, Language and Hearing Research, 50*(5), 1157.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809-823.

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication, 44*(1), 5-18.

Bien, N., Ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *Neuroimage, 59*(1), 663-672.

Blau, V., Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2008). Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European journal of neuroscience, 28*(3), 500-509.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience, 16*(5), 805-816.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*(5312), 593-596.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649-657.

Calvert, G. A., & Lewis, J. W. (2004). Hemodynamic Studies of Audiovisual Interactions. In S. C. Calvert G.A., Stein B. (Ed.), *The Handbook of Multisensory Processes* (pp. 483-502). Cambrige: MIT Press.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1001-1010.

Chandrasekaran, C., & Ghazanfar, A. A. (2009). Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. Journal of *Neurophysiology, 101*(2), 773-788.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception, 9*(6), 719-721.

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research, 1242,* 136-150.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*(6577), 66-68.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169.

Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*(1), 66-75.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *The Journal of Neuroscience, 25*(20), 5004-5012.

Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences, 110*(5), 1959-1963.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3.*

Grant, K. W., & Greenberg, S. (2001). *Speech intelligibility derived from asynchronous processing of auditory-visual information.* Paper presented at the AVSP 2001-International Conference on Auditory-Visual Speech Processing.

Grant, K. W., Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony. *Speech Communication, 44*(1), 43-53.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex, 18*(7), 1560-1574.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*, 467.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS biology, 8*(8), e1000445.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-141.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry:* Lawrence Erlbaum.

Massaro, D. W. (1997). *Perceiving talking faces: From speech perception to a behavioral principle:* Mit Press.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America, 100,* 1777.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748 .

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*(10), 3215-3229.

Miller, J. L. (1977). Properties of feature detectors for VOT: The voiceless channel of analysis. The Journal of the Acoustical Society of America, 62, 641.

Müller, E., & MacLeod, G. (1982). Perioral biomechanics and its relation to labial motor control. *The Journal of the Acoustical Society of America, 71*(S1), S33-S33.

Munhall, K., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd & B. D. (Eds.), *Hearing by Eye II* (pp. 123-139): Sussex: Taylor and Francis.

Munhall, K., & Vatikiotis-Bateson, E. (2004). Spatial and Temporal Constraints on Audiovisual Speech Perception. In S. C. Calvert G.A., Stein B. (Ed.), *The Handbook of Multisensory Processing* (pp. 177-188). Cambridge, Massachusetts: The MIT Press.

Pandey, P. C., Kunov, H., & Abel, S. M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *Journal of Auditory Research, 26*(1), 27-41.

Parise, C. V., & Spence, C. (2009). "When Birds of a Feather Flock Together": Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes. *PloS one, 4*(5), e5664.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245-255.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition, 92*(3), B13-B23.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses:* The MIT Press.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research,* 1-17.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage, 44*(3), 1210-1223.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Summerfield, A. (1987). Some preliminaries to a theory of audiovisual speech processing. In B. D. R. Campbell (Ed.), *Hearing by eye* (pp. 58-82): Hove, UK: Erlbaum Associates.

Van Atteveldt, N., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex, 17*(4), 962-974.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*(2), 271-282.

Van Atteveldt, N. M., Blau, V. C., Blomert, L., & Goebel, R. (2010). fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC neuroscience, 11*(1), 11.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vatakis, A., Maragos, P., Rodomagoulakis, I., & Spence, C. (2012). Assessing the effect of physical differences in the articulation of consonants and vowels on audiovisual temporal perception. *Frontiers in Integrative Neuroscience, 6*, 71.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain research, 1111*(1), 134-142.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Attention, Perception, & Psychophysics, 69*(5), 744-756.

Vitkovitch, M., & Barber, P. (1994). Effect of video frame rate on subjects' ability to shadow one of two competing verbal passages. *Journal of Speech, Language and Hearing Research, 37*(5), 1204.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871-884.

Wallace, M., Wilkinson, L., & Stein, B. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246-1266.

Welch, R., & Warren, D. (1986). Intersensory interactions. In K. R. Boff, et al., (Ed.), *Handbook of perception and human performance.* (Vol. 1, pp. 25.21-25.36): Wiley.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638.

Wiener, F., & Miller, G. A. (1946). Some characteristics of human speech. *Transmission and reception of sounds under combat conditions. Summary Technical Report of Division 17, National Defense Research Committee,* 58-68.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*(2), 198-210.

Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual Input Enhances Selective Speech Envelope Tracking in Auditory Cortex at a "Cocktail Party". *The Journal of Neuroscience, 33*(4), 1417-1426.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*(3), 151-161.

CHAPTER 7

OSCILLATORY PHASE SHAPES SYLLABLE PERCEPTION

Corresponding Manuscript:
Ten Oever, S., Sack, A.T. (2015). Oscillatory phase shapes perception. *Proceedings of the National Academy of Sciences of the United States of America, 112*(52), 15833-15837

Abstract

The role of oscillatory phase for perceptual and cognitive processes is being increasingly acknowledged. To date little is known about the direct role of phase in categorical perception. Here, we show in two separate experiments that the identification of ambiguous syllables that can either be perceived as /da/ or /ga/ is biased by the underlying oscillatory phase as measured with EEG and sensory entrainment to rhythmic stimuli. The measured phase difference in which perception is biased towards /da/ or /ga/ exactly matched the different temporal onset delays in natural audiovisual speech between mouth movements and speech sounds, which lasts 80 ms longer for /ga/ than for /da/. These results indicate the functional relationship between pre-stimulus phase and syllable identification and signify that the origin of this phase relationship could lie in exposure and subsequent learning of unique audiovisual temporal onset differences.

Introduction

In spoken language, visual mouth movements naturally precede the production of any speech sound and therefore serve as a temporal prediction and detection cue for identifying spoken language [(Campbell, 2008), but also see (Schwartz & Savariaux, 2014)]. Different syllables are characterized by unique visual-to-auditory temporal asynchronies (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009; Ten Oever, Sack, Wheat, Bien, & Van Atteveldt, 2013). For example /ga/ has a 80 ms longer delay as /da/ and this difference aids categorical perception of these syllables (Ten Oever, et al., 2013). We propose that neuronal oscillations might carry the information to dissociate these syllables based on temporal differences. Multiple authors have proposed (Luo & Poeppel, 2007; Peelle & Sommers, 2015; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008) – and it has been demonstrated empirically (Luo & Poeppel, 2007; Perrodin, Kayser, Logothetis, & Petkov, 2015; Van Atteveldt, Murray, Thut, & Schroeder, 2014) - that at the onset of visual mouth movements ongoing oscillations in auditory cortex align [see (Besle et al., 2011; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Mercier et al., 2015) for non-speech phase reset], providing a temporal reference frame for the auditory processing of subsequent speech sounds. Consequently, auditory signals fall on different phases of the aligned oscillation depending on the unique visual-to-auditory temporal asynchrony, resulting in a consistent relation between syllable identity and oscillatory phase.

We hypothesized that this consistent "phase-syllable" relationship results in ongoing oscillatory phase biasing syllable perception. More specifically, the phase at which syllable perception is mostly biased should be proportional to the visual-to-auditory temporal asynchrony found in natural speech. A naturally occurring /ga/ has an 80 ms longer visual-to-auditory onset difference than a naturally occurring /da/ (Ten Oever, et al., 2013). Consequently, the phase difference between perception bias toward /da/ or /ga/ should match 80 ms, which can only be established with an oscillation with a period greater than 80 ms, that is, any oscillation under 12.5 Hz. The apparent relevant oscillation range is therefore theta, with periods ranging between 111-250 ms (4-9 Hz).

This oscillation range has already been proposed as a candidate to encode information and seems specifically important for speech perception (Kayser, Ince, & Panzeri, 2012; Peelle & Davis, 2012).

To test this hypothesis of oscillatory phase biasing auditory syllable perception in the absence of visual signals, we presented ambiguous auditory syllables which could be interpreted as /da/ or /ga/ while recording EEG. In a second experiment we used sensory entrainment (thereby externally enforcing oscillatory patterns) to demonstrate that entrained phase indeed determines whether participants identify the presented ambiguous syllable as being either /da/ or /ga/.

Results

Experiment 1

Psychometric curves: first, we created nine morphs between a /da/ and /ga/ by shifting the third formant frequency of a recorded /da/ from around 3000 to 2600 Hz (figure 1A). We determined the individual threshold at which participants would identify a morphed stimulus 50% as /da/ and 50% as /ga/ by repeatedly presenting the nine different morphs and participants had to indicate their percept (see SI experimental procedures for details). Indeed, 18 out of 20 participants were sensitive to the manipulation of the morphed stimulus and psychometric curves could be fitted reliably (figure 1B; average explained variance of the fit was 92.7%, standard deviation of 0.03). The other two participants were excluded from further analyses.

Consistency of phase differences: we used the individually determined most ambiguous stimuli to investigate whether ongoing theta phase prior to stimulus presentation influenced the identification of the syllable. Therefore we presented both the unambiguous /da/ (stimulus 1) and /ga/ (stimulus 9) and the ambiguous stimulus while recording EEG. Data was epoched -3 to 3 sec around syllable onset. To ensure that post-stimulus effects did not temporally smear back to the pre-stimulus interval [see e.g.

(Zoefel & Heil, 2013)] we padded all data points after zero with the amplitude value at zero. For every participant we extracted the average phase for each of the syllable types for the -0.3 to 0.2 sec interval. There were four syllable types: the /da/ and /ga/ of the unambiguous sounds and the ambiguous sound either perceived as /da/ or /ga/. Then, we determined the phase difference between /da/ and /ga/ for both the unambiguous and ambiguous condition. In the ambiguous condition pre-stimulus phase is hypothesized to bias syllable perception and this should be reflected in a consistent phase difference between the perceived /da/ and /ga/. During the unambiguous condition phase in the pre-stimulus

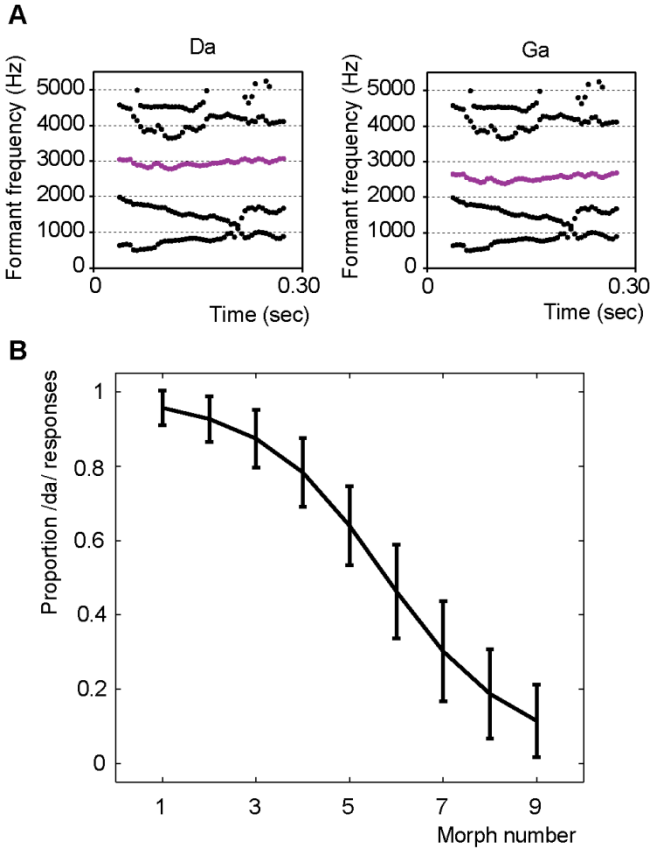


Figure 1. Results from the morphed /daga/ stimuli. A) Stimulus properties of the used /da/ and /ga/ stimulus. Only the third formant differs between the two stimuli (purple line). B) Average proportion /da/ responses for the 18 participants in Experiment 1. Error bars reflect the standard error of the mean.

time windows should mostly reflect random fluctuations as participants are unaware of the identity and arrival time of the upcoming syllable and participants generally identified stimulus 1 as /da/ stimulus 9 as /ga/, resulting in a low consistency of the phase difference. Note however that in principle phase differences are possible in this condition as we did exclude trials in which participants identified the unambiguous syllables as the syllable at the other side of the morphed spectrum. The mean resultant vector lengths (MRVL) of the phase difference between /da/ and /ga/ were calculated and Monte-Carlo simulations with a cluster-based correction for multiple comparisons were used for statistically testing. A higher MRVL indicates a higher phase concentration of the difference. We found that the ambiguous sounds had a significantly higher MRVL before sound onset (-0.25 to -0.1 ms) around 6 Hz (cluster statistics = 19.821, p = 0.006; figure 2A and B). When repeating the

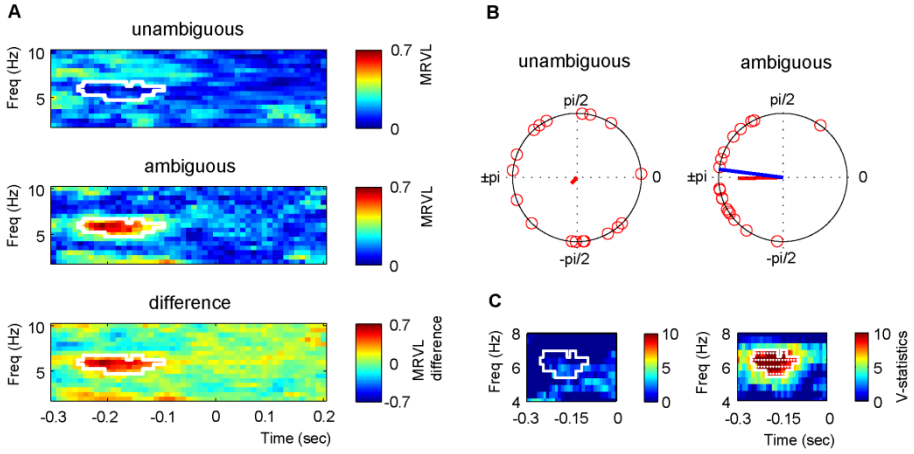


Figure 2. Pre-stimulus phase differences. A) The mean resultant vector length (MRVL) over participants for the phase difference between /da/ and /ga/ for the unambiguous sounds and for the phase difference between perceived /da/ and /ga/ for the ambiguous sounds. The white rectangle indicates the region of significant differences. B) Phase differences of individual participants at 6 Hz at -0.18 sec for the unambiguous and ambiguous sounds. The blue line indicates the 80 ms expected difference. The red line indicates the strength of the MRVL. C) The V-statistics testing whether the phase differences are significantly non-uniformly distributed around 80 ms for all significant points at the MRVL analysis. The white rectangles indicate at which time and frequency point the analysis was performed (note the difference in the x and y axes between A and C). White dots indicate significance.

analysis including a wider frequency spectrum (1-40 Hz) the same effect was present (cluster statistics = 18.164, p = 0.030), showing the specificity of the effect for theta. Since any phase estimation requires integration of data over time, the significant data appears distant from the onset of the syllable. For example, the 6 Hz phase angle is calculated using a window of 700 ms (to ensure the inclusion of multiple cycles of the theta oscillation). The closer the center of the estimation is to an abrupt change in the data (such as a stimulus or the data padding to zero), the more the estimation is negatively influenced by the "post-change data" [see e.g. (Zoefel & Heil, 2013)].

80ms phase differences: a second hypothesis was that the phase difference of the ambiguous stimuli judged as /da/ vs. /ga/ would match 80 ms, consistent with the visual-to-auditory onset difference between /da/ and /ga/ found in natural speech (Ten Oever, et al., 2013). Therefore we took all the significant time and frequency points in the first analysis and tested whether the phase difference of all participants was centered around 80 ms (the blue line in figure 2B corresponds to an 80 ms difference). This is typically done with the V-test that examines the non-uniformity of circular data centered around a known specific mean direction. We found that the ambiguous phase differences indeed centered around 80 ms for almost all tested data points, while for the unambiguous sounds no such phase concentration was present (figure 2C).

From figure 2B it is evident that there is a consistent phase difference over participants between /da/ and /ga/ for the ambiguous sounds. When looking at the consistency of the phases of the individual syllables /da/ and /ga/ this consistency drops (compare figure 2B with figure S1B). Statistical testing confirmed that that /da/ and /ga/ phases seemed distributed randomly (figure S1C). At this point we cannot differentiate whether this effect occurs due to volume conduction of the EEG or individual latency differences for syllable processing [see also (Lakatos, et al., 2007)]. When repeating this analysis for each participant we did find a significant (uncorrected) consistency for multiple

participants and a significant different phase between /da/ and /ga/ (figure S2; for only two participants this effect survived correction for multiple comparisons).

The current reported effects could not be explained by any eye movements (no significant differences between conditions) or any artefacts due to the data padding (figure S3).

Experiment 2

To investigate whether neuronal entrainment results in oscillatory identification patterns we experimentally induced theta phase alignment using sensory entrainment (de Graaf et al., 2013; Henry & Obleser, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008) in twelve different participants. In this experiment auditory stimuli of broadband noise (white noise band-pass filtered between 2.5 and 3.1 kHz, 50 ms length) were repeatedly presented (presumably entraining underlying oscillations at the presentation rate) after which ambiguous sounds were presented at different stimulus onset asynchronies (SOA's; 12 different SOA's fitting exactly 2 cycles). If ongoing phase is important for syllable identification, the time course of identification should oscillate at the presentation rate. Indeed, the time course of identification showed a pattern varying at the presentation rate of 6.25 Hz (figure 3A). To test the significance of this effect we calculated the relevance value (Fiebelkorn et al., 2011). This value is calculated by 1) fitting a sinus to the data and 2) multiplying the explained variance of the fit with the variance of the predicted values. In this way the relevance statistic gives less weight to models that have a fit with a flat line. Thereafter, we performed bootstrapping on the obtained relevance values (of the average curve) to show that of the 10,000 fitted bootstraps only 2.83% had a more extreme relevance value (figure 3B), suggesting that indeed syllable identity depends on theta phase.

Three control experiments were performed. In the first two experiments the frequency specificity of the effect was investigated by changing the presentation rates of the entrainment train to 1 and 10 Hz. In a third experiment we wanted to rule out the possibility that the effect already occurs at a lower perceptual level instead of the syllable

identification level. Therefore, we band-passed filtered the syllables between 2.5 and 3.1 Hz, maintaining the formant frequency at which the two syllables differ, but distorting syllable perception. Participants had to indicate whether they felt the sound was of high or low frequency (this experiment will from now on be called frequency control). As a reference for what was considered a high or low frequency the band-passed filtered

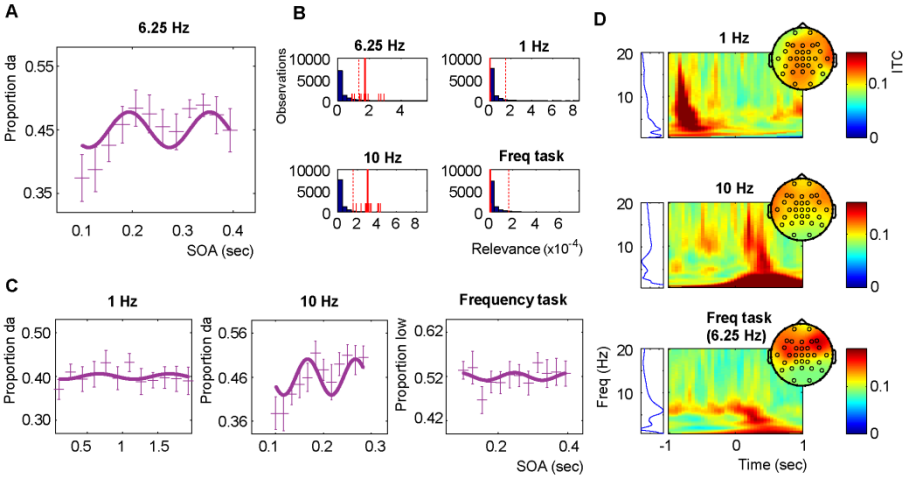


Figure 3. Results from experiment 2. A) Grand average proportion /da/ of all the participants with the respective error bars reflecting the within-subject standard error of the mean (plusses; vertical extend reflects the error bars) and the fitted 6.25 Hz sinus (solid line). B) The bootstrap histograms for the relevance statistics for all four conditions. The long solid and dotted red lines represent the relevance value of that dataset and the 95 percentile of all bootstrapped values, respectively. The short solid lines indicate the twelve relevance values when iteratively taking out one participant. C) The grand average of all participants with the respective error bars reflecting the within-subject standard error of the mean (plusses; vertical extend reflects the error bars) for the three different control conditions used in the experiment and their respective best fitted sinus (solid line). D) The inter-trial coherence (ITC) plots for all three entrainment frequencies. Zero indicates entrainment offset. The left inset indicate the ITC averaged in the -0.5-0 interval (ITC range 0.08-0.12). All the conditions show a peak at the respective entrainment frequency. However, for 1 Hz an evoked response of the last entrainment stimulus is present (around -0.8 sec). For 10 Hz, and in a lesser extend for 6.25 Hz, evoked responses to the target stimuli are present post-stimulus (around 0-1 sec). This effect only arises in these frequencies as the interval a target presented is much narrower as for the 1 Hz.

stimulus number 1 and 9 were both presented at random order at the beginning of the trial.

Results show that for both the 1 Hz and the frequency control no sinus could be fitted reliable (figure 3B and 3C; $p = 0.80$ and $p = 0.69$ respectively). In contrast, for 10 Hz a sinus could be reliably fitted ($p = 0.011$). For all three presentation frequencies there was entrainment at the expected frequency (figure 3D).

Discussion

In the current study, we investigated whether ongoing oscillatory phase biases syllable identification. We presented ambiguous auditory stimuli while recording EEG and revealed a systematic phase difference before auditory onset between the perceived /da/ and /ga/ at theta frequency. This phase discrepancy corresponded to the 80 ms difference between the onset delays of the speech sounds /da/ or /ga/ with respect to the onset of the corresponding mouth movements found in natural speech (Ten Oever, et al., 2013). Moreover, we could show that syllable identification depends on the underlying oscillatory phase induced by entrainment to a 6.25 Hz or 10 Hz presented stimulus train of broadband noise. These results reveal the relevance of phase coding for language perception and provides a flexible mechanism for statistical learning of onset differences and possibly for the encoding of other temporal information for optimizing perception.

Audio-visual learning results in phase coding

The human brain is remarkably capable of associating events that repeatedly occur together (Fiser & Aslin, 2001; Summerfield & Egner, 2009), representing an efficient neural coding mechanism for guiding our interpretation of the environment. Specifically, when two events tend to occur together, they will enhance the neural connections between each other, consequently increasing the detection sensitivity of one event in case the associated event is present (Hebb, 2002). We propose that this

could also work for temporal associations. In a previous study we showed that the onset between mouth movements and auditory speech signals differs between syllables, and that this influences syllable identification (Ten Oever, et al., 2013). For example, a naturally occurring /ga/ has an 80 ms bigger visual-to-auditory onset difference than a naturally occurring /da/ [figure 4A; (Ten Oever, et al., 2013)]. Recent theories propose that visual cues benefit auditory speech processing by aligning ongoing oscillations in auditory cortex such that the 'optimal' high excitable period coincides with the time point at which auditory stimuli are expected to arrive, thereby optimizing their processing [figure 4B; (Mercier, et al., 2015; Schroeder & Lakatos, 2009; Van Atteveldt, et al., 2014)]. If this indeed occurs, different syllables should be consistently presented at different phases of the reset oscillation (the green and blue line in figure 4B). A similar mechanism has also been proposed by Peelle and Davis (Peelle & Davis, 2012). As humans (or rather our brains) likely (implicitly) learn this consistent association between phase and syllable identity, one could hypothesize that neuronal populations coding for different syllables may begin to prefer specific phases, biasing syllable perception at corresponding phases even when visual input is absent (figure 4C). The current data indeed supports this notion as we show that the phase difference between /da/ and /ga/ fits 80 ms. The exact cortical origin of this effect cannot be unraveled with the current data, but we would expect to find these effects in auditory cortex.

Generalization of this mechanism

Temporal information is not only present in (audio-visual) speech. Therefore, any consistent temporal relationship between two stimuli could be coded in a similar vein as demonstrated here. For example, the proposed mechanism should also generalize to auditory only settings as any temporal differences caused by articulatory processes should also influence the timing of individual syllables within a word, for example, the second syllable in "baga" should arrive at a later time point as "bada". It is an open question how these types of mechanisms generalize to situations in which speech is faster or slower. However, it is conceivable that when speaking faster the visual-to-auditory onset differences

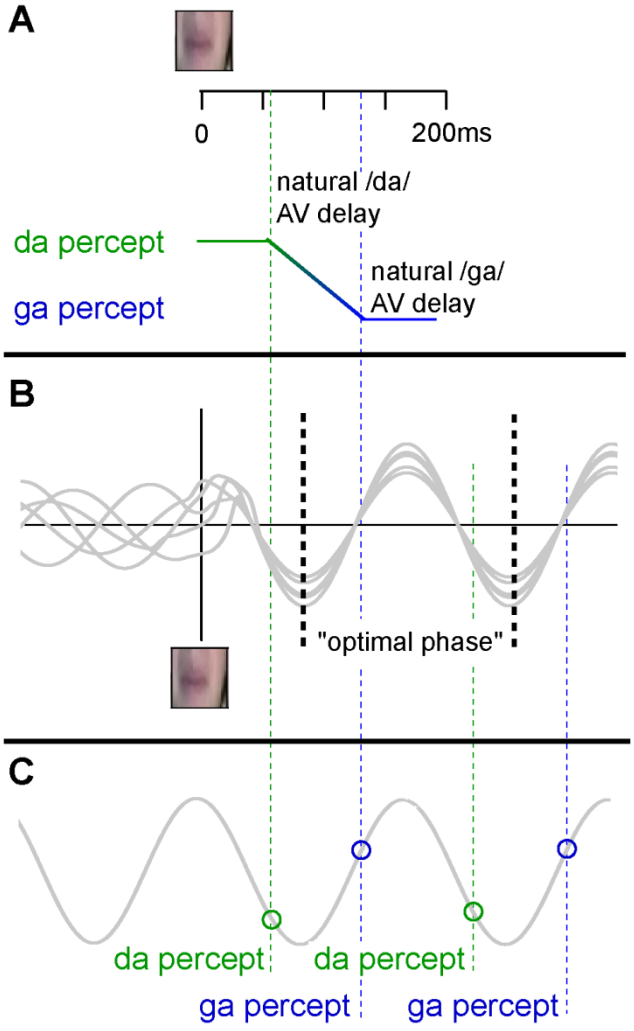


Figure 4 Proposed mechanism for theta phase sensitization. A) Dependent on the natural visual-to-auditory delay voiced-stop consonants are identified as a /da/ or a /ga/ after presenting the same visual stimulus3. B) When visual speech is presented ongoing theta oscillations synchronize creating an optimal phase (black dotted line) at which stimuli are best processed. The phase at which a /da/ or a /ga/ in natural situations is presented is different (green and blue line respectively) caused by the difference in visual-to-auditory delay. C) Syllable perception is biased at phases at which /da/ and /ga/ are systematically presented in audio-visual settings even when visual input is absent.

between /da/ and /ga/ also reduce, thereby also changing their expected phase difference. It has already been shown that cross-modal mechanisms rapidly update changing temporal statistics in the environment (Fujisaki, Shimojo, Kashino, & Nishida, 2004), by for example changing the oscillatory phase relationship between visual and auditory regions (Kösem, Gramfort, & van Wassenhove, 2014).

Our results show that during 10 Hz entrainment an oscillatory pattern of syllable identification is present. This frequency is slightly higher than what is generally considered theta. This likely reflects that the brain flexibly adapt to the changing environment, for example when facing a person that speaks very fast. Thus, although under 'normal' circumstances the effect seems constrained to theta (as shown in experiment 1), altering the brain state by entraining to higher frequencies still induces the effect and shows the flexibility of this mechanism.

Excitability versus phase coding

Much research has focused on the role of oscillations to systematically increase and decrease the excitability levels of neuronal populations (Jensen, Gips, Bergmann, & Bonnefond, 2014; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010; Schroeder & Lakatos, 2009). In this line of reasoning, speech processing is enhanced by aligning the most excitable phase of an oscillation to the incoming speech signal (Peelle & Sommers, 2015; Schroeder, et al., 2008). Intuitively, our results seem in contrast with this idea as it appears that neuronal populations coding for separate syllables have phase specific responses. However, it could also be considered possible that one neuronal population biases identification in the direction of one syllable, succeeding this bias when excited and failing when suppressed. This interpretation is less likely considering that the exact phases at which syllable identification was biased varied over participants. Therefore, the phase at which identification is biased towards one syllable does not always fall on the most excitable point of the oscillation for each participant (unless the phases of the measured EEG signal are not comparable over participants). Considering that there are individual differences in the lag between stimulus presentation and

brain response [see e.g. (Henry & Obleser, 2012)], it would also follow that the phase at which syllable identification is biased does not match over participants. However, more research is needed to irrefutably demonstrate that different neuronal populations code information preferably at a specific oscillatory phase [see (Watrous, Fell, Ekstrom, & Axmacher, 2015)].

Conclusion

Temporal associations are omnipresent in our environment and it seems highly unlikely that this data is ignored by our brain when information has to be ordered and categorized. The current study demonstrated that oscillatory phase shapes syllable perception and this phase difference matches temporal statistics in the environment. To determine whether this type of phase sensitization is a common neural mechanism it is necessary to investigate other types of temporal statistics. Especially since it could provide a mechanism for separating different representations (Fell & Axmacher, 2011; Jensen, et al., 2014; Kayser, Montemurro, Logothetis, & Panzeri, 2009) and offers an efficient way of coding time differences (Chakravarthi & VanRullen, 2012). Future research has to investigate whether also other properties are encoded in phase, revealing the full potential of this type of phase coding scheme.

Materials and methods

In total 40 participants took part in our study (20 per experiment). All participants give written informed consent. The study was approved by the local ethical committee at the Faculty of Psychology and Neuroscience at Maastricht University. Detailed methods are described in the SI materials and methods.

Acknowledgments

This study was supported by a grant from the Dutch Organization for Scientific Research (NWO; grant number 406-11-068). The authors

declare no competing financial interests. We thank Giancarlo Valente for useful suggestion for the analysis. Kirsten Petras and Helen Lückmann helped improving the final manuscript.

References

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience, 31*(9), 3176-3185.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1001-1010.

Chakravarthi, R., & VanRullen, R. (2012). Conscious updating is a rhythmic process. *Proceedings of the National Academy of Sciences, 109*(26), 10599-10604.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436.

De Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., & Thut, G. (2013). Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. *PloS one, 8*(3), e60035.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience, 12*(2), 105-118.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., & Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *The Journal of Neuroscience, 31*(27), 9971-9981.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499-504.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*(7), 773-778.

Hebb, D. O. (2002). *The organization of behavior: A neuropsychological theory:* Psychology Press.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, 109*(49), 20095-20100.

Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences, 37*(7), 357-369.

Kayser, C., Ince, R. A., & Panzeri, S. (2012). Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS computational biology, 8*(10), e1002717.

Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron, 61*(4), 597-608.

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92,* 274-284.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001-1010.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition, 115*(1), 186-191.

Mercier, M. R., Molholm, S., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Foxe, J. J. (2015). Neuro-Oscillatory Phase Alignment Drives Speeded Multisensory Response Times: An Electro-Corticographic Investigation. *The Journal of Neuroscience, 35*(22), 8546-8557.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex, 68,* 169-181.

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proceedings of the National Academy of Sciences, 112*(1), 273-278.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Schwartz, J.-L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag, *PLoS Computational Biology, 10*(7), e1003743.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Ten Oever, S., Sack, A., Wheat, K. L., Bien, N., & Van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology, 4.*

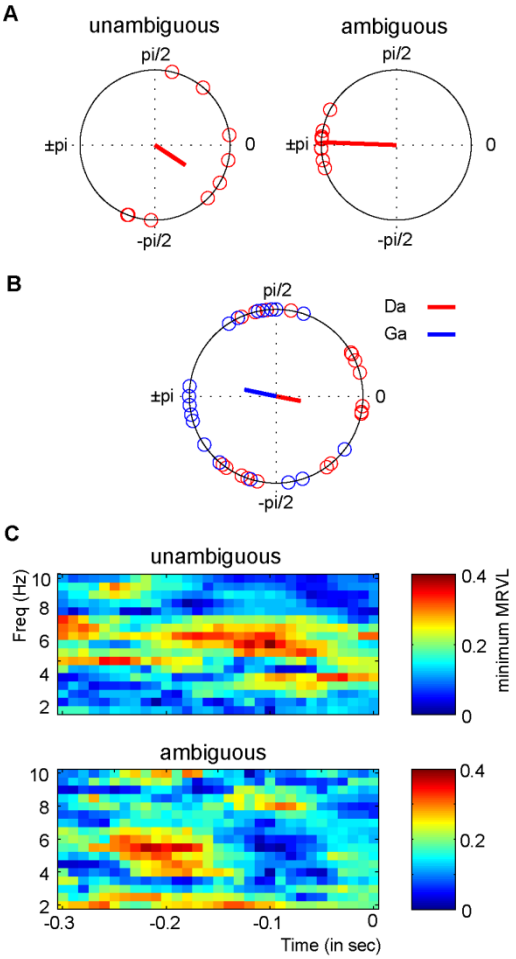
Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron, 81*(6), 1240-1253.

Watrous, A. J., Fell, J., Ekstrom, A. D., & Axmacher, N. (2015). More than spikes: common oscillatory mechanisms for content specific neural representations during perception and memory. *Current Opinion in Neurobiology, 31*, 33-39.

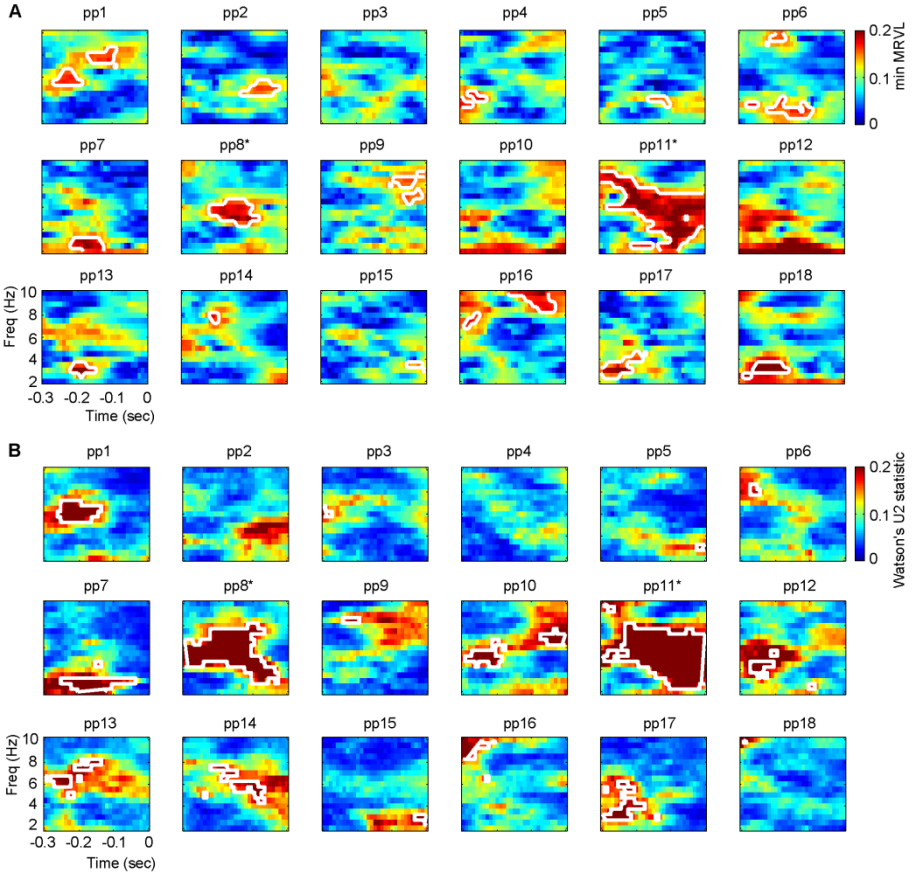
Zoefel, B., & Heil, P. (2013). Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Frontiers in Psychology, 4.*

Supporting Information

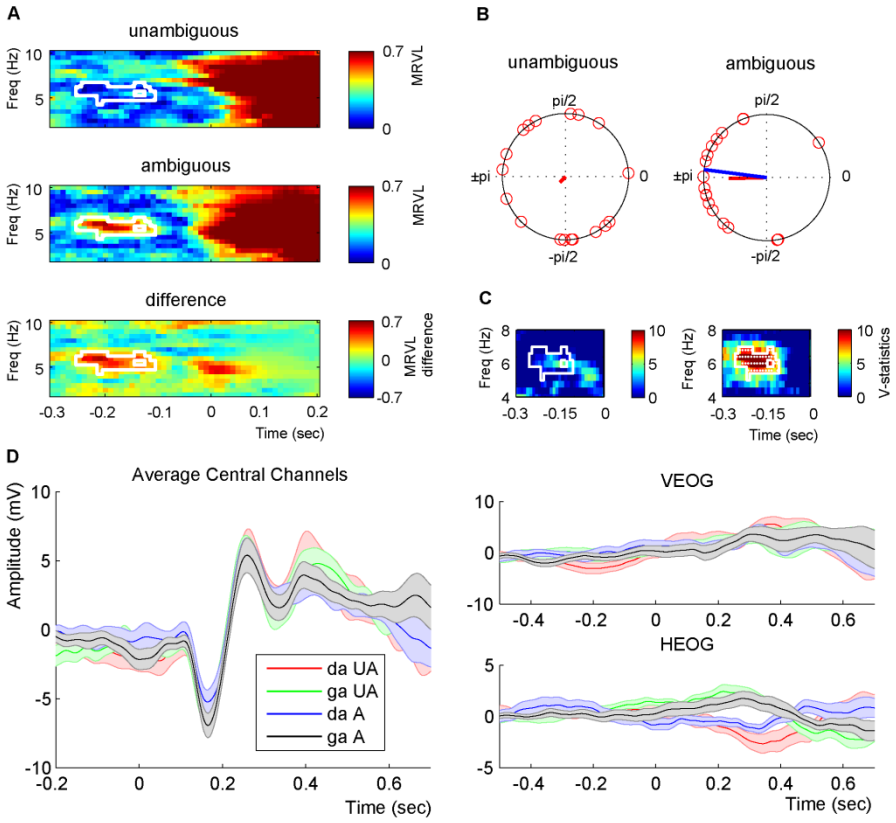
Supplementary figures



Supplementary figure 1. Channel and phase consistency. A) The phase difference between /da/ and /ga/ of all individual channels (averaged over participants). For the ambiguous sounds there is a very strong consistency for the phases of the separate channels. B) For the ambiguous condition the mean phases are plotted for each participant for /da/ (red) and /ga/ (blue) separately. Although there is some phase consistency, it is considerably less strong as the consistency of the difference (compare with figure 2B). C) The minima of the MRVL of /da/ and /ga/ for both the unambiguous and ambiguous condition.



Supplementary figure 2. Single subject analysis. A) For each participant the lowest inter-trial coherence (ITC) of the two syllable types per time/frequency data point is displayed. White rectangles indicate areas of significant minimum ITC (uncorrected). B) For each participant the Watson's U2 statistics comparing the mean phase of /da/ and /ga/ is displayed. White rectangles indicate areas of a significant Watson's U2 value (uncorrected). The asterisks indicate the participants where the effect survived cluster-based corrections for multiple comparisons.



Supplementary figure 3. Pre-stimulus phase differences for the original non-padded data and event related potentials (ERP's). A) The mean resultant vector length (MRVL) over participants for the phase difference between /da/ and /ga/ for the unambiguous sounds and for the phase difference between perceived /da/ and /ga/ for the ambiguous sounds. The white rectangle indicates the region of significant differences. B) Phase differences of individual participants at 6 Hz at -0.18 sec for the unambiguous and ambiguous sounds. The blue line indicates the 80 ms expected difference. The red line indicates the strength of the MRVL. C) The V-statistics testing whether the phase differences are significant non-uniformly distributed around 80 ms for all significant points at the MRVL analysis. The white rectangles indicate on which time and frequency point the analysis was performed. White dots indicate significance. D) The ERP's are displayed of the four different syllable types (the /da/ (red) and /ga/ (green) of the unambiguous (UA) sounds and ambiguous (A) sound either perceived as /da/ (blue) or /ga/ (black)). ERP's are displayed for the average of the nine used electrodes (left), the vertical electrooculogram (VEOG; right top), and horizontal electrooculogram (HEOG; right bottom). Shaded areas indicate the standard error of the mean. No significant differences were found for the ERP.

Materials and Methods

Experiment 1

Participants: a total of 20 participants (9 male; age range: 18-32 years, mean age: 25) participated in the study. All participants reported to have normal hearing and normal or corrected to normal vision. The participants were native Dutch speakers. All gave written informed consent prior to the study. The study was approved by the local ethical committee at the Faculty of Psychology and Neuroscience at Maastricht University. Participants received monetary compensation for participating.

Stimuli: stimuli were created by morphing a recorded auditory /da/ to an auditory /ga/ using the synthesis function of the program Praat (Boersma & Weenink, 2013), similar to Bertelson et al. (Bertelson, Vroomen, & de Gelder, 2003). First, the sound was resampled to 11 kHz. To extract the different formants a linear predictor was created using 10 linear-prediction parameters (which would extract up to maximally 5 formants), using a sliding moving window of 25 ms, estimating the parameters every 5 milliseconds. To shift the recorded /da/ to /ga/, we moved the third formant from the original frequency band of 3 kHz down to 2.6 kHz in steps of -19 Mel (Bertelson, et al., 2003), resulting in nine stimuli morphed from /da/ to /ga/ (85 dB). Stimuli were presented via ER-30 insert earphones (Etymotic Research Inc., Elk Grove Village, IL, USA).

Procedure: first, we assessed individual psychometric curves by measuring the exact point at which participants reported to hear a /da/ or /ga/ at 50% of the trials, respectively. To this end, the nine stimuli on the da-ga spectrum were presented in random order while participants indicated whether they heard /da/ or /ga/. The inter stimulus interval (ISI) was 1.6, 1.8, or 2 sec, which presentation order was random. In total 135 trials were presented. A cumulative Gaussian was fitted to the data using the fitting toolbox modelfree v 1.1. (Zchaluk & Foster, 2009), implemented in MATLAB (mathworks). In the second part of this study, these individually calibrated parameters were used for EEG measurements. Here, only the three stimuli from the middle of the

individual psychometric curve as well as the two extremes were presented. Two out of twenty participants had a shifted psychometric curve, such that the point at which they detected /da/ 50% of the time was at the highest stimulus number used (stimulus 9) and were therefore excluded. Again participants had to indicate whether they heard /da/ or /ga/. The ISI was 3, 3.2, and 3.4 sec, which presentation order was random. For the two extremes a total of 80 trials and for the ambiguous stimuli a total of 120 trials were presented, divided in two blocks. During both tasks participants had to fixate to a white cross presented on a black background. Presentation software was used for stimulus delivery (Neurobehavioral Systems, Inc., Albany, NY).

EEG acquisition and preprocessing: EEG data were recorded with online filters of 0.1-200 Hz and a sampling rate of 500 Hz. The BrainAmp MR Plus EEG amplifier (BrainProducts GmBh, Munich, Germany) was used as amplifier, and BrainVision Recorder (BrainProducts, GmBh, Munich, Germany) for recording. Nine central Ag-AgCl electrodes were positioned on the head (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). The ground electrode was AFz, and the reference the tip of the nose. Four additional electrodes for eye movements were used (lateral to both eyes, below and above the left eye). All electrodes were positioned using Ten20 paste (Weaver, Aurora, USA). Impedance was kept below 15 kOhm (5 kOhm for ground and reference). Only nine electrodes were used as the experiment was part of a more extended set-up. However, only the EEG data is shown here.

All data processing was done with Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and the circular toolbox (Berens, 2009). Data was epoched from -3-3 around stimulus onset. Then, bad channels were removed (and replaced with the average of the other remaining recording channels). For most participants no channels were replaced (for 14 participants no channels were replaced, for 3 participants 1 channel was replaced, and for 1 participant 2 channels). Eye blinks were removed using the function scrls_regression of the eeglab plugin AAR [(Gómez-Herrero et al., 2006) filter order: 3, forgetting factor: 0.999, sigma: 0.01,

precision: 50], after which data was resampled to 200 Hz. Finally, with visual inspection trials with extreme variance were removed.

Data analyses

Consistency of phase differences: the complex Fourier spectra of individual epochs were extracted with Morlet wavelets between 2 and 10 Hz (with the amount of cycles used linearly increasing from 1.4 at 2 Hz to 7 at 10 Hz), stepsize of 0.5 Hz (time points of interest -0.3 to 0.2 seconds in steps of 0.01 sec), after which the phases were obtained. To ensure that post-stimulus effects did not temporally smear back to the pre-stimulus interval we padded all data points after zero with the amplitude value at zero. To investigate whether found effects were not due to this data padding we repeated the same analysis with the original data set. Then, we calculated for each individual the mean phase for each condition collapsed over channels. We had four conditions: 1) /Da/ of the unambiguous sounds (stimulus number 1; only trials that were identified as /da/ were included), 2) /ga/ of the non-ambiguous sounds (stimulus number 9; only trials that were identified as /ga/ were included), 3) ambiguous syllables that were identified as /da/ (on average 47.2 trials (SD = 18.5) per participant), and 4) ambiguous syllables that were identified as /ga/ (on average 69.9 trials (SD = 17.3) per participant). To ensure that effects were not due to difference in trial amounts we randomly drew trials of the three conditions with highest trial amount. The drawn amount corresponded to the trial amount of the condition with the lowest trial amount. We repeated this procedure 100 times, and took the average of this as our final phase estimate. To test whether the phase differences between /da/ and /ga/ were concentrated at a specific phase before sound onset we calculated the phase difference between /da/ and /ga/ for each time and frequency point for both ambiguous and unambiguous sounds. We expected that for the ambiguous sounds the pre-stimulus phase difference should correspond to the 80 ms difference found in the previous study. For the unambiguous sounds a less strong pre-stimulus effects was predicted since participants are unaware of the identity of any upcoming stimulus and therefore phase fluctuations should be random. Note that we did exclude trials were participants

indicated /ga/ for stimulus number 1 (mean amount of trials removed 2, range: 0-7 trials) and /da/ for stimulus number 9 (mean amount of trials removed 2.78, range: 0-13 trials), so this could cause small pre-stimulus phase differences. To test whether there was indeed such a difference in phase concentration dependent on condition we calculated the mean resultant vector length (MRVL) of the phase difference between /da/ and /ga/ for both conditions over the participants. The MRVL varies between 0 and 1 and the higher the value the more consistent the phase difference over participants. We used Monte Carlo simulations implemented in Fieldtrip to statically test that the difference between the MRVLs was not due to random fluctuations (two-sided test with 1,000 repetitions). Data was corrected for multiple comparisons using cluster based correction implemented in Fieldtrip (parameters: cluster alpha of 0.05, using the maximal sum of all the time and channel bins as dependent variable. P-values reflect two-sided p-values in all reported analysis).

80ms phase differences: for all the time and frequency points at which a significant difference in MRVL was found we investigated whether the mean phase indeed corresponded to the 80 ms difference that was expected. Therefore, we performed for both conditions V-tests for non-uniformity with a specified mean direction [corresponding to the 80 ms difference (Zar, 1998)] again using the average phase per participant per syllable averaged over channels. This test investigates whether the phase difference between /da/ and /ga/ is non-uniformly distributed around 80 ms phase difference. The performed tests were corrected for multiple comparisons via False Discovery Rate [FDR; (Benjamini & Yekutieli, 2001)].

Consistency of individual syllable types :we were interested whether the mean phase of individual syllable types had a strong phase consistency at similar frequency and time points. This analysis would indicate that not only the phase difference is consistent over participants, but also the exact phase at which perception is biased towards /da/ or /ga/. We again calculated the MRVL, but this time for each syllable type (again using the mean phase per participant and syllable type). Then, we took the

minimum MRVL of /da/ and /ga/ for each time and frequency point for the ambiguous and unambiguous sounds separately. To statistically test the strength of the coherence we permuted the labels of the /da/ and /ga/ for each condition separately in Monte Carlo simulations (1,000 repetitions). This analysis provides a map at which time and frequency point it is unlikely that such a high minimum MRVL is found, and thus reflects data points with high MRVL for both /da/ and /ga/. Data was corrected for multiple comparisons using the same cluster-size based correction as above.

ERP analysis: we were interested whether there any differences between syllable types in the post-stimulus event related potentials (ERP) and whether eye channels could explain the pre-stimulus effects. To calculate the ERP we first bandpass filtered the data between 0.5 and 20 Hz using a second order Butterworth filter. Then we averaged the data of the four syllable types for each participant. To calculate the VEOG we took the mean of the eye channel below and above the eye (multiplying the channel above the eye with -1). The same was done for the HEOG (using the right and left eye channel. Here the right eye channel was multiplied with -1). We tested significant differences between /da/ and /ga/ for both the ambiguous and nonambiguous condition separately using paired samples t-tests. Data were corrected for multiple comparisons using FDR.

Consistency of individual syllable types – individual participants: we performed the same analysis as in the section "Consistency of individual syllable types" for each individual participant. Instead of using the average phase, we took the individual phases of all trials (averaged over channels). This analysis would inform us whether for individual participants phases are consistently centered at one phase value. First, we calculated the inter trial coherence (ITC; which is mathematically equivalent to the MRVL but is the common used terminology for consistency over trials instead of constancy over average phases) for both /da/ and /ga/ (only for the unambiguous trials as we already showed that the effect is weaker for the ambiguous sounds). Then we again took the

minimum ITC as our test statistic. The strength of these minima were against statistically tested using Monte Carlo simulations (1,000 repetitions) using cluster based corrections for multiple comparisons. For this analysis (and the one in the next section) we did not randomly drew trials when there were unequal trial amounts as for some participants this would result in very little trials to estimate the ITC. However, it is shown that if the variance in the two groups to be compared do not vary considerably the chance of a type I error does not change much (Mewhort, Kelly, & Johns, 2009). The variance difference between the conditions was extremely low (the ratio being on average 1:0.96 with a range of 1:0.94 to 1:1.04).

Phase differences for individual participants: to test whether for individual participants there was a significant phase difference between perceived /da/ and /ga/ we performed the Watson's U2 test for equal means to test the phase difference between /da/ and /ga/ for all time and frequency points for the phases of individual trials for each participant. Monte Carlo simulations and cluster based correction for multiple comparisons were used for statistical testing.

Experiment 2

Participants: a total of 14 participants (one author; 4 male; age range: 18-31 years, mean age: 22.2) participated in the study. All participants reported to have normal hearing and normal or corrected to normal vision. They all gave written informed consent prior to the study. The study was approved by the local ethical committee at the Faculty of Psychology and Neuroscience at Maastricht University. For the control studies we were able to re-invite 6 of the original participants, but also had to recruit 6 new participants (2 male; age range: 20-31, mean age: 25.0). Participants got a monetary compensation for participating. Two of the original 14 participants were excluded from analysis since they had a ceiling effect in their performance, and only heard /da/ for almost all auditory stimuli.

Stimuli: the same syllable stimuli were used as in Experiment 1. Additionally, to entrain the system to sensory stimuli we presented white noise, band-pass filtered between 2.5 and 3.1 kHz [second order Butterworth; (Hari, Hämäläinen, & Joutsiniemi, 1989; Rees, Green, & Kay, 1986)]. This filter was chosen since it included the both the third formant frequencies of /da/ and /ga/, increasing the chance that the correct regions will be entrained as entrainment has been shown to be frequency specific (Lakatos et al., 2013). Stimuli were 50 ms long. Different entrainment trains lasted 2, 3 or 4 seconds at a presentation rate of 6.25 Hz. This rate was chosen since then 80 ms would correspond to exactly half a period. After the stimulus train finished the ambiguous syllable stimulus was presented at SOA's ranging from 0.1 to 0.58 sec in steps of 0.0267 sec (exactly fitting 2 cycles of 6.25 Hz). The stimuli were presented via headphones (stimuli in this experiment were presented around 60 dB). Three control tasks were implemented. In the first control experiment the presentation rate of the stimulus train was at 10 Hz instead of 6.25 Hz and the SOA's were ranging from 0.1 to 0.28 in steps of 0.017 sec. In the second control the presentation was 1 Hz. SOAs were ranging from 0.1 to 1.93 in steps of 0.167 sec. For this condition the entrainment trains lasted 4, 5 or 6 seconds to ensure that enough stimuli were presented to entrain the system. The third control was identical as the original experiment with the exception that the syllable stimuli were band-passed filtered between 2.5 and 3.1 kHz to only include the frequency range in which the two syllables differ. For this task the middle stimulus was used for each participant instead of individual tailored stimuli. The task of the participant was to indicate whether they believed the sound was of high or low frequency. At the start of each trial filtered syllable number 1 and 9 were presented to give a reference of what was 'high' and 'low'. The order of presentation of these stimuli was randomized.

Procedure: the same psychophysical procedure was used as in Experiment 1 to find the most ambiguous stimulus on the da-ga spectrum for individual participants. The stimulus closest to the 50th percent threshold was used for the main experiment. In the main experiment three blocks

were presented in which the trains of band-passed noise were presented. In total there were 324 trials (27 trials per SOA, 108 per entrainment length). All trials were presented at random order. During the experiment participants were required to fixate a white cross presented on a black background. Stimuli were again presented via Presentation software (Neurobs) and participants performed the experiment in a sound-shielded room. The three control experiments were presented after each other in random order. For the rest the procedure was the same as in the original experiment. For 7 control participants EEG was recorded at the same time (the stimuli were here presented through speakers instead of headphones).

Data analysis: first, the proportion of /da/ identification was calculated for every SOA, thereby creating a time course of proportion /da/ identification. We expected that this time course would follow an oscillatory pattern at 6.25 Hz. Hence, we fitted a sinus at a fixed frequency to the averaged behavioral time course using the function lsqnonlin in MATLAB. The test statistic calculated was the so-called relevance. This statistic is calculated by multiplying the explained variance of the model by the total variance of predicted values; thereby also endorsing predicted models that have a high variance instead of relatively flat lines. To check the likelihood of finding a bigger relevance value than the one in our own dataset we bootstrapped the labels of the phase bins to create randomized time courses and relevance values (n =10,000). We performed bootstrapping on the average instead of the individual time courses due to a lack of power in the individual data. For the three control analysis the same analysis was performed.

Data Analysis – EEG: all recording parameters and preprocessing steps were the same as in Experiment 1, with the exception that data was recorded from the full scalp using 31 electrodes, using the easycapM22 set-up (excluding electrode, TP9 and 10, but including channels C1, C3, and CPz). Data was epoched to -3 to 3 around entrainment train offset. Then, we extracted the ITC by extracting the angle from the complex

Fourier transform calculated via Morlet Wavelets (4 cycles included for the estimate) for each of the three control conditions (1 Hz, 10 Hz, and the frequency control).

Supplementary references:

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics,* 1165-1188.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software, 31*(10), 1-21.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification a mcgurk aftereffect. *Psychological Science, 14*(6), 592-597.

Boersma, P., & Weenink, D. (2013). *Praat: a system for doing phonectics by computer* (Version 5.3.56).

Gómez-Herrero, G., De Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Van Huffel, S., et al. (2006). *Automatic removal of ocular artifacts in the EEG without an EOG reference channel.* Paper presented at the Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic.

Hari, R., Hämäläinen, M., & Joutsiniemi, S. L. (1989). Neuromagnetic steady-state responses to auditory stimuli. *The Journal of the Acoustical Society of America, 86*(3), 1033-1039.

Lakatos, P., Musacchia, G., O'Connell, M., Falchier, A., Javitt, D., & Schroeder, C. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*(4), 750-761.

Mewhort, D., Kelly, M., & Johns, B. T. (2009). Randomization tests and the unequal-N/unequal-variance problem. *Behavior research methods, 41*(3), 664-667.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011,* 1.

Rees, A., Green, G., & Kay, R. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hearing research, 23*(2), 123-133.

Zar, J. H. (1998). *Biostatistical Analysis* (4 ed.). Englewood Cliffs, New Jersey: Prentice Hall.

Zchaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics, 71*(6), 1414-1425.

CHAPTER 8

OSCILLATORY PHASE SHAPES SYLLABLE REPRESENTATIONS

Corresponding Manuscript:
Ten Oever, S., Hausfeld, L., Correia, J.M., Van Atteveldt, N., Formisano, E. & Sack, A.T. (under revision at NeuroImage). A 7T fMRI study investigating the influence of oscillatory phase on syllable representations.

Abstract

Complex information is represented in the brain by a distributed pattern of activation from different neurons that each codes a subset of the features of the external input. At the same time, intrinsic features of the brain have been shown to influence stimulus categorization. For example, we have shown that the identification of an ambiguous syllable that can either be perceived as /da/ or /ga/ is biased by ongoing oscillatory phase. This suggests that phase is a cue for the brain to determine syllable identity and this cue could be an element of the representation of these syllables. If so, activation patterns for /da/ should be more unique when the syllable is presented at the /da/ biasing (i.e. its "preferred") phase as one additional /da/ feature is present in the "input". To test this hypothesis we presented *non*-ambiguous /da/ and /ga/ syllables at either their preferred or non-preferred phase (using sensory entrainment) while measuring 7T fMRI. Using multivariate pattern analysis we show that syllable decoding performance is higher when syllables are presented at their preferred compared to their non-preferred phase using activation patterns from auditory regions. These results suggest that phase information increases the distinctiveness of /da/ and /ga/ activation patterns and that sensory entrainment could be a method to investigate the effects of oscillatory phase on fMRI signals.

Introduction

The role of oscillatory phase for perception and cognition is becoming increasingly clear (Kayser et al., 2009; Fell and Axmacher, 2011). While many studies have focused on the role of phase for stimulus detection [e.g. (Fiebelkorn et al., 2013; Henry et al., 2014; Ten Oever et al., 2015)], oscillatory phase also influences the categorization of stimuli (Watrous et al., 2015; Ten Oever and Sack, in press). We have for example shown that ongoing oscillatory phase as measured with electroencephalography (EEG) determines whether an ambiguous syllable that can either be perceived as /da/ or /ga/ is identified as one or the other syllable (Ten Oever and Sack, in press). In the same study, we entrained oscillatory patterns in the brain to rhythmically presented sounds after which the same ambiguous syllable was presented at different onset delays (figure 1A). We found that depending on the delay (and thus the underlying phase) participants more likely identified the syllable as /da/ or /ga/ (figure 1B). This suggests that oscillatory phase is a cue for syllable identification and each syllable has one "preferred" phase.

Patterns of activation in the (auditory) cortex have been shown to reflect distributed representations of speech (Formisano et al., 2008; Staeren et al., 2009; Mesgarani and Chang, 2012; Tsunada and Cohen, 2014). These representations likely reflect the collective activation of numerous neurons active for different features that determine the external speech input. Generally, the more distinct two different types of input are (e.g. by having multiple cues that differentiate the inputs), the more distinct their activation patterns [or representations; (Hausfeld et al., 2014)]. As oscillatory phase is a cue for syllable identification, it might also enhance the distinctiveness of the representation of a syllable. Accordingly, we would predict that when both /da/ and /ga/ are presented at their preferred phase, their activation pattern is more distinct compared to when both syllables are presented at their non-preferred phase.

Multi-variate pattern analysis (MVPA) in functional magnetic resonance imaging (fMRI) has been used successfully to discriminate between distributed speech representations (Formisano et al., 2008; Kilian-Hütten et al., 2011), and seems to be more sensitive than classical univariate approaches to dissociate these distributed patterns of activation

(Haxby et al., 2001; Haynes and Rees, 2006). We used this method to investigate whether discrimination performance between /da/ and /ga/ would be better when both syllables were presented at their preferred compared to non-preferred phase. This would show that phase increases the distinctiveness of the representation or activation patterns of /da/ and /ga/.

We induced brain oscillations by repeatedly presenting auditory stimuli at a 6.25 Hz rate (i.e. auditory entrainment), similar as in our previous study (Ten Oever and Sack, in press). *Non*-ambiguous /da/ or /ga/ syllables were presented at differing delays after the entrainment finished, either corresponding to the syllable's preferred or non-preferred

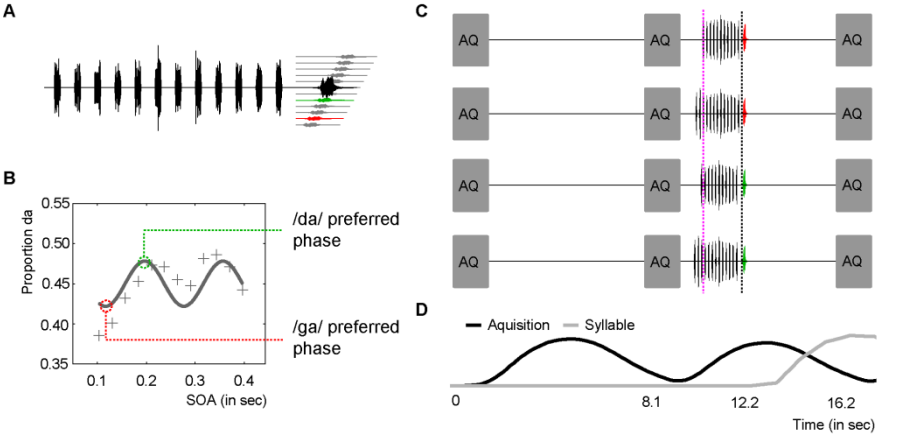


Figure 1. Previous results and stimulation protocol. A) Entrainment stimulus after which a syllable is presented at different intervals. B) The results from the previous study [adapted with permission from (Ten Oever and Sack, in press)]. The red and green SOAs represent the preferred phase for /ga/ and /da/ respectively. C) The stimuli were always presented in the silent gap after acquisition (AQ). Four different stimuli presentations of /ga/ are visualized in the figure: at a stimulus onset asynchrony (SOA) of 120 ms with an entrainment train of n = 11, an SOA of 120 ms with n = 13, an SOA of 200 ms with n = 11, and at an SOA of 200 ms with n = 13. While the syllable types (red and green lines) are always presented at 12.2 (or 12.28) seconds after the first acquisition (see black dotted line) the entrainment trains (black) start at different time points dependent on condition (see pink dotted line). D) The predicted BOLD response to the acquisition noise (black) and the syllable (grey) is displayed. Due to the long TR the response to the acquisition noise is reduced while estimating the peak BOLD response of the syllable.

phase (see red and green lines in figure 1). Then, with MVPA we calculated the accuracy of syllable identity discrimination. We found significantly better performance when both syllables were presented at their preferred phase compared to their non-preferred phase. These results show that syllable representations in auditory regions are processed at a preferred oscillatory phase and indicate the potential of fMRI to study oscillatory patterns.

Methods

Participants

Ten healthy native Dutch speakers participated in the study (4 male, age range: 26-32, mean age: 29.1). One participant was left-handed. The study was approved by the local ethical committee at Maastricht University. Participants gave written informed consent prior to participation and filled out the safety screening from the Scannexus MRI facilities at Maastricht University. Participants received monetary compensation for participating. One participant was excluded from the analysis as the full fMRI session was not completed.

Stimuli and experimental procedures

In each trial first an entrainment sequence was presented, which consisted of band-passed noise-bursts (2.5 kHz-3.1 kHz, 50 ms) at a presentation rate of 6.25 Hz. The entrainment sequences were 11, 12, or 13 stimuli long to reduce temporal expectations of the arrival time of the syllable. After the train finished, the sound of either the syllable /da/ or /ga/ was presented. The original syllable used was a /da/ pronounced by a Dutch female speaker, lasting approximately 300 ms. This syllable was then morphed into a /ga/ by changing the third formant frequency from a mean frequency of 3.0 kHz to 2.6 kHz using the program Praat (Boersma and Weenink, 2013; Ten Oever and Sack, in press). The syllables were presented after the entrainment sequence either 120 or 200 ms after the onset of the last noise burst (per run per condition 6 trials), which corresponded with the preferred phase of /ga/ and /da/ respectively (Ten Oever and Sack, in press). In another condition the middle /da/ -/ga/

morph was presented at either 120 or 200 ms (per run per condition 6 trials), but with reversed audio (data not shown). In four additional trials per run (condition type randomly selected) the last stimulus sequence had a wider filtered broadband noise (2.2 kHz-3.4 kHz). Participants were required to press a button when they heard this stimulus sequence. These trials were not analyzed.

Scanning parameters

MRI data was collected on a 7-tesla Siemens Magnetom scanner with a body gradient system with a whole brain coil at the Scannexus facilities, Maastricht, The Netherlands. Anatomical images were acquired via a T1 weighted MPRAGE sequence (TR = 3100 ms; TI = 1500 ms; TE = 2.25 ms; 0.6 mm isotropic) and a proton density (PD) weighted sequence with the same parameters (except the TR = 1440) not using the inversion module. This sequence was acquired to remove field inhomogeneities to improve image quality by dividing the T1 weighted image by the proton density weighted image (Van de Moortele et al., 2009). Five functional runs with 84 TRs were acquired for all participants. A blood oxygenation level-dependent (BOLD)-sensitive echo-imaging (EPI) sequence was used (matrix = 128*128; field of view = 192*192 mm^2; 66 slices; TR = 8100 ms; TE = 19 ms; acquisition time = 1.4 s resulting in a voxel size of 1.5*1.5*1.5mm^3) with a GRAPPA acceleration factor of 2 (Griswold et al., 2002). Moreover, two slices were acquired simultaneously via an interleaved multiband sequence to improve the speed of acquisition (Moeller et al., 2010). To correct for the direction of acquisition 2 EPI sequences of 5 TRs were collected using both the anterior-to-posterior and posterior-to-anterior direction [main functional runs were acquired using the anterior-to-posterior acquisition direction; (Bowtell et al., 1994; Jezzard and Balaban, 1995; Andersson et al., 2003)].

EPI sequences are inherently noisy, thereby challenging auditory research in the scanner. Even more troubling for the current paradigm is that the EPI sequence contains a strong rhythmic component as the separate images are acquired. To ensure that entrainment only occurs to our presented stream and not to the scanner noise we used a sparse sampling paradigm with a repetition time of 8100 ms. In this way we could position our stimuli in between two acquisitions such that our

stimulus of interest (the syllable) would be presented 4 or 3.92 s before and 4.1 or 4.02 s after the start of the image acquisition, thereby collecting the data around the peak of the BOLD response while decreasing the signal related to scanner noise (figure 1C and D). At the following acquisition interval no stimuli were presented to ensure that the signal would recover the baseline. As the syllable positioning was fixed at either 4000 or 3920 ms prior to acquisition, the onset of the entrainment train was slightly different, depending on the specific condition.

Data preprocessing

Data preprocessing was performed with BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands) and FSL (www.fmrib.ox.ac.uk). For anatomical data, the reconstructed MPRAGE T1 weighted images were divided by the images by the PD images to reduce inhomogeneities of the signal (Van de Moortele et al., 2009). An additional inhomogeneity correction was performed in Brainvoyager and the images were resampled to 0.5 mm isovoxel resolution and rotated to ACPC space. Then we performed automatic grey-white matter segmentation in FSL and manually adjusted the segmentation in Brainvoyager. A grey matter cortical mask of the temporal lobe was created to reduce the amount of voxels present in the multivariate pattern classification.

Functional images were motion corrected and temporal high-passed filtered using three cosine cycles and a linear trend regressor for each run separately. Slice acquisition timing was corrected with a sinc-weighted interpolation. In FSL we used the TOPUP function to correct for susceptibility induced distortions caused by the acquisition direction to improve alignment with the anatomical data. Then images were co-registered with the anatomical data.

Data analysis

Univariate analysis: Due to our long TR we only had 2 data points to model the BOLD response. Therefore, we estimated the activation patterns for each stimulus by calculating the proportion of signal change subtracting the activity of a single data point directly after the stimulus

from the data point before the stimulus (baseline) and further dividing by this baseline. We repeated this calculation for all the stimuli, providing us with an overall activation pattern for each participant that was later used for the MVPA analysis. To obtain a group map of activation we performed cortex based alignment of the surface maps (Goebel et al., 2006) and a random effect GLM using a step function as predictor for each sound condition (with conditions /da/ time point 120, /da/ time point 200, /ga/ time point 120, /ga/ time point 200, reverse, and control) and a separate predictor for each run.

MVPA analysis: All MVPA analyses were performed in ACPC space. Support vector machines (SVM) were used to decode the multivariate activation patterns. As a first analysis we tested whether we could reliably decode syllable identity irrespective on which time point on the entrainment the syllables were presented. Training data consistent of randomly picking 96 out of the 120 trials (48 per syllable); the remaining trails were used for testing. Ten cross-validations were performed (with replacements). Features consisted of the proportion of signal change as described above per voxel and per trial. Standardized z-scores were calculated over each run. Excessive amounts of features can harm classification performances due to overfitting and it is therefore necessary to have the optimal amount of features present in your classification (Norman et al., 2006). Therefore we repeated the classification using between 50 and 2500 most active voxels (overall activation) in 15 logarithmically spaced steps and extracted the best classification (Kilian-Hütten et al., 2011). Voxels with activation patterns stronger than 5 standard deviations of the mean were never included as this high activity pattern likely does not arise from neuronal activity, but more likely from bigger veins [see e.g. (Lee et al., 1995; Turner, 2002)]. To investigate whether the classification performance was above chance level we performed permutation tests. Syllable labels of both the training and testing were permuted 100 times and the exact same analysis was performed. Then we compared using a one-sided Wilcoxon signed rank test whether the original classification performance was higher than the average permuted labels for all participants. The same analysis was performed to classify the two time points irrespective of the identity of the syllable.

Our main hypothesis was that decoding performance would increase when syllables are presented at their "preferred" phase. Therefore, we split the dataset in two parts to perform separate classifications: 1) syllables presented at their "preferred" (/ga/ at 120 ms and /da/ at 200 ms) and 2) syllables presented at their "non-preferred" phase (/da/ at 120 ms and /ga/ at 200 ms). The rest of the analysis was the same as above except that model training (testing) was based on 48 (12) trials. As a final test to investigate whether one specific time point/phase would have a higher classification performance we repeated the analysis, but performed the /da/-/ga/ classification when both syllables were presented at 120 ms or when both were presented at 200 ms.

To investigate the spatial consistency of the voxels used for the classification we created group discriminate maps for the /da/-/ga/ classification when syllables were presented at their "preferred" phase. These maps represent the shared cortical locations that contributed to the discrimination of the syllables. We created these maps in two different ways. First, we created a map using all the voxels that went into the final optimized classification. This map shows the overlap over participants of all the voxels that were used for classification; however it does not dissociate which voxels contributed more to the discrimination. Moreover, the voxel size varies over participants. Therefore, we created a second map in which only the 150 most discriminating voxels where included. 150 voxels were chosen as it corresponded with the amount of voxels used for the classification of the participant with the least voxels in the initial feature selection. All maps were then transformed to the surface representation of one participant after cortex based alignment (Goebel et al., 2006).

Results

Univariate analysis

The activation maps of syllable presentation versus baseline showed a bilateral network of activation mainly in primary auditory and auditory association cortex. Additionally, parts of the right cingulate motor areas were active (figure 2). This indicates that our sparse sampling method was successful in eliciting reliable brain responses to spoken syllables. Other

areas known to be activated by syllables or phonemes (Liebenthal et al., 2005; Hickok and Poeppel, 2007; Desai et al., 2008), such as inferior frontal cortex and insula, showed responses when using a more liberal threshold. Any direct contrast between /da/ or /ga/ or between the two time points did not result in any significant difference [as estimated using false discovery rate (Benjamini and Yekutieli, 2001)].

Multivariate analysis

/Da/ versus /ga/: In a first step we wanted to replicate the finding that syllable identity can be decoded from fMRI BOLD patterns (Formisano et al., 2008). Moreover, this finding would show that our paradigm of sparse sampling can be successfully used to perform classification. We found a mean classification accuracy of 0.578 (figure 3, left panel). Statistical testing of this performance by permuting the labels of the syllables indicated that the accuracies of the original labels was higher than the performance for permuted labels (Z = 38; p = 0.038). Note that the empirical chance level for classification is higher as 0.5 as we optimized the amount of voxels selected for the classification by using the classification with the best performance. However, permutation testing controls for this enhanced empirical chance level (Moeller et al., 2010).

Time point comparison: In a second step we classified the two time points (120 vs 200 ms) irrespective of the syllable identity. Overall classification

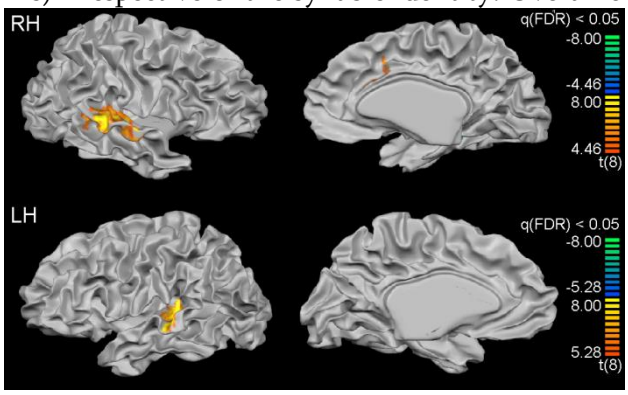


Figure 2. Univariate results. Overall activation map of all presented syllables versus baseline as measured with a random effect GLM. Results are presented on the surface of one participant after cortex based alignment to this participant.

performance was 0.574 which was higher compared to classification performance with permuted labels (Z = 32; p = 0.049).

/Da/ verus /ga/ at the preferred versus non-preferred phase: Our main expectation was that classification performance should be higher when syllables are presented at their preferred phase. This analysis is orthogonal to the previous two analyses as each syllable is presented in equal amounts at both phases. We split the data in two and repeated the /da/ versus /ga/ classification either when the syllables were both presented at their preferred or at their non-preferred phase. We found a higher /da/-/ga/ classification performance (0.605) for the preferred compared to the non-preferred phase (0.559; figure 4A, top panel; Z = 41; p = 0.016). Moreover, we found that only for the preferred phase the classification performance was higher as for the permuted labels (Z = 38; p = 0.037 and Z = 20; p = 0.410 for preferred and non-preferred phase, respectively).

/Da/ versus /ga/ at 120 or 200 ms: To test whether there was one specific phase that increased classification accuracy, we repeated the previous analysis, but testing /da/-/ga/ classification when both syllables were presented at 120 ms or both presented at 200 ms. There was a classification performance of 0.578 and 0.550 when both syllables were presented at 120 ms or 200 ms, respectively. The classification performance did not significantly differ (Z = 28; p = 0.563; two-sided). Additionally, both classifications were not significantly different from the

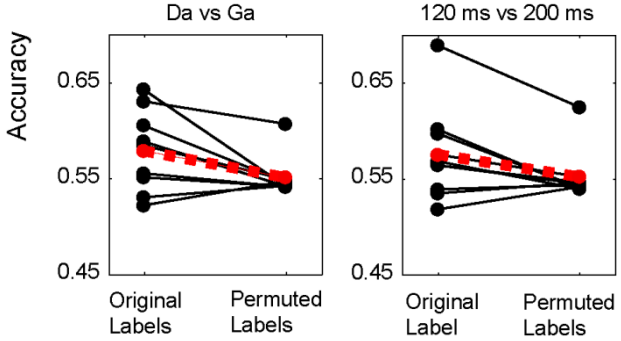


Figure 3. Classification performance. Classification performance for each participant (black line) and the average (red dotted line) for the both original labels and permuted labels for the contrast /da/ versus /ga/ (left panel) and 120 versus 200 ms SOA (right panel).

permuted labels (Z = 29; p = 0.248 and Z = 26; p = 0.367 for 120 and 200 ms, respectively).

Finally, we wanted to calculate the interaction between the factors /ga/ phase and /da/ phase. To do so with a non-parametric test we performed a singed rank test between two difference scores: 1) the difference in /da/-/ga/ classification performance between the preferred and non-preferred phase and 2) the difference in /da/-/ga/ classification

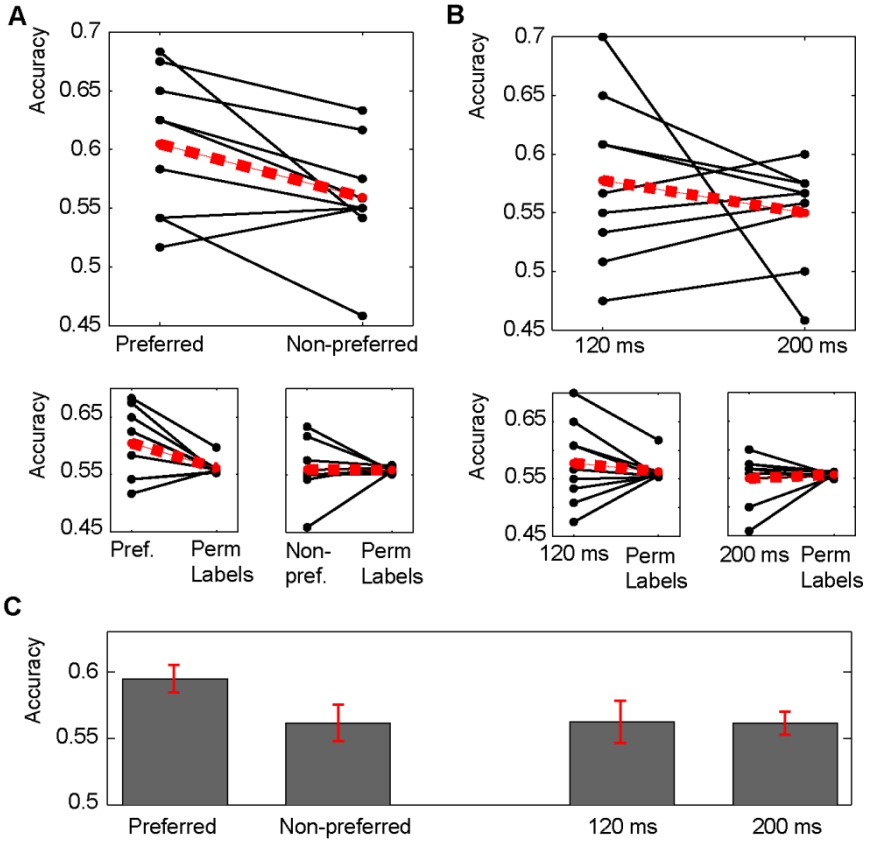


Figure 4. Classification performance main contrasts. A) /da/-/ga/ classification performance for each participant (black line) and the average (red dotted line) for the preferred and non-preferred phase (top panel). Bottom two panels reflect the comparisons with the actual labels of the preferred phases (left) and non-preferred phases (right) and their respective permuted labels. B). /da/-/ga/ classification performance when syllables where presented at the early or late time point. C) The average classification performance excluding one outlier participant. Error bars represent the within subject standard error of the mean.

performance between 120 and 200 ms. Initially, there was no significant effect (Z = 32; p = 0.15). However, one participant had an extreme value in the 120-200 ms classification difference that was more than two standard deviations from the average (see figure 4B. One participant has a difference of almost 0.25). When removing this participant, the interaction was significant (Z = 32; p = 0.027; see figure 4C).

Discriminative maps: Figure 5A shows the overlap of all the voxels that were used for the "preferred" phase classification. These were the voxels having the highest percent signal change. The amount of voxels was dependent on the participant as it was individually tailored to

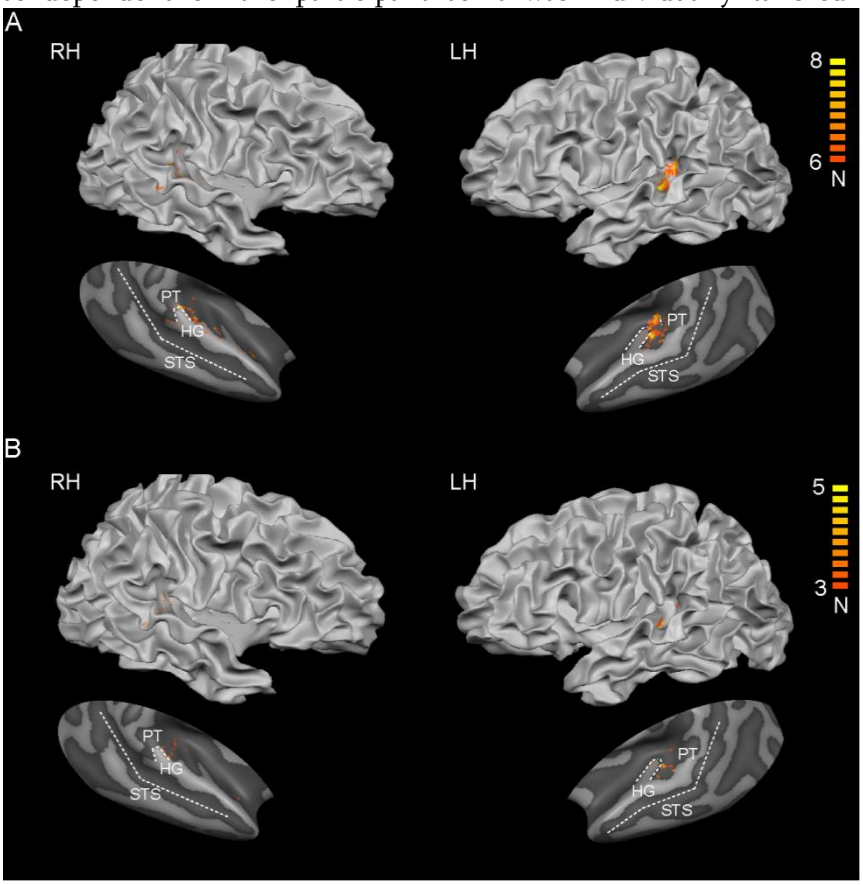


Figure 5. Discriminative maps. A) The spatial overlap over participants of all voxels used in the optimized classification. Color indicates the amount of participants. B) The spatial overlap over participants when only using the 150 most discriminative voxels. The overlap is highly reduced.

optimize the classification. It is clear that only a small proportion of voxels had overlap for more than 6 out of 9 participants, mostly overlapping around the left Heschl's sulcus (HS) and in the right hemisphere anterior of Heschl's gyrus (HG) at the first transverse sulcus (FTS). This overlap was reduced when looking at the 150 most discriminative voxels (figure 5B). Voxels only had showed overlap for 3 to 5 participants. Supplementary figure 1 shows for each individual the 150 most discriminative voxels that resulted in the highest classification performance. Although most participants have their most discriminative voxels around the main auditory regions including bilateral superior temporal gyrus (STG), FTS, planum temporale (PT), HG, and HS, the exact distribution varied across participants.

Discussion

In the current study, we investigated whether oscillatory phase information changes the distinctiveness of syllable representations as measured with fMRI. This study was based our previous results showing that syllable identification of an ambiguous stimulus (either perceived as /da/ or /ga/) is biased when it is presented at a specific oscillatory phase (Ten Oever et al., 2013; Ten Oever and Sack, in press). We used the same sensory entrainment paradigm and presented /da/ and /ga/ stimuli either at their "preferred" or "non-preferred" phase to investigate whether phase information would change the fMRI activation patterns to these syllables. As we hypothesized, we found that /da/-/ga/ classification (with MVPA) was more accurate when both syllables were presented at their preferred compared to non-preferred phase. These results verify that syllable processing is phase dependent and show that this information can be extracted even with slow fluctuating BOLD responses.

Syllable dependent phase processing

Phase coding has been a proposed as a mechanism to represent information in the brain (Fries, 2005; Jensen et al., 2014; Watrous et al., 2015). Different electrophysiological studies have shown that adding phase information to classifiers aid classification performance (Kayser et

al., 2009; Lopour et al., 2013). Moreover, neuronal populations coding for similar representations seem to communicate with each other by synchronizing their firing rates to a specific phase (O'Keefe and Recce, 1993; Fries, 2005; Lisman and Jensen, 2013). Since specific syllable representations prefer specific oscillatory phases (Ten Oever and Sack, in press) neuronal populations coding for one syllable might become active when another syllable is presented at their preferred phase. This could be reflected in for example /da/ sensitive neurons being active if a /ga/ syllable is presented at a /da/ preferred phase. Alternatively, /da/ sensitive neuronal populations might have more robust processing for specific phases. In either way, syllable representations are more distinctive from each other when syllables are presented at their preferred phase.

Methodological considerations

The noisy scanner environment makes any type of auditory experiment difficult (Cho et al., 1997; Griswold et al., 2002). In the current design we choose to overcome this problem by having a very long repetition time and only trying to sample the peak of the BOLD response. This of course has its drawbacks as the data amount is little and, if participants have a particularly slow or fast BOLD response, it might not sample the optimal point. However, we were still able to find normal activation patterns to auditory stimuli and above chance level classification performance. The increased signal-to-noise ratio of 7T MRI might have helped to increase the overall activation levels. Moreover, the BOLD response to the scanner noise that normally accompanies the BOLD response to the auditory target stimuli might be significantly reduced (Bandettini et al., 1998; Talavage et al., 1999). This shows the added value of high field fMRI and the feasibility of this type of silent paradigms [see also (Amaro et al., 2002; Zaehle et al., 2007)]. For most experiments this type of sampling is not necessary, but if the rhythmic auditory patterns of the scanner noise are too intrusive for the specific experimental set-up, the proposed acquisition scheme represents one option to overcome this limitation of fMRI.

Spatial overlap

Our discriminative maps (figure 5 and supplementary figure 1) indicated a limited spatial overlap of the voxels used for the syllable classification. The only area that showed clear overlap when using all the voxels used for optimal classification was left Helschl's sulcus bordering the Planum Temporale (PT). Left Helschl's sulcus is involved in the primary auditory analysis and largely part of the belt area (Moerel et al., 2014). It is normally sensitive to a broader tuning width of sounds (Rauschecker et al., 1995; Hackett et al., 1998; Moerel et al., 2013) and the most lateral part of the Helschl's sulcus also seems speech/voice sensitive (Belin et al., 2000; Moerel et al., 2014). In contrast, PT is viewed as a computational hub in which complex spectrotemporal inputs are matched to stored memories of auditory objects (Griffiths and Warren, 2002). It has been shown that this area plays an important role discriminating the perceived identity of an ambiguous syllable (Kilian-Hütten et al., 2011). However, in the study of Kilian-Hütten and colleagues (2011) the discriminative voxels extended more to the posterior part of PT, including sensory-motor integration areas (Hickok and Poeppel, 2007), while our voxels are located more anteriorly. In sum, it seems that the most discriminate areas in our study include areas that perform a higher order acoustic transformation linking the acoustic input to stored auditory categories (Obleser and Eisner, 2009). On a critical note, it could be that the high activation of these broadly tuned areas is partly induced by the broadband noise used in the entrainment. Moreover, individual discriminative maps are much more diverse and include more widespread areas, covering almost all auditory and auditory association areas.

Conclusion

In this study, we showed that oscillatory phase contributes to the distinctiveness of the representation of /da/ and /ga/. These results add to a growing literature showing the role of oscillatory phase in perception and cognition (Lakatos et al., 2008; Cravo et al., 2011; Peelle and Davis, 2012; Jensen et al., 2014). Furthermore, it indicates that intrinsic properties in the brain might be an essential part of the representation of categorical information. FMRI has thus far not been used to investigate

influences of oscillatory phase. We are one of the first to demonstrate that slow fluctuating BOLD patterns are indeed sensitive to this type temporal manipulation. This opens a new way to investigate phase, using the high spatial resolution that fMRI provides. This is important, as the phase coding mechanism that we demonstrate might be a unique strategy of the brain to memorize and organize perceptual input and future studies should aim to unravel the principles of this mechanism.

References

Amaro E, Williams S.C., Shergill S.S., Fu C.H., MacSweeney M., Picchioni M.M., Brammer M.J., & McGuire P.K. (2002). Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. *Journal of magnetic resonance imaging, 16*(5), 497-510.

Andersson J.L., Skare S., & Ashburner J. (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage, 20*(2), 870-888.

Bandettini P.A., Jesmanowicz A., Van Kylen J., Birn R.M., & Hyde J.S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magnetic Resonance in Medicine, 39(3),* 410-416.

Belin P., Zatorre R.J., Lafaille P., Ahad P., & Pike B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benjamini Y., & Yekutieli D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics,* 1165-1188.

Boersma P., & Weenink D. (2013). Praat: a system for doing phonectics by computer (Version 5.3.56).

Bowtell R., McIntyre D., Commandre M., Glover P., & Mansfield P (1994). *Correction of geometric distortion in echo planar images.* Paper presented at the Soc. Magn. Res. Abstr, p 411.

Cho Z., Park S., Kim J., Chung S., Chung S., Chung J., Moon C., Yi J., Sin C., & Wong E. (1997). Analysis of acoustic noise in MRI. *Magnetic resonance imaging, 15*(7), 815-822.

Cravo A.M., Rohenkohl G., Wyart V., & Nobre A.C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *Journal of Neurophysiology, 106*(6), 2964-2972.

Desai R., Liebenthal E., Waldron E., & Binder J.R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience, 20*(7), 1174-1188.

Fell J., & Axmacher N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience, 12*(2), 105-118.

Fiebelkorn I.C., Snyder A.C., Mercier M.R., Butler J.S., Molholm S., & Foxe J.J. (2013). Cortical cross-frequency coupling predicts perceptual outcomes. *Neuroimage, 69,* 126-137.

Formisano E., De Martino F., Bonte M., & Goebel R. (2008). "Who" Is Saying" What"? Brain-Based Decoding of Human Voice and Speech. *Science 322*(5903), 970-973.

Fries P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9*(10), 474-480.

Goebel R., Esposito F., & Formisano E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically

aligned group general linear model analysis and self-organizing group independent component analysis. *Human brain mapping, 27*(2), 392-401.

Griffiths T.D., & Warren J.D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25*(7), 348-353.

Griswold M.A., Jakob P.M., Heidemann R.M., Nittka M., Jellus V., Wang J., Kiefer B., & Haase A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic resonance in medicine, 47*(6), 1202-1210.

Hackett T., Stepniewska I., & Kaas J. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *Journal of Comparative Neurology, 394*(4), 475-495.

Hausfeld L., Valente G., & Formisano E. (2014). Multiclass fMRI data decoding and visualization using supervised self-organizing maps. *NeuroImage, 96,* 54-66.

Haxby J.V., Gobbini M.I., Furey M.L., Ishai A., Schouten J.L., & Pietrini P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haynes J.D., & Rees G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523-534.

Henry M.J., Herrmann B., Obleser J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences, 111*(41), 14935-14940.

Hickok G., Poeppel D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Jensen O., Gips B., Bergmann T.O., & Bonnefond M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences, 37*(7), 357-369.

Jezzard P., & Balaban R.S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic resonance in medicine, 34*(1), 65-73.

Kayser C., Montemurro M.A., Logothetis N.K., & Panzeri S. (2009.) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron, 61*(4), 597-608.

Kilian-Hütten N., Valente G., Vroomen J., & Formisano E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience, 31*(5), 1715-1720.

Lakatos P., Karmos G., Mehta A.D., Ulbert I., Schroeder C.E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Lee A.T., Glover G.H., & Meyer C.H. (1995). Discrimination of Large Venous Vessels in Time-Course Spiral Blood-Oxygen-Level-Dependent Magnetic-Resonance Functional Neuroimaging. *Magnetic resonance in medicine, 33*(6), 745-754.

Liebenthal E., Binder J.R., Spitzer S.M., Possing E.T., & Medler D.A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex,15*(10), 1621-1631.

Lisman J.E., & Jensen O. (2013). The theta-gamma neural code. *Neuron, 77*(6), 1002-1016.

Lopour B.A., Tavassoli A., Fried I., & Ringach D.L. (2013). Coding of Information in the phase of local field potentials within human medial temporal lobe. *Neuron, 79*(3), 594-606.

Mesgarani N., & Chang E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*(7397), 233-236.

Moeller S., Yacoub E., Olman C.A., Auerbach E., Strupp J., Harel N., & Uğurbil K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine, 63*(5), 1144-1153.

Moerel M., De Martino F., & Formisano E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in neuroscience, 8.*

Moerel M., De Martino F., Santoro R., Ugurbil K., Goebel R., Yacoub E., & Formisano E. (2013). Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *The Journal of Neuroscience, 33*(29), 11888-11898.

Norman K.A., Polyn S.M., Detre G.J., & Haxby J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-430.

O'Keefe J., & Recce M.L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*(3), 317-330.

Obleser J., & Eisner F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences, 13*(1), 14-19.

Peelle J.E., & Davis M.H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3.*

Rauschecker J.P., Tian B., & Hauser M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science, 268*(5207), 111-114.

Staeren N., Renvall H., De Martino F., Goebel R., & Formisano E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology, 19*(6), 498-502.

Talavage T.M., Edmister W.B., Ledden P.J., & Weisskoff R.M. (1999). Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Human brain mapping, 7*(2), 79-88.

Ten Oever S., Sack A.T. (in press). Oscillatory phase shapes syllable perception. Proceedings of the National Academy of Sciences.

Ten Oever S., Van Atteveldt N., & Sack A.T. (2015). Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. *Journal of Cognitive Neuroscience, 27*(9), 1811-1822.

Ten Oever S., Sack A., Wheat K.L., Bien N., & Van Atteveldt N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology, 4.*

Tsunada J., & Cohen Y.E. (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. *Frontiers in neuroscience, 8.*

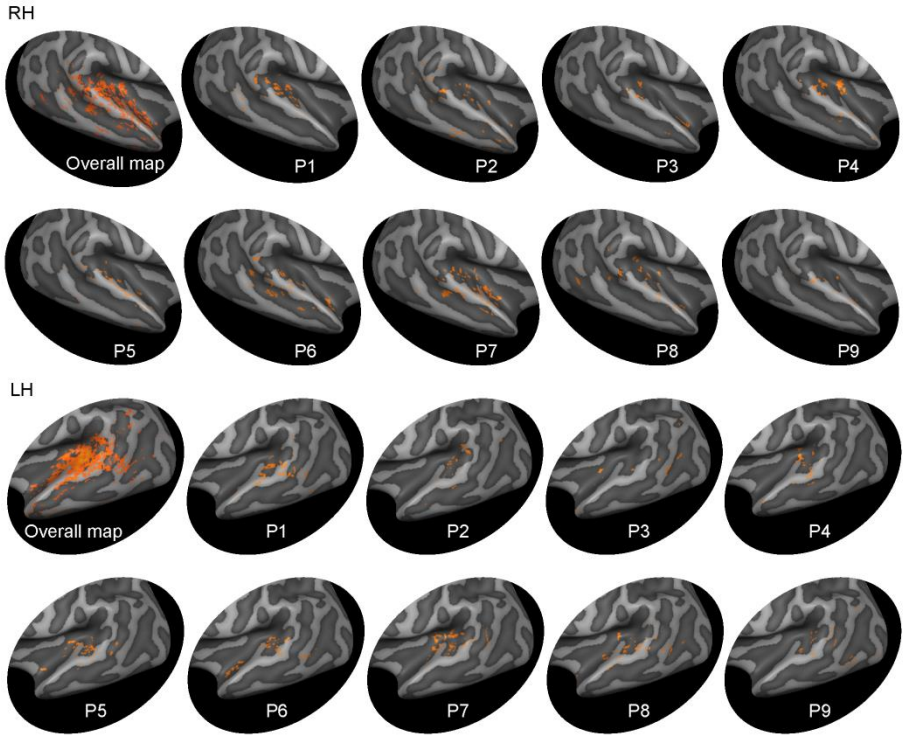
Turner R. (2002). How much cortex can a vein drain? Downstream dilution of activation-related cerebral blood oxygenation changes. *NeuroImage, 16*(4), 1062-1067.

Van de Moortele P.F., Auerbach E.J., Olman C., Yacoub E., Uğurbil K., & Moeller S. (2009). T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *Neuroimage, 46*(2), 432-446.

Watrous A.J., Fell J., Ekstrom A.D., & Axmacher N. (2015). More than spikes: common oscillatory mechanisms for content specific neural representations during perception and memory. *Current opinion in neurobiology, 31,* 33-39.

Zaehle T., Schmidt C.F., Meyer M., Baumann S., Baltes C., Boesiger P., & Jancke L. (2007). Comparison of "silent" clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *Neuroimage, 37*(4), 1195-1204.

Supporting Information



Supplementary figure 1: Individual discriminative maps. The top 150 discriminative voxels for each participant are shown for the right and left hemisphere. The map labeled "overall map" shows all the top 150 discriminative voxels of all participants combined on one map. All voxels are plotted on the surface map of one representative participant after cortex based aligned.

Supplementary table 1.

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Accuracy	0.675	0.542	0.683	0.542	0.517	0.583	0.625	0.625	0.65
Amount of voxels	618	1891	50	153	1891	50	267	2500	116

Individual maximum accuracy levels with corresponding amount of voxels.

CHAPTER 9

SUMMARY AND DISCUSSION

The work in this thesis focuses on the influence of temporal associations between audio-visual stimulus pairs on perception. In the first part of the thesis the perceptual benefit of temporal associations acquired during one experimental session either through rhythmicity or temporal cueing is explored. In the second part the temporal association between the onset of mouth movements and onset of speech sounds and the consequences for behavior and syllable coding are central.

Part I: Short term temporal statistics

Temporal regularities in the environment have been shown to have behavioral benefits both decreasing reaction times through temporal preparation (Correa, Lupiáñez, Milliken, & Tudela, 2004; Coull & Nobre, 1998; Los & Van der Burg, 2013; Niemi & Näätänen, 1981) as well as improving perception in discrimination (Ellis & Jones, 2010; Jones, Moynihan, MacKenzie, & Puente, 2002; Mathewson, Fabiani, Gratton, Beck, & Lleras, 2010) and detection tasks (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Rohenkohl, Cravo, Wyart, & Nobre, 2012). In chapter 2 we show that this detection improvement is also acquired even when the temporal regularity itself is not yet perceived. Moreover, adding multiple types of temporal information (rhythmicity as well as temporal cueing) improves detection even more, even though one type of temporal information would already have been sufficient to estimate the temporal arrival exactly. This last effect is likely mediated by the fact that temporal estimates are never fully accurate (Eisler, 1976) and adding multiple types of not fully accurate estimates collectively improve perception. This finding relates to cue integration and Bayesian models of perception which state that the brain combines multiple perceptual cues to optimize perception (Ernst & Bülthoff, 2004; Fetsch, DeAngelis, & Angelaki, 2013; Knill & Pouget, 2004). Each sensory cue is typically weighted by its own reliability, improving perception in an optimal fashion. Conclusively, the benefit of combining multiple types of temporal information seems to be similar compared to other types of sensory cues (also see Elliott, Wing, & Welchman, 2014).

The optimized percept in Bayesian models is not only influenced by directly preceding information. Instead, information from multiple time scales is integration and updated online to arise to an optimal

percept (Kim, Basso, 2010; Montagnini, Mamassian, Perrinet, Castet, Masson, 2007; Ernst & Bülthoff, 2004). In chapter 4, there are two types of temporal information on two different time scales. Firstly, the rhythmic structure of the entrainment stimuli aids participants to attend to moments in time that they expect a stimulus to occur in the rhythm. Secondly, there are different lengths of the entrainment train, thus participants have an expectation of the entrainment train continuing. We show that temporal estimations of stimulus occurrence are influenced by expectations whether a rhythmically presented stream of stimuli will continue and thereby shows that the estimates do not just respond to the immediately preceding stimuli, but instead estimates are updated continuously to improve perception (Friston, 2011; Siegel, Buschman, & Miller, 2015; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010).

Rhythmicity of stimuli is picked up by the brain by resonating the neuronal responses to the presented rhythmic stimuli (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008). We also find this resonance effect both in chapter 3 and 4. Chapter 3 shows that these resonating properties can even be present when 1) the participant is unaware of the stimulus stream and 2) the stimuli do not induce any direct measurable evoked response. This chapter shows that it is the alignment of oscillatory phase and not the change in amplitude of an oscillation that drives the entrainment effect (Makeig et al., 2002). This alignment could be a mechanism to attend to specific moments in time, thereby increasing the sensitivity of detection at that time point (Schroeder & Lakatos, 2009). This has a benefit as there is one specific most excitable phase on an oscillation, where neurons are more likely to fire with less input (Buzsáki & Draguhn, 2004; Lakatos et al., 2005). The entrainment is also highly influenced by context. Stefanics et al. (2010) have shown stronger entrainment effects for more probable events. In chapter 4 we further elaborate on this finding, showing that expectancy of entrainment continuation also influences the strength of the entrainment. These results highlights that the brain's response is not a direct reflection of the environmental input, but instead entrainment to environmental stimuli is mediated by mechanisms that predict whether the rhythmic input will continue.

During temporal cueing the brain also seems to optimize perception at the time at which a stimulus is expected by using oscillatory

mechanisms. In chapter 5 we show low frequency oscillatory responses when temporal information is predicted over a longer time window. This effect occurs as there is an optimal detection window on the oscillation period when stimuli are closer to threshold (Buzsáki & Draguhn, 2004; Lakatos et al., 2005). By resetting lower frequencies this window is broadened, effectively increasing the window at which stimuli are optimally processed, which is beneficial when stimuli are predicted to occur for a longer time period (also see Wilsch, Henry, Herrmann, Maess, & Obleser, 2015). This is further verified since the phase of low frequency oscillations also determines whether a target will be detected or not. These findings fit closely to mechanisms proposed for temporal attention during rhythms (Schroeder & Lakatos, 2009; Peele & Davis, 2012) as they indicate that the phase of oscillations is reset to optimally improve detection also during temporal cues. Additionally, they indicate that the frequency of the phase reset seems to depend on the temporal statistics in the environment.

Part I of this thesis shows that stimulus detection is pro-actively optimized by using the temporal regularities in the environment both afforded via rhythmicity and temporal cueing. It is shown that oscillatory brain responses change both their phase and frequency to align the most excitable phase of the oscillation to the time point that stimuli are expected. These results highlight that attention can be directed in time (Coull, Frith, Büchel, & Nobre, 2000) and that these mechanisms are mediated via oscillatory properties in the brain (Schroeder & Lakatos, 2009; Zion Golumbic, Poeppel, & Schroeder, 2012).

Part II: Long term temporal statistics during audio-visual speech

Some temporal associations between audio-visual stimuli can be quite consistent. Especially in speech there are many temporal consistencies that have behavioral relevance to optimize speech perception both in audio only (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Poeppel, 2003; Rosen, 1992) and audiovisual speech (Myers, 1971; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Zion Golumbic et al., 2012). In chapter 6 we show that there is a consistent relationship between the onset of mouth movements and the onset of a speech sound. This relationship is unique to specific syllables [see also (Chandrasekaran,

Trubanova, Stillittano, Caplier, & Ghazanfar, 2009)] and is used to perceive syllable identity. These results indicate that audio-visual temporal cues do not only indicate when auditory information will occur, but also aid the identification and categorization of syllables (the what). These results go beyond the function of temporal information to drive attention in time (Nobre, Correa, & Coull, 2007), but show that temporal information itself is a cue for categorizing stimuli.

The neuronal coding of these consistent temporal relationships has not been investigated thoroughly. The unique feature of varying temporal visual-to-auditory delays for different syllables provides an exclusive starting point to investigate whether temporal information in audiovisual settings can influence the neuronal coding of syllables. In chapter 7 we show that syllable identification of an ambiguous syllable (either perceived as /da/ or /ga/) is biased when the syllable is presented at a specific phase. The phase difference between whether the ambiguous syllable will be identified as /da/ or /ga/ exactly matches the audio-visual delay difference between these two syllables and indicates that the temporal information afforded by audiovisual presentation is transferred to the coding of the auditory information. This is further supported in chapter 8 where we show that temporal cortex can decode syllable identity better when the syllables are presented at their preferred phase. These results show that phase information is integral part of the representations of these syllables.

The idea of a temporal code that codes different representations on oscillatory phase is not new (see Bernstein, 1967 for one of the first accounts). For example, O'Keefe and Recce (1993) found that specific neuronal populations coding for locations in the environment fire at specific phases of the ongoing theta oscillation in the hippocampus. Neuronal populations coding for different locations thus have one specific preferred phase of firing. More recently, the role of phase for content representation in the cortex is being uncovered (Kayser, Montemurro, Logothetis, & Panzeri, 2009; Lopour, Tavassoli, Fried, & Ringach, 2013; Panzeri, Petersen, Schultz, Lebedev, & Diamond, 2001; Watrous, Fell, Ekstrom, & Axmacher, 2015). It seems evident from these data that indeed different representations fire synchronously at one specific phase. On the one hand, this coding scheme binds the different features that the neurons are coding for (Crick & Koch, 1998; Engel & Singer, 2001; Fries,

2005; Singer & Gray, 1995). On the other hand, it provides a clear separation between different representations and this separation should improve the memory for these representations (Fell & Axmacher, 2011; Lisman, 2005; O'Keefe and Recce, 1993; Singer, 1999; Watrous et al., 2015).

The separation of representations by phase is very attractive. However, up to date it is not clear how specific neuronal populations start preferring one specific phase over another. Most accounts state that the changing excitability during an oscillatory period is the driving force for any phase influence on perception or cognition (Buzsáki & Draguhn, 2004; Giraud & Poeppel, 2012; Jensen, Gips, Bergmann, & Bonnefond, 2014; Lisman & Jensen, 2013; Schroeder & Lakatos, 2009). For example, it has been proposed that attention mechanisms reset the phase of the oscillation to be most excitable at the time point that stimuli are expected (Schroeder & Lakatos, 2009); as is supported by part I of this thesis. However, coding information is very different from this type of temporal attention. To incorporate excitability in the coding of information, Jensen et al. (2014) proposed that more salient input is processed earlier on the duty cycle of the oscillation as they can still reach an action potential even though the potential is further from threshold. This mechanism provides an intuitive manner by which different representations are ordered on the oscillatory period. This mechanism can also explain how sequential locations are ordered: the first upcoming location will be at the least excitable point as it is the most salient representation. However, it would predict that dependent on the saliency of the external stimuli, the associated phase would change. Consequently, the coding of specific representations on specific phases is not possible in this type of coding scheme.

Excitability might be an intuitive way to code information, but the reported effects of preferred oscillatory phase for different syllables seems in contrast with this notion (chapter 7 and 8). There is no reason to believe that one syllable should be preferentially processed at the more excitable part of the cycle compared to another syllable. Moreover, we did not find a clear consistency in the phases coding the individual syllables over participants, suggesting that there is not one specific phase that codes for one syllable. This effect might be partly driven by variances in brain anatomy of the participants: different anatomies could shift the

phase as measured with EEG compared to the phase measured at the source of the oscillation. Nevertheless, it is striking that the phase difference between /da/ and /ga/ matches the exact temporal difference in visual-to-auditory delays found in natural speech between these syllables. The presence of a clear difference implies that at least one syllable representation is not preferentially processed at the most excitable phase of the oscillation. Our results instead suggest another way in which specific representations start preferring a specific phase: it is a consequence of the temporal associations that are present in the environment. As visual mouth movements are presented oscillatory patterns in auditory regions reset (Perrodin, Kayser, Logothetis, & Petkov, 2015). Dependent on the specific temporal delay to the auditory speech sound that is unique for each syllable, syllables are presented at a specific phase. This consistent sequence of events leads to the association of specific syllables to specific phases. In this scheme the phase of syllable coding is a consequence of the temporal association between stimulus pairs and becomes an integral part of the representation. This would suggest that it is not excitability, but temporal relationships in the environment that guide at which phase neurons start preferentially firing (also see Kösem, Gramfort, & van Wassenhove, 2014).

There is an increasing amount of data showing that there might be a temporal phase code by which representations are stored (Fell & Axmacher, 2011; Lisman & Jensen, 2013). Part II of this thesis suggests that this phase code might consist of the wiring of temporal associations via statistical learning. This coding provides a unique and natural way to categorize information, thereby optimizing perceptual processes. In the future experiments directly testing the learning phase of temporal associations need to be conducted to investigate the evolvement of the coding of these associations. These experiments could account for part of the phase coding effects found in the literature (Kayser et al., 2009; Lopour et al., 2013; O'Keefe, & Recce, 1993; Watrous et al., 2015) and would explain how temporal information required for speech perception is stored (Giraud & Poeppel, 2012; Peelle & Davis, 2012).

Concluding remarks

Temporal information is omnipresent in our environment, but as we use this information implicitly it is hard to imagine the influence that this information has on our perception and thereby behavior. In the current thesis we show that the use of temporal information cannot be underestimated. Ultimately, perceiving the environment around us requires the attentiveness to the changes in our environment (James, 1886; Myers, 1971). Stationarity does not convey new information, but instead it is the temporal dynamics in the environment to which any living creature has to learn to adapt and usefully interact.

References

Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science, 348*(6241), 1352-1355.

Bernstein, N. (1967). *The Coordination and Regulation of Movement.* London: Pergamon Press.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304*(5679), 1926-1929.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436.

Correa, A., Lupiáñez, J., Milliken, B., & Tudela, P. (2004). Endogenous temporal orienting of attention in detection and discrimination tasks. *Attention, Perception, & Psychophysics, 66*(2), 264-278.

Coull, J., Frith, C., Büchel, C., & Nobre, A. (2000). Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia, 38*(6), 808-819.

Coull, J., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience, 18*(18), 7426-7435.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *The Journal of Neuroscience, 33*(9), 4002-4010.

Crick, F., & Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex, 8*(2), 97-107.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin, 83*(6), 1154.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society of London B: Biological Sciences, 281*(1786), 20140751.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception, & Psychophysics, 72*(8), 2274-2288.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*(1), 16-25.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience, 12*(2), 105-118.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience, 14*(6), 429-442.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in cognitive sciences, 9(10), 474-480.

Friston, K. (2011). Prediction, perception and agency. *International Journal of Psychophysiology, 83*(2), 248-252.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517.

James, W. (1886). The perception of time. *The Journal of speculative philosophy,* 374-407.

Jensen, O., Gips, B., Bergmann, T. O., & Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends in Neurosciences, 37*(7), 357-369.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313-319.

Kayser, C., Montemurro, M. A., Logothetis, N. K., & Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron, 61*(4), 597-608.

Kim, B., & Basso, M. A. (2010). A probabilistic strategy for understanding action selection. *The Journal of Neuroscience, 30*(6), 2340-2355.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712-719.

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *Neuroimage, 92*, 274-284.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology, 94*(3), 1904-1911.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*(5872), 110-113.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431.

Lisman, J. E. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus, 15*(7), 913-922.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron, 77*(6), 1002-1016.

Lopour, B. A., Tavassoli, A., Fried, I., & Ringach, D. L. (2013). Coding of Information in the phase of local field potentials within human medial temporal lobe. *Neuron, 79*(3), 594-606.

Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1612.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. (2002). Dynamic brain sources of visual evoked responses. *Science, 295*(5555), 690-694.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition, 115*(1), 186-191.

Montagnini, A., Mamassian, P., Perrinet, L., Castet, E., & Masson, G. S. (2007). Bayesian modeling of dynamic motion integration. *Journal of Physiology-Paris, 101*(1), 64-77.

Myers, G. E. (1971). William James on time perception. *Philosophy of Science,* 353-360.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89*(1), 133.

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology, 17*(4), 465-470.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*(3), 317-330.

Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M., & Diamond, M. E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron, 29*(3), 769-777.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3.*

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proceedings of the National Academy of Sciences, 112*(1), 273-278.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245-255.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *The Journal of Neuroscience, 32*(24), 8424-8428.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336*(1278), 367-373.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106-113.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20*(2), 172-176.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience, 18*(1), 555-586.

Singer, W. (1999). Time as coding space? *Current opinion in neurobiology, 9*(2), 189-194.

Stefanics, G., Hangya, B., HernÃ¡di, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *The Journal of Neuroscience, 30*(41), 13578-13585.

Watrous, A. J., Fell, J., Ekstrom, A. D., & Axmacher, N. (2015). More than spikes: common oscillatory mechanisms for content specific neural representations during perception and memory. *Current Opinion in Neurobiology, 31*, 33-39.

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2015). Slow-delta phase concentration marks improved temporal expectations based on the passage of time. *Psychophysiology, 52*(7), 910-918.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, 122*(3), 151-161.

VALORIZATION ADDENDUM

> Valorization: "*the process of creating value from knowledge, by making knowledge suitable and/or available for social (and/or economic) use and by making knowledge suitable for translation into competitive products, services, processes and new commercial activities*" (Maastricht promotie regelement, 2013).

Valorization of this thesis

In this part of the thesis I am supposed to write about the value of the knowledge that I created with my research. Since 2014 this is an obligatory part of all theses that are produced at Maastricht University. At a first sight it sounds fair to make your research valuable by "*making knowledge suitable and/or available for social use*" as tax payers are investing in us. However, what does this exactly mean?

I would interpret the definition "*making knowledge suitable and/or available for social use"* as implying that everybody in society can access the knowledge I created ("availability"). This thesis will be publicly available at the website of Maastricht University from the day of my defense on. Moreover, anybody can get a copy if they approach me. So the availability of my research is confirmed. Valorization seems to be an easy task. But I am forgetting one word here, the word "*suitable*". This would mean that the knowledge I am creating should be at a level at which society can understand it. This is going to be trickier as this depends on people's background knowledge. I operationalize "*suitable*" when people outside my research field can understand it. I guess this would require that I would rewrite my empirical chapters in short understandable language (I also refer to lolmythesis.com for others trying this). Fair enough:

Chapter 2: Detecting soft sounds is easier when you know when a sound is arriving.

Chapter 3: Brain waves track sounds when they are presented rhythmically.

Chapter 4: Brain waves only track sounds when you know that a sound will occur in the first place.

Chapter 5: Brain waves are slower when you are unsure about the exact point in time when the sound will occur.

Chapter 6: People use timing information created by mouth movements in syllables to identify them.

Chapter 7: Different syllables are represented at different time points on a brain wave.

Chapter 8: Some regions in the brain respond to sounds. The same regions respond different when an identical sound is presented to different time points on a brain wave.

Although this is a very brief summary, it is in more or less simple terms what all the chapters were about ("suitably available" knowledge). Anybody can use this for any social and/or economic purpose. If more detailed information is required I would be happy to communicate this further. It seems that I am done with this chapter. Let's double-check with the guidelines in the regulations whether I fulfilled the requirements:

"Five questions are provided that can guide candidates in writing this addendum

1) *(Relevance) What is the social (and/or economic) relevance of your research results (i.e. in addition to the scientific relevance)?*
2) *(Target groups) To whom, in addition to the academic community, are your research results of interest and why?*
3) *(Activities)/Products) Into which concrete products, services, processes, activities or commercial activities will your results be translated and shaped?*
4) *(Innovation) To what degree can your results be called innovative in respect to the existing range of products, services, processes, activities and commercial activities?*
5) *(Schedule & Implementation) How will this/these plan(s) for valorization be shaped? What is the schedule, are there risks involved, what market opportunities are there and what are the costs involved?"*

This seems relatively far from the core definition of valorization as defined in the regulations, in my opinion. While in the original definition

I am required to make my knowledge *available* at a suitable level for society, these guidelines require me to also decide on *how* this knowledge is utilized by deciding the relevance, possible products, and even when this is all going to be made. I thought that as a scientist I was required to create knowledge. Just like an artist is creating art, a scientist creates knowledge (science literally means knowledge). Other professions and society should profit from this knowledge, I fully agree with this. But according to the operationalization of valorization above I should not only create the knowledge, but also implement it in society. Moreover, it seems that I should even let my research be guided by this, because how could I answer these questions if my research is not directly relevant for society in the first place? This would mean a goodbye for the pure scientist, the creator of knowledge and a hello to the periodontist, the creator of relevance.

Valorisatie van onderzoek als taak van de universiteiten?

More than ten years the government has been promoting valorization in the university. It first affected the universities in 2005 when the letter "*valorisatie van onderzoek als taak van de universiteiten*" appeared. In this letter it is explained that valorization is one of the core tasks of university, next to creating education and doing scientific research. It is a basic statement that universities should think about what type of research they are doing and how this information can be conveyed to the society (which in principle I do not oppose to). But let me just elaborate on the specification of this letter since this letter is one of the starting forces why I am writing this section in my thesis in the first place. The letter starts out as follows (freely translated): there is no discussion that a significant part of academic research should not be aimed for any direct or indirect societal use, but to maintain and contribute to worldwide scientific developments. The agenda is not decided by any societal question, but by the research possibilities. ("*Buiten discussie staat dat een belangrijk deel van het universitaire onderzoek niet primair gericht moet zijn op direct of indirect maatschappelijk nut, maar op het bijhouden van en bijdragen aan wereldwijde wetenschappelijke ontwikkelingen. De agenda wordt echter niet bepaald door de maatschappelijke vraag, maar door de onderzoeksmogelijkheden*"). This type of research was defined as "offer based": research done to create scientific developments

("*aanbodgedreven: het wordt gedaan omdat er snelle wetenschappelijke ontwikkelingen zijn of te verwachten zijn*"). Next to this offer based research, universities were required to provide "question based" research: research aimed for answering societal questions ("*het geven van antwoord op maatschappelijke vragen. Het kan zowel gaan om vragen van bedrijven als van de overheid en van niet-commerciële maatschappelijke organisaties*"). Let us ignore the poorly chosen definition as all (proper) research is based on a research question and agree that it is valid to have research performed for scientific developments as well as societal relevant questions and that a balance has to be found. Here comes the problem of this letter that universities face: universities are required to prioritize within the offer based research, research that has the possibility of creating a synergy between business and societal parties, and the subsequent possibilities for economical and societal valorization, and responsibly report the results of this "prioritization". ("*Wij verzoeken de universiteiten en onderzoekinstellingen dan ook om in hun strategische plannen aandacht te besteden aan de mogelijkheden om bij de prioriteitsstelling binnen het aanbodgedreven onderzoek de mogelijkheden tot het scheppen van synergie met het bedrijfsleven en maatschappelijke partijen, en de daaruit voortvloeiende mogelijkheden tot economische en maatschappelijke valorisatie, expliciet mee te wegen en om in hun verantwoording te rapporteren wat daarvan het resultaat is geweest.*"). I freely interpreted this as the proposal that although offer based research should not be guided by societal questions, universities should fund offer based research that anyway answers these societal questions.

How did the minister think about the practicalities of this? We, scientists come up with research questions in our "offer based", non-societal relevant manner and offer this research to the world. The universities just fund whatever by chance also seems to answer a societal question. And the scientific research questions are magically uninfluenced by this "prioritization" scheme of the university. Let us be honest here, if universities truly implement this suggestion funding only applies to scientists with research questions that have societal relevance. Thus, no more offer based research, but only societal relevant research.

Is it a problem if we lose any fundamental research not aimed to answer societal questions? Yes, definitely. Many great inventions

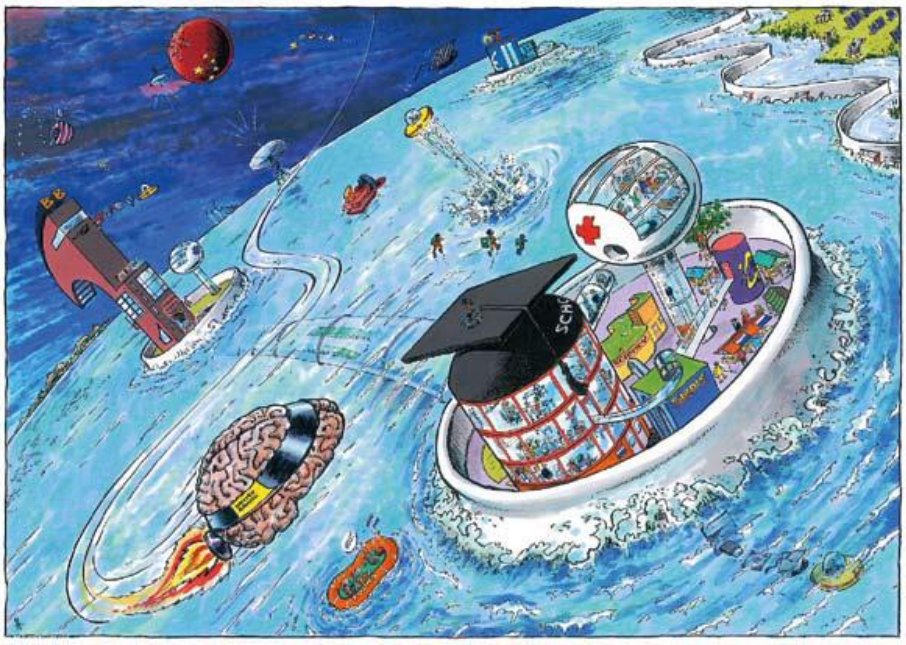
appeared without the aim of any societal benefit (penicillin, electricity etc.), and even the minister seems to agree with this in her letter. I will not go into detail here why basic research not guided by societal questions is relevant as many before me provide convincing answers to this. I freely refer to:

- Curiosity creates cures: The value and impact of basic research, National Institute of General Medical Sciences, National Institutes of Health.
- ICSU position statement: The value of basic scientific research", International Council for Science, December 2004.
- Liz Karagianis - MIT Spectrum (21 April 2015). "How discovery science is reinventing the world - MIT News". MIT News.
- Karen Kashmanian Oates – The Importance of Basic Research – Huffington Post

If you are not convinced of the role of basic research I would like to see you try to solve cardiac arrest without knowing how the heart works in the first place. The point that is important here is that there is no logical way to infer which knowledge will be relevant in the future and which knowledge will not be relevant. At what level of processing should we know all details of the heart to solve all heart diseases? Is the level of arteries enough, or should we go for the level of genes, or single atoms? Who knows what level will be sufficient? (see also a great and entertaining discussion at the cognitive neuroscience meeting 2016 about a related topic: https://www.youtube.com/watch?v=uSbNRyY2QH0). Is the researcher investigating the blood flow in the left anterior descending artery required to develop the cure for cardiac arrest? I say no, this person is merely obliged to share his knowledge within the community so that we know how this detailed piece of knowledge could aid in figuring out heart functioning as a whole.

Reflection: Knowledge has to circulate

In 2009 another beautiful document appeared from the government (Van voornemens naar voorsprong: Kennis moet circuleren). They created a vision what valorization would bring us in 2016. The following picture provided their vision. I leave this up for your own interpretation.



Now, we are in 2016. So this is the perfect time to reflect on valorization as it is implemented. What has valorization in universities brought? Of course not much that this picture provides (although I would have liked to have a flying brain), but I guess that could have been expected in 2009. Valorization as it is implemented now forces PhD students to think about how their research is relevant for society in a five-to-eight page document at the end of their thesis. It forces bachelor students just learning about research for the first time to report on the relevance of their intended research. So already during their studies they are drilled that research is about societal questions. It requires every grant application to contain speculations and operationalizations of the possible impact of our research even if we are not qualified at all to actually implement the possible societal relevant output. Finally, less and less funding is available for fundamental research.

Are there only bad consequences of valorization? No, I think there are some clear positive changes that we should focus on. I think the main positive sign that is happening is focusing on the "*availability*" of knowledge. This is in the end the core of valorization for fundamental research to me. The Dutch government has been pushing publishing

agencies to publish research "open access". This means that anybody can access your publication, instead of letting universities pay for specific journals, by which the normal public does not have any access. Moreover, many journals have been pushing you to rephrase your research findings in layman's language. Although I would prefer that they call these sections "Layman's message" instead of "significance", they still provide a means to have a better understanding of the main message of a paper. These advances make it possible for societal or business partners to *use* the created knowledge. Knowledge exchange might be one of the core problem is modern-day science. We scientists have specialized knowledge about one small fraction of science. The exchange of this knowledge is still very difficult. Especially considering that different research fields are speaking with their own language. It is necessary to invest in ways to improve communication with each other.

I fully agree that knowledge is there to be exchanged, indeed to circulate. However, let the different parties do what they are experts in. Basic scientists to create knowledge; applied scientists, businessmen and societal institutions to use this knowledge and implement it for products, programs etc. (also read about why many researchers unconsciously are busy with valorization in the first place https://pure.knaw.nl/portal/files/1514072/Hoe_onderzoekers_werkelijk_valoriseren.pdf). Let us focus on the exchange between these different groups instead of forcing basic scientists to perform all these steps. Create transparency and a platform to exchange knowledge at a suitable level such that qualified people get the relevant knowledge (the suitability level would vary according to the background knowledge of the exchanging parties. The one provided in this section would of course not suit many applications much more than my mom understanding a bit better what I do). Leave space (and funding) for basic research that has no direct societal influence and leave the research questions to the scientists and the implementation of knowledge to the implementers (which of course can include qualified scientists). Maybe then we can create our flying brain in 2025.

ACKNOWLEDGEMENTS

Getting a PhD is certainly a journey on which you meet a lot of people that all in their own way have a significant impact on this piece of work you see in front of you. I will try to make an attempt to show some gratitude to all the great people I met the last years.

First of all, I want to thank my supervisors Alex and Nienke. Alex, I am very happy you took on this challenge with me when I asked you to write this NWO grant. Your open spirit to let me try the research path I wanted to take really suited with my personality and needs. You are a great writer, great scientist, and a great person. Nienke, although you were in Amsterdam during the supervision, you were always avaible and have a great mind for good research. You gave me a jump start in New York and had significant impact on my whole PhD path. Thank you both for all your input and making this happen!

Mehrdad, you are the best thing that I walked into when returning to Maastricht. The last years have been amazing. I now finaly recrify an erroneous statement in "*Behind the Seen*": you are officially *de beste*. Thanks for helping me and putting up with me working so much. Much more great things are on our path.

The Cognitive Neuroscience department in Maastricht is a great place to do research. The interactive atmosphere is something to treasure and I hope it stays the same although the department is growing. All seniors, Elia, Rainer, Bernadette, Peter, Beatrice, Kamil are and should be fighting for this. Thank you creating this great atmosphere. Fren, thank you for all the conversations, I learned a lot more about EEG from you. Also thanks to all my collegeas at the "Brain and Cognition", aka the TMS group for great interactive and informative meetings (also for non-scientific topics). Tahnee, Rosanne, Dennis, Nina, Christianne, Lukas, Helen, Felix, Tom, Geraldine, Teresa, Jasmina, Franzi and others: you are great. Also other colleages you are always ready for scientific and non-scientific drinks/conversations. Thank you Rebecca, Lars, Kiki, Joao, Martin, Gojko, Michelle, Nathalie, Mario, Victoria, Tobi and many others. My different office mates, Tamara, Lisa, Tabea, Dennis, and Nina. Thank you for coping with me. Dennis thank you for the welcome in my new office and the many hand-and-craft smiley faces to light up the morning. Kirsten, you have grown into a great researcher. You are officially at the point that I don't understand your matlab scripts and science talk anymore.

Most PhD students at FPN might see my name at some of the PhD academy emails that I have been sending around all the time. To all PhDs: go to their events! The people there are great. Thanks to all (at some point) members of the board there for organizing amazing events and making my life in Maastricht great. Daan, Alejandro, Elisa, Alisson, Pieter, Dorijn, Roxanne, Hendrik and many more.

The best thing in Maastricht is the improv group. I have never met such a crazy, amazing group of people. Please don't change and let's go crazy many more times! Nevena, Gabri, Anna, Dorijn, Jessie, Paola, Jan, Joanne, Despina, Mare, Adam, Petor, Eveline, Mare, Zubin, Bas, Latifa, Frederik, Lukasz, Alan, Norbert, Masha, Paul, Francesco, Mari, Nordin, Burcu, Yeliz and many more.

Maastricht is small, but has great things and great people. I can't image what we have created the last years. We made a movie, we are going to make a play. International dinner people, thanks for making great food, although the original purpose was lost somewhere on the way ☺. Tess, Paula, Giacomo, Alexandros, Julie, Joost, Jo, Yara, Sean, Artemis, Alejandro, Cheng, Shuan, Elisa and more, keep on with this great food exchange. Swing Out Maastricht, you have created something great so quickly. Karen, Claire, Catherina, Georgina, Hannah, I'm proud of you and be proud of yourself! Thank you for teaching me so far. And so many others, Ehsan, Saba, Shyam, Menica etc..

Also outside of Maastricht there are great people. Two of these chapters have their roots in New York. Elana, you are my science goddess from NY. It must have been horrible waiting for my poor matlab skills to develop. Thanks for the patience. Charlie, thank you for all the input, letting me stay at your house when I just arrived, and pushing my research career forward (and a memorable farewell barbeque). David, it was a pleasure crashing at your lab and recording all the data. You have a great group of people around you and you have an amazing research style. Thank you people from NY, Johanna, Luc, Gwyneth, Keith, Xianbing, Lucia, Saskia, Molly, Michael, George and more. Especially my NY budy Jaco, you made this time great.

Ook thuis moeten ze me missen. Afra, Angelique, Cris, Dion. Fab Five, onze traditionele levens-analyses zijn onevenaarbaar. Jullie krijgen een 10! Mijn prettig gestoorde familie. Mams, Pim, Maarten, en Marlou. Het is altijd heerlijk om thuis te zijn, waar alles rust en eenheid is.

Spelletjes zijn altijd dichtbij. Mam, ik ben zo trots op hoe je met alles omgaat en altijd probeert alles wat ik doe te kunnen volgen. Oma, wat moeten we zonder jou. Sorry dat ik er zo weinig ben, maar ik was dit boek aan het schrijven. Het zal hopelijk mooi in de kast staan. Andere familie van Ten Oever en Simons, dank. Nieuw-Wehl is een geweldig plaats.

Nu als laatst de persoon die dit alles zal moeten missen. Paps, je was er niet meer tijdens mijn CN onderzoeksleven. Maar zonder jou was dit niet gelukt. Je wordt gemist, zeker tijdens mijn verdediging. Je zou het geweldig hebben gevonden. Paps, je bent een fantastisch persoon en de beste vader. Dank, dank, dank je voor alles.

PUBLICATIONS

AND CURRICULUM VITAE

Publications

Ten Oever, S., Romei, V., Van Atteveldt, A., Soto-Faraco, S., Murray, M., Matusz, P (2016). The COGs (Context-Object-Goals) in multisensory processing, Experimental Brain Research, 234(5):1307-23

Petras, K., **Ten Oever, S.**, & Jansma, B. (2016). The effect of distance on moral engagement: event related potentials and alpha power are sensitive to perspective in a virtual shooting task. Frontiers in Psychology, 6:2008

Ten Oever, S., Sack, A.T. (2015). Oscillatory phase shapes syllable perception. Proceedings of the National Academy of Sciences of the United States of America, 112(52):15833-7

Ten Oever, S., Van Atteveldt, N.[1], & Sack, A.T.[1] (2015). Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. Journal of Cognitive Neuroscience, 27(9):1811-22

Ten Oever, S., Schroeder, C.E., Poeppel, D., Van Atteveldt, N., Zion-Golumbic, E. (2014). The influence of temporal regularities and cross-modal temporal cues on auditory detection. Neuropsychologia, 63:43-50.

Ten Oever, S., Sack, A.T., Wheat, K.L., Bien, N., Van Atteveldt, N. (2013). Audiovisual onset differences are used to determine syllable identity for ambiguous audiovisual stimulus pairs. Frontiers in Psychology, 4:331.

Ten Oever, S.[1], Bien, N.[1], Goebel, R., Sack, A.T. (2012). The sound of size: Investigating the neural correlates of synesthesia in the normal population by combining TMS, EEG and psychophysics. Neuroimage, 59(1):663-72.

Blokland A., **Ten Oever, S.**, Van Gorp, D., Van Draanen, M., Schmidt, T., Nguyen, E., Krugliak, A., Napoletano, A., Keuter, S., Klinkenberg, I. (2012). The use of a test battery assessing affective behavior in rats: order effects. Behavioral Brain Research, 228(1):16-21.

[1]Contributed equally to the work.

Submitted Manuscripts:

Ten Oever, S. & Sack, A.T. Sensory entrainment effects increase with varying entrainment lengths (under review at Journal of Cognitive Neuroscience).

Ten Oever, S., Hausfeld, L., Correia, J.M., Van Atteveldt, Al., Formisano, E., Sack, A.T. Oscillatory phase shapes syllable representations: A 7T fMRI study using sensory entrainment (under revision at NeuroImage)

Ten Oever[1], S., De Graaf[1], T.A., Bonnemayer, C., Ronners, C., Sack., A.T., & Riecke., L. Stimulus presentation at specific neuronal oscillatory phases experimentally controlled with tACS (under review at Journal of Neuroscience Methods)

Manuscripts in Preparation:

Ten Oever, S., Schroeder, C.E., Poeppel, D., Van Atteveldt, N., Zion-Golumbic, E. Evidence for entrainment to sub-threshold rhythmic auditory stimuli.

De Graaf, T.A.[1], Duecker, F.[1], Y. Stankevich, **Ten Oever, S.**, Sack, A.T. Seeing in the Dark: Phosphene thresholds for open versus closed eyes in the absence of visual inputs.

[1] Contributed equally to the work

About the Author

Sanne ten Oever was born on October 9th 1989 in Zevenaar, the Netherlands. She completed her high school education at the St. Ludger College in Doetinchem in 2007 and then studied the Bachelor Psychology at Maastricht University. During this Bachelor she published her first paper together with Dr. Nina Bien and Prof. Dr. Alexander Sack. She finished this study with summa cum lauda in 2010. She continued her studies at Maastricht with the Research Master in Cognitive and Clinical sciences. During these studies she has spend time at the group of Prof. Dr. Charlie Schroeder (Columbia University) and Prof. Dr. David Poeppel (New York University) in New York City. Under guidance of Dr. Elana Zion-Golumbic she did a project studying how temporal information influences auditory perception. She finished her studies with cum laude in 2012. Her interest in research stayed and she successfully applied together with Prof. Dr. Alexander Sack for a grant to continue her PhD. Prof. Dr. Alexander Sack and Dr. Nienke van Atteveldt were her supervisors during this PhD. She finished this PhD in 2016 at is currently working as a PostDoc for the Brain and Language group under guidance of Bernadette Jansma.