Lexical and audiovisual bases of perceptual adaptation in speech

Each chapter of this dissertation has addressed an aspect of a perceptual strategy known as recalibration or retuning, a process through which listeners can learn to adapt to a speaker by attending to information other than the auditory signal itself. These sources can include the lip-movements of the speaker (also known as audiovisual cues) or the listener's own lexical knowledge, which can assist them in making assumptions as to what the speaker is most likely to be saying. Repeated experience with pairings between an ambiguous auditory signal and these contextual sources can shift boundaries between phoneme categories (Bertelson, Vroomen, & De Gelder, 2003; Norris, McQueen, & Cutler, 2003) and thereby allow the listener to understand a speaker with more ease (Sjerps & McQueen, 2010).

Summary

In Chapter 2, lexical retuning and audiovisual recalibration were compared with a novel paradigm where listeners switched between the two forms of perceptual learning within a single session. Switching did not lead to significant cost in learning effects, compared to groups that received only one type of cue. Audiovisual recalibration effects were stronger than lexical retuning, in a switching group and a single exposure group, but this was most likely due to the design of the study which contained short exposures in two possible acoustic directions, unlike most previous studies of lexical retuning (Cutler, Eisner, McQueen, & Norris, 2010). Nevertheless, listeners were able to show alternating forms of perceptual learning, indicating that both audiovisual recalibration and lexical retuning are flexible even under constrained conditions. The obtained results may reflect how listeners can switch between informative contextual sources depending on the needs of the listening situation.

In Chapter 3, lexical and audiovisual cues were combined to investigate whether and how the combination of cues would lead to perceptual shifts. The combined cues produced effects that were similar in magnitude to audiovisual recalibration effects, but were larger than lexical effects. Again, a constrained design was employed (with short and alternating exposure/test blocks), so lexical cues may have led to diminished effects with the atypical design, but lexical and audiovisual cues also did not additively combine to induce perceptual boundary shifts. Rather, the combination of cues led to effects larger than lexical retuning alone and comparable to audiovisual recalibration. The pattern of results suggests that lexical and audiovisual cues do not operate together when inducing categorical shifts, and the two types of cues may be relied upon for different purposes.

Finally, in Chapter 4, lexical retuning and audiovisual recalibration were compared in an fMRI study, to pinpoint the neural correlates underlying the two processes, and to identify how much overlap they share. Once again, an alternating blocked design was used in order to have participants undergo both forms of perceptual learning with two phonemes within a short window of time. During exposure blocks, audiovisual and lexical cues elicited similar patterns of activity in the temporal cortex, across Heschl's gyrus (HG), planum temporale (PT), superior temporal gyrus (STG) and sulcus (STS). These regions are involved in acoustic and phonemic processing (HG/PT/STG) as well as higher-level syllabic and semantic information (STG/STS) (Buchsbaum, Hickok, & Humphries, 2001; Formisano, De Martino, Bonte, & Goebel, 2008; Jäncke, Wüstenberg, Scheich, & Heinze, 2002). Significant activation was also found in the inferior parietal lobule (IPL) and the insula, but audiovisual exposure blocks specifically led to activation in the occipital cortex, between V1 and V2. Similarly, during test blocks, when listeners were undergoing recalibration or retuning resulting from the preceding audiovisual or lexical cues, activation was observed in HG and STG/STS in the temporal cortex, as well as IPL and insula. During audiovisual test blocks, significant activity was still evoked in the occipital cortex (V1/V2), even while no visual stimuli were presented. In addition, a number of regions defined by activity during the exposure blocks showed distinct differences in the degree of activation between high and low recalibration (i.e. more or fewer responses in the same direction as the bias contained in the prior exposure block). These regions included temporal, occipital, insular, and motor clusters, but only showed the high-low distinction for audiovisual test blocks, while no regions were significantly distinguishable for lexical test blocks. Overall, results showed that the areas of the brain involved in lexical retuning and audiovisual recalibration overlap in many respects especially within the auditory cortex, but audiovisual recalibration seems to trigger a specific reactivation of the occipital cortex, which suggests the involvement of mental imagery (i.e. re-activation of visual representations from short-term memory) during shifts (Pearson, 2019). A network of regions across the brain also appears responsible for effectively shifting the category boundary, involving both low-level acoustic/phonetic processing, and higher-level cross-modal and semantic processing.

Taken together, the outcomes of these studies have clarified some of the similarities and differences between lexical retuning and audiovisual recalibration. In Chapter 2, retuning and recalibration were both found to be flexible, as listeners proved capable of switching between them, but lexical retuning can be limited in a design where blocks rapidly alternate between exposure and test, and between two different phonemes. However, this difference in effect size may represent differences in the typical applications of the processes, where audiovisual cues may be more suitable for short-term, situation specific learning (a noisy environment) whereas lexical cues may be more applicable to long-term, speaker-specific learning (unfamiliar accent, unusual pronunciations). Chapter 3 identified how retuning and recalibration seem to differ and do not additively combine to enhance aftereffects. It appears that lexical and audiovisual cues operate across different networks, and that there are domain-specific aspects of the phoneme categories that they tap into, which may prevent the cues from being utilized simultaneously. In addition, listeners do not seem to benefit from the combination of cues if one cue type is sufficiently informative; for example, the audiovisual cue may have already indicated to the listener what the ambiguous phoneme was most likely to be, then the lexical cue may not have provided any additional guidance. If two possible phoneme candidates are visually identical (such as /b/ and /p/), then lexical information may be more useful, but if two phonemes are visually different (such as /p/ and /t/), then audiovisual cues may be more helpful. Listeners most likely utilize whichever cue is fastest and most reliable in the given situation. Chapter 4 delineated the neural activity underlying retuning and recalibration, and both processes engaged areas across the temporal cortex that are known to be involved in rudimentary acoustic processing, such as HG, STG/STS, and PT. Both retuning and recalibration also showed patterns of reactivation between exposure and test, as many of the same areas activated by the exposure blocks, when listeners where presented with either the audiovisual or lexical stimuli, were also activated by the test blocks, when only ambiguous phonemes were presented in a categorization task. However, the observed neural activity also points to modality-specific contributions, as audiovisual recalibration recruits the visual cortex, while lexical retuning largely relies on the speech network both within auditory cortex, and in other related areas such as IPL and insula.

Discussion

This dissertation sought a cohesive explanation of the various forms of perceptual adaptation, but a number of questions still remain unanswered and must be taken into consideration in order to bridge the gap in understanding between the two processes. The three studies revealed some of the limitations in perceptual adaptation studies, so future studies may benefit by circumventing these drawbacks accordingly. Many of these potential restrictions involved the stimulus construction, the study design, and the confines of an fMRI study. However, the findings across the three studies also elucidated some of the processes involved in speech perception, and how theories of speech perception may or may not be equipped to explain what perceptual adaptation entails.

Stimulus construction & design

The three studies used largely similar approaches to measure perceptual shifts, with alternating blocks of exposure and test, containing only six or eight stimuli, and with the phoneme bias also changing throughout the experimental session. This design, derived from a previous study (van Linden & Vroomen, 2007), allowed us to compared retuning and recalibration under the same constraints, as well as efficiently testing two forms of perceptual learning in two directions within the same session (Figure 1).