

1 Background

Listeners of a language are faced with a large degree of variability when interacting with their fellow language users. Speakers can have different sizes, different genders, and different backgrounds that make speech sound categories, at first blush, overlapping in distribution and harder to separate in acoustic domains. Despite this, listeners maintain a large degree of perceptual constancy, through speaker normalization and perceptual learning. Ambiguous vowel sounds can be mapped to different categories depending on the vowels of the preceding context (Ladefoged and Broadbent, 1957), where the previously-unheard ambiguous tokens is normalized to the vowel space present in the context. Perceptual learning refers to the updating of distributions corresponding to a sound category for a particular speaker.

1.1 Perceptual learning

Norris et al. (2003) began the investigation into perceptual learning in speech. Previous work has demonstrated perceptual learning in the visual domain (Color hue reference?) and (I think auditory as well). Norris et al. (2003) exposed Dutch listeners to a fricative halfway between /s/ and /f/ at the ends of words and nonwords and tested their categorization on a fricative continuum from 100% /s/ to 100% /f/. Listeners exposed to the ambiguous fricative at the end of words shifted their categorization behaviour, while those exposed to them at the end of nonwords did not. The exposure using words was further differentiated by the bias introduced by the words. Half the tokens ending in the ambiguous fricative formed a word if the fricative was interpreted as /s/ but not if it was interpreted as /f/, and the others were the reverse. Listeners exposed only to the /s/-biased tokens categorized more of the /f/-/s/ continuum as /s/, and listeners exposed to /f/-biased tokens categorized more of the continuum as /f/. The ambiguous fricative was associated with either /s/ or /f/ dependent on the bias of the word, which led to an expanded category for that fricative at the expense of the other category.

In addition to lexically-guided perceptual learning, unambiguous visual cues to sound identity can cause perceptual learning as well (referred to as perceptual recalibration in that literature). In Bertelson et al. (2003), an auditory continuum from /aba/ to /ada/ was synthesized and paired with a video of a speaker producing /aba/ and a video of /ada/. Participants first completed a pretest that identified the maximally ambiguous step of the /aba/-/ada/ auditory continuum. In eight blocks, participants were randomly exposed to the ambiguous auditory token paired with video for /aba/ or with the video for /ada/. Following each block, they completed a short categorization test. These categorization tests showed clear perceptual recalibration effects, such that the participant was more likely to respond with /aba/ if they had been exposed to video of /aba/ paired with the ambiguous token in the preceding block, and likewise for /ada/.

van Linden and Vroomen (2007) compared the perceptual recalibration effects from the visual lipread paradigm (Bertelson et al., 2003) to the standard lexically-guided perceptual learning paradigm (Norris et al., 2003). Lipread recalibration and perceptual learning effects had comparable size, lasted equally as long, were enhanced when presented with a contrasting sound, and were both unaffected by periods of silence between exposure and categorization. The effects did not last through prolonged testing in this study, unlike in other studies (Kraljic and Samuel, 2005; Eisner and McQueen, 2006).

1.2 Retention

Kraljic and Samuel (2005) looked at the strength of the perceptual learning effect across various types of unlearning tasks. Native English listeners were exposed to ambiguous fricatives in between /s/ and /f/ in words that biased interpretation toward either /s/ or /f/ in a lexical decision task either spoken by a female voice or a male voice. Participants that completed an /s/-/f/ categorization task immediately following exposure and also those that had a 25 minute visual Unlearning task between exposure and categorization showed a large perceptual learning effect. Additionally, participants that had an Unlearning task involving auditory input from the same speaker as they were exposed to showed attenuated perceptual learning effects when the speech contained correct versions of the ambiguous fricatives, i.e. correct /s/ tokens when they were exposed to the ambiguous sibilant in all /s/ words in the exposure task. The perceptual learning effect in this condition was almost eliminated. Participants that had an Unlearning task involving auditory input that did not have any /s/ or /f/ tokens showed perceptual learning effects comparable to those with no Unlearning task. When the Unlearning task involved auditory input from a different speaker than exposure, perceptual learning effects were as robust as when there was no Unlearning task.

Eisner and McQueen (2006) looked at the stability of perceptual learning. Dutch participants first completed a categorization task for a continuum of /s/ to /f/ and then listened to a story that contained ambiguous fricative between /s/ and /f/, with one version having the ambiguous fricative in /s/ contexts and the other having it in /f/ contexts. Immediately following, they completed another /s/ to /f/ categorization task. Finally, 12 hours later, they completed the categorization task once again. Half the participants started at 9 am and completed the final categorization task at 9 pm the same day, and the other half started at 9 pm and completed the final categorization at 9 am the following day. Perceptual learning effects were found in both categorization tasks following exposure, with no significant differences between the two times. The perceptual learning effects found were robust across time and across intervening language exposure, as the those tested at 9 am and 9 pm likely had many interactions with other people in between.

Perceptual learning effects are robust across time. Given no other interactions with a given speaker, a listener's behaviour for that speaker's speech will remain constant over the course of a day, and probably longer. However, the perceptual system remains plastic. If a listener is exposed to a speaker trait in the first interaction, but the trait disappears in later interactions, the perceptual system will reattenuate to the newer

evidence.

1.3 Generalization

1.3.1 Generalizing to new speakers

Kraljic and Samuel (2007) looked at different sound contrasts and the ability of participants to generalize across speakers. The first contrast was voicing between /t/ and /d/. When exposed to ambiguous stops between /t/ and /d/ corresponding to /d/ for one speaker, and ambiguous stops corresponding to /t/ for another, perceptual learning effects corresponded to the most recent exposure, regardless of whether the test voice was the same as that recent exposure voice. The second contrast they used was sibilants of /s/ and /ʃ/. For this contrast, they found the speaker-specificity effect found in the literature, where recency of the exposure did not affect the perceptual learning for a particular voice. They argue that the differences between these two contrasts lies in the relative indexical information carried by the sibilants and the lack of much indexical information carried in the stops.

Eisner and McQueen (2005) looked at what manipulations could cause perceptual learning could generalize to additional talkers. Using a Dutch lexical decision task with ambiguous fricatives in between /s/ and /f/, they tested categorization on a continuum between /s/ and /f/ from two talkers. When exposed to one talker's ambiguous fricative and tested on another speaker's continuum, no perceptual learning effect was found. There were two conditions where the fricatives in the exposure and categorization came from the same speaker, but the speakers of the carrier speech. In one, the ambiguous fricative from the categorization speaker's productions was spliced into the exposure tokens, and in the other, the exposure talker's /s/ to /f/ continuum were spliced with vowels from the categorization talker. In these two conditions, they did find a perceptual learning effect across speakers, despite reports by the participants of hearing different speakers, suggesting a sound-specificity effect in addition (or potentially causing) speaker-specificity effects.

Reinisch et al. (2013) looked at perceptual learning of speaker characteristics across languages. When exposed to a native Dutch speaker speaking English with ambiguous fricatives between /s/ and /f/, participants show a perceptual learning effect when categorizing Dutch minimal pairs differing only in whether one sound is an /f/ or /s/ spoken by the same speaker. When exposed to the same native Dutch speaker speaking Dutch words and nonwords with the same ambiguous fricative, native Dutch listeners and L2 Dutch listeners (native German speakers) show comparable perceptual learning effects.

Reinisch and Holt (2013) Holt exposed native English listeners to a female Dutch-accented voice speaking words and nonwords, where words ending in /s/ or /f/ instead ended either in a natural token or an ambiguous fricative between /s/ and /f/. Listeners showed a perceptual learning effect for the exposure talker as well as another female Dutch-accented voice, showing cross-speaker generalization. However, when categorizing a continuum from /s/ to /f/ of a male Dutch-accented voice, listeners did not generalize the perceptual learning effect, but there was a slight perceptual learning effect for the

first block of categorization tokens for the male voice that disappeared in later blocks. Pretests of the continua revealed that the male voice continuum was heavily biased towards /s/, with steps 1-4 of the continuum consistently heard as /s/. For the exposure female voice, Steps 1-4 were already ambiguous between /s/ and /f/, with the proportion /s/ response less than 60% for all steps. The new female voice was responded to similar to the exposure female voice, but with less ambiguity for the initial step. When the continua for the exposure female voice and the male voice were more closely matched by removing the first four steps of the male voice and removing 4 steps along the continuum of the female voice, generalization of the perceptual learning effects matched those for the new female voice. These findings suggest that the perceptual system leverages known distributions that may not be speaker specific. When new tokens fall outside that known distribution, as in the male voice categorization, perceptual learning for that voice occurred using the relatively abundant unambiguous examples for that speaker to adjust the perceptual system for that talker.

Kraljic and Samuel (2005), in addition to the findings reported in in Section 1.2, also found an asymmetry in how listeners generalized perceptual learning. Across all of their experiments, categorization of the same voice as exposure produced perceptual learning. When the continuum from /s/ to /ʃ/ from the male voice was categorized following exposure to the female voice, perceptual learning effects were consistently present. When the voice of the exposure and categorization was reversed, the consistency of perceptual learning effects disappeared. They performed an acoustic analysis of the spectral means for the categorization items and the exposure items for each talker. The spectral means of the voices for the continuum steps were well separated, but the exposure items were much closer together. The spectral mean of the ambiguous sibilant for the female voice fell in between the ranges of the female continuum steps and the male continuum steps, but the spectral mean exposure ambiguous sibilant for the male voice was squarely in the range of the male continuum steps. The asymmetry of speaker generalization can then be attributed to the female exposure ambiguous items being acoustically (and therefore perceptually) similar enough to generalize the perceptual learning for the female voice to the male voice.

From these studies, the clearest thread through all of them is the degree of acoustic similarity or perceptual similarity between the exposure stimuli that prompts perceptual learning and whatever continuum the listeners are tested on. Stops and fricative do not actually appear to behave categorically different, per se, but their differences in behaviour can be attributed to the fact that stops are acoustically more similar to each other in general than fricatives are.

1.3.2 Generalizing to other categories and contexts

Kraljic and Samuel (2006) looked at two sound contrasts (/d/-/t/ and /b/-/p/) that shared a feature, namely voicing. Participants were exposed to ambiguous in between /d/ and /t/ in English words and then tested on two continua with two different voices. The categorization continuum that was presented first, from /b/ to /p/, showed the stronger perceptual learning effects than the second continuum from /d/ to /t/, even

though participants were exposed to ambiguous tokens between /d/ and /t/. Both old voices and new voices for the categorization task produced perceptual learning effects.

Mitterer et al. (2013) exposed participants to either words ending with /r/ or /l/ or to words beginning with /r/ or /l/, and tested each group’s perceptual learning behaviour on continua from /r/ to /l/ in both initial and final positions. If listeners were exposed to an ambiguous sound between /r/ and /l/ at the ends of words, they showed a perceptual learning effect for the continuum that ended with an /r/ or /l/, but not for the continuum that began with an /r/ or an /l/. The inverse behaviour was observed for the group that was exposed to an ambiguous sound between /r/ and /l/ at the beginnings of words. They argue that the category being affected is not a context-insensitive phoneme, but rather a context sensitive allophone. Jesse and McQueen (2011) found that listeners did indeed generalize across positions, but their stimuli were the fricatives /s/ and /f/, and the same physical fricatives were the same across all positions.

Reinisch et al. (2014) utilized a visually-guided recalibration paradigm from Bertelson et al. (2003) to test whether acoustic cues to place of articulation could be recalibrated with unambiguous visual feedback. In Experiment 1, they trained listeners by exposing one group to the maximally ambiguous token between /aba/ and /ada/ with either a video of /ada/ or /aba/ and another group to the maximally ambiguous token between /ibi/ and /idi/ with either a video of /ibi/ or /idi/. In vowel context of /a/, the consonant portion of the token was silenced so that only formant cues to place of articulation would be available. In the vowel context of /i/, the consonant part was not silenced, but since formant transitions in the /i/ context are the same for labials and dentals, the only cues to place of articulation would be in the consonant itself. Afterwards, when they categorized auditory only continua of /aba/ to /ada/ and /ibi/ to /idi/ with the same silencing of the consonant in the /a/ context. They found a recalibration effect for the continuum participants were exposed to, but no generalization effect to the novel continuum. Experiment 2 followed the same setup as Experiment 1, with the same /aba/ to /ada/ as in Experiment 1, but replaced the /ibi/ to /idi/ continuum with a /ama/ to /ana/ continuum. As in Experiment 1, clear recalibration effects were found for the continuum participants were exposed to, but not for the novel continuum. Experiment 3 followed the same procedure but used a /ubu/ to /udu/ continuum instead of an /ibi/ to /idi/ continuum. The /u/ context provides formant transition cues to place of articulation, so the consonant was silenced for this continuum as well. As before, there were perceptual recalibration effects for the continuum participants were exposed to, but not to the novel continuum. Across all of these, there is a context specificity effect.

Kraljic et al. (2008a) exposed participants to ambiguous sibilants between /s/ and /ʃ/ in two different contexts. In one, the ambiguous sibilants were intervocalic, and in the other, they occurred as part of a /str/ cluster. Participants exposed to the ambiguous sound intervocalically showed a perceptual learning effect, while those exposed to the sibilants in /str/ environments did not. The sibilant in /str/ often surfaces closer to [ʃ] in many varieties of English, due to coarticulatory effects from the other consonants in the cluster, but the coarticulatory effects for merging /s/ and /ʃ/ are much weaker in intervocalic position. They argue that the interpretation of the ambiguous sound is

done in context of the surrounding sounds, and only when the pronunciation variant is unexplainable from context is the variant learned and attributed to the speaker, see also Kraljic et al. (2008b). In addition to a continuum of /asi/ to /afi/, they also tested a continuum from /astri/ to /aftri/, and found comparable perceptual learning effects across both continua for those exposed to the intervocalic ambiguous sibilants, but no perceptual learning effects on either continua for the other condition, showing a context insensitivity absent in other studies.

1.4 Current contribution

Perceptual learning effects require two important aspects to be involved in the exposure phase. There must be some ambiguous acoustic aspect to be learned, and there must be an unambiguous link between the ambiguous acoustics and a sound category. This link can be provided through bottom-up information, such as in the visual domain (Bertelson et al., 2003), or it can be top-down information, such as the lexical status of the tokens (Norris et al., 2003). Most of the literature within the domain of lexically-guided perceptual learning has been on the presence or absence of perceptual learning in the categorization phases, with manipulations to the categorization phase to test for generalization. The question that I am interested in is in the linking of ambiguous tokens to sound categories. Can manipulations in the exposure phase that reduce the reliability of the linking cause significant differences in perceptual learning?

In this dissertation, I will induce manipulations of lexical bias and attention in Experiments 1 and 2, and manipulations of sentence predictability and attention in Experiment 3. Lexical bias has been shown to affect phoneme categorization tasks (Ganong, 1980), and can be manipulated by position of the ambiguous sound in the word and attention Pitt and Szostak (2012). Sentence predictability, has likewise been found to affect phoneme categorization tasks similar to lexical bias (Borsky et al., 1998), and can be manipulated by the preceding words in the sentence (Kalikow et al., 1977). The interaction of sentence predictability with attention has not been explicitly studied, but the interaction should be similar to lexical bias.

2 Experiments 1 and 2 motivation

2.1 Lexical bias

Lexical bias is the primary way through which perceptual learning occurs in the experimental literature. Lexical bias, also known as the Ganong Effect, refers to the tendency for listeners to interpret a speaker’s production as a meaningful word rather than a nonsense word. For instance, given a continuum from a nonword like *dask* to word like *task* that differs only in one sound, listeners in general are more likely to interpret any step along the continuum as the word endpoint rather than the nonword endpoint (Ganong, 1980). This bias is exploited in perceptual learning studies to allow for noncanonical, ambiguous productions of a sound to be linked to pre-existing sound categories.

Studies looking at lexical bias use a phoneme categorization task, where participants identify a sound at the beginning or end of a word from two possible options that vary in one dimension. For instance, given a continuum from /t/ to /d/, participants are asked to identify the sound at the beginning or end as either /t/ or /d/. Lexical bias effects are calculated based on two continua, where words are formed at opposite ends. For the /t/ to /d/ continua, one would form a word at the /t/ end, such as *task* and one would form a word at the other end, such as *dash*. Lexical bias effects are then calculated from the different categorization behaviour for these two continua.

Lexical bias varies in strength according to several factors. First, the length of the word in syllables has a large effect on lexical bias, with longer words showing stronger lexical bias than shorter words (Pitt and Samuel, 2006). Continua formed using trisyllabic words, such as *establish* and *malpractice*, were found to show consistently large lexical bias effects than monosyllabic words, such as *kiss* and *fish*. Pitt and Samuel (2006) also found that lexical bias from monosyllabic words show greater response to experimental manipulations than trisyllabic words.

Pitt and Szostak (2012) used a lexical decision task with a continuum of fricatives from /s/ to /ʃ/ embedded in words differing in the position of a sibilant. They found that ambiguous fricatives earlier in the word, such as *serenade* or *chandelier*, lead to greater nonword responses than the same ambiguous fricatives embedded later in a word, such as *establish* or *embarrass*. In the experiments and meta-analysis of phoneme identification results presented in Pitt and Samuel (1993), they found that, for monosyllabic word frames, token-final targets produce more stable results than token-initial targets.

Pitt and Szostak (2012) additionally investigated the role of attention in modulating lexical bias. When listeners were told that the speaker's /s/ and /ʃ/ were ambiguous and to listen carefully to ensure correct responses, they were less tolerant of noncanonical productions across all positions in the word. That is, participants attending to the speaker's sibilants were less likely to accept the modified production as a word than participants given no particular instructions about the sibilants.

Lexical bias is affected by processing conditions. In their first experiment, Mattys and Wiget (2011) found that lexical bias has a larger effect when listeners have more cognitive load than when they have no cognitive load from concurrent tasks. In their second experiment, the same cognitive load and non-cognitive load conditions were present, but listeners had to respond to the phoneme identification task within 1500 ms. In this experiment, the lexical bias across cognitive load conditions was identical and similar to the cognitive load condition of experiment 1. In experiment 3, listeners could only respond *after* 1500 ms, and the effect of cognitive load re-emerged with similar magnitude to experiment 1. In the remainder of their experiments, they examined whether cognitive load increased word activation or decreased sensory information, the two primary possible explanations for the pattern of results found in experiment 1. They found that semantic priming was not significantly different across cognitive loads, but fine-grained discrimination suffered under cognitive load conditions. Thus, the effect of cognitive load observed is a function of impoverished fine detail leading to a greater reliance on lexical information. However, the timecourse for the incorporation of these different sources

of information is interesting, with lexical information playing a larger role earlier on in processing and sensory signal information being incorporated later on. Earlier responses are more likely to be lexically biased than later responses.

3 Experiment 3 motivation

3.1 Semantic predictability

The type of contextual predictability used in this experiment is known as semantic predictability. Sentences are highly predictable when they contain words prior to the final word that points almost definitively to the identity of the final word. For instance, the sentence fragment *The cow gave birth to the...* from Kalikow et al. (1977) is almost guaranteed to be completed with the word *calf*. On the other hand, a fragment like *She is glad Jane called about the...* is far from having a guaranteed completion, other than having the category of noun. Semantic predictability has been found to have effects on both production and perception, which will be discussed in the two following sections

3.1.1 Acoustic realization

In general, semantic predictability has a reductive effect on the acoustics of words. In Scarborough (2010), words produced in highly predictable frames tend to be shorter in duration and have less dispersed vowel realizations. This semantic predictability effect did not interact with neighbourhood density, a lexical factor that proxies for the amount of lexical competition a word has. Words with many neighbours, and therefore had lessened lexical predictability, had longer durations and more dispersed vowel realizations. For both the lexical and the semantic predictability, high predictability led to less distinct word realizations, and low predictability led to more distinct word realizations, independently of each other.

In a study looking at semantic predictability across dialects, Clopper and Pierrehumbert (2008) found that not all dialects realize the effects of semantic predictability the same. For the Southern dialect of American English, the results were much the same as in Scarborough (2010), showing temporal and spectral reduction in high predictability environments. However, speakers in the Midland dialect showed no such effect, and speakers of the Northern dialect showed more extreme Northern Cities shifting in the high predictability environment.

3.1.2 Intelligibility

Despite the temporal and spectral reduction found in high predictability contexts, high predictability sentences are generally more intelligible. Sentences that form a semantically coherent whole have higher word identification rates across varying signal-to-noise ratios (Kalikow et al., 1977). However, when words at the ends of predictive sentences are excised from their context, they tend to be less intelligible than words excised from non-predictive contexts (Lieberman, 1963).

3.2 Similarity to lexical bias

A key pillar of the motivation underlying this experiment is the notion that semantic predictability is similar to, yet distinct from, lexical bias/lexical predictability. Scarborough (2010) found that lexical predictability and semantic predictability did not interact and had the same effect directions, but different effect sizes, with lexical predictability having a larger effect than semantic predictability.

One line of research has attempted to show whether the semantic predictability effects are part of perceptual and phonetic processing or if they result following phonetic processing. Connine and Clifton (1987) established different reaction time profiles for perceptual and postperceptual processes in categorizing a continuum. With lexical bias, response times to the end points of a /d/ to /t/ continuum show no difference whether the continuum forms a word at one end point (such as *dice* to *tice*) or at the other (such as *dype* to *type*). However, at the boundary between /d/ and /t/, reaction times are faster when a subject responds consistent to the bias (i.e. interprets an ambiguous word *?ice* as *dice*). For postperceptual processes, in this case a monetary payoff for responding either /d/ or /t/ along a continuum that has nonwords at both ends, the pattern is reversed, where reaction times were faster for responses consistent with the monetary bias at the end points of the continuum, but no such difference was found for the category boundary. Both biases produced similar categorization patterns, such that the participants biased toward /d/, either lexically or monetarily, categorized more of the continuum as /d/, so the principle difference between the two biases was in the reaction time profile.

Two studies that have looked at semantic predictability in this paradigm are Connine (1987) and Borsky et al. (1998), and they found conflicting results. In Connine (1987), the reaction time profile aligned more with the monetary bias in Connine and Clifton (1987), with reaction time differences located at the end point and not the category boundary, but Borsky et al. (1998) found the reverse with a profile more similar to lexical bias.

3.3 Attention?

I don't know of any work that has specifically looked at attention (to acoustics in particular) in the context of semantic predictability.

References

- Bertelson, P., Vroomen, J., and De Gelder, B. (2003). Visual Recalibration of Auditory Speech Identification: A McGurk Aftereffect. *Psychological Science*, 14(6):592–597.
- Borsky, S., Tuller, B., and Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America*, 103(5 Pt 1):2670–2676.

- Clopper, C. G. and Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, 124(3):1682–1688.
- Connine, C. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 538:527–538.
- Connine, C. M. and Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of experimental psychology. Human perception and performance*, 13(2):291–299.
- Eisner, F. and McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2):224–238.
- Eisner, F. and McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4):1950.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1):110–125.
- Jesse, A. and McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception.
- Kalikow, D., Stevens, K., and Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. ... *Journal of the Acoustical Society of America*, 61(5).
- Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008a). Accommodating variation: dialects, idiolects, and speech processing. *Cognition*, 107(1):54–81.
- Kraljic, T. and Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, 51(2):141–78.
- Kraljic, T. and Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2):262–268.
- Kraljic, T. and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1):1–15.
- Kraljic, T., Samuel, A. G., and Brennan, S. E. (2008b). First impressions and last resorts: how listeners adjust to speaker variability. *Psychological science*, 19(4):332–8.
- Ladefoged, P. and Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1):98–104.
- Lieberman, P. (1963). Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech. *Language and Speech*, 6(3):172–187.

- Mattys, S. L. and Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, 65(2):145–160.
- Mitterer, H., Scharenborg, O., and McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, 129(2):356–361.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech.
- Pitt, M. and Szostak, C. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, (April 2013):37–41.
- Pitt, M. A. and Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of experimental psychology. Human perception and performance*, 19(4):699–725.
- Pitt, M. A. and Samuel, A. G. (2006). Word length and lexical activation: longer is better. *Journal of experimental psychology. Human perception and performance*, 32(5):1120–1135.
- Reinisch, E. and Holt, L. L. (2013). Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. *Journal of experimental psychology. Human perception and performance*, 40(2):539–555.
- Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of experimental psychology. Human perception and performance*, 39(1):75–86.
- Reinisch, E., Wozny, D. R., Mitterer, H., and Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, 45:91–105.
- Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. *Laboratory phonology*, pages 575–604.
- van Linden, S. and Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of experimental psychology. Human perception and performance*, 33(6):1483–1494.