CHAPTER 1

Speech Perception

FRANK EISNER AND JAMES M. MCQUEEN

INTRODUCTION

What Speech Is

Speech is the most acoustically complex type of sound that we regularly encounter in our environment. The complexity of the signal reflects the complexity of the movements that speakers perform with their tongues, lips, jaws, and other articulators in order to generate the sounds coming out of their vocal tract. Figure 1.1 shows two representations of the spoken sentence ***The sun melted the snow***—an oscillogram at the top, showing variation in amplitude, and a spectrogram at the bottom, showing its spectral characteristics over time. The figure illustrates some of the richness of the information contained in the speech signal: There are modulations of amplitude, detailed spectral structures, noises, silences, bursts, and sweeps. Some of this structure is relevant in short temporal windows at the level of individual phonetic segments. For example, the vowel in the word ***sun*** is characterized by a certain spectral profile, in particular the location of peaks in the spectrum (called "formants," the darker areas in the spectrogram). Other structures are relevant at the level of words or phrases. For example, the end of the utterance is characterized by a fall in amplitude and in pitch, which spans several segments. The acoustic cues that describe the identity of segments such as individual vowels and consonants are referred to as *segmental* information, whereas the cues that span longer stretches of the signal such as pitch and amplitude envelope and that signal prosodic structures such as syllables, feet, and intonational phrases are called *suprasegmental*.

Acoustic cues are transient and come in fast. The sentence in Figure 1.1 is spoken at a normal speech rate; it contains five syllables and is only 1.3 seconds long. The average duration of a syllable in the sentence is about 260 ms, meaning that information about syllable identity comes in on average at a rate of about 4 Hz, which is quite stable across languages (Giraud & Poeppel, 2012). In addition to the linguistic information that is densely packed in the speech signal, the signal also contains a great deal of additional information about the speaker, the so-called paralinguistic content of speech. If we were to listen to a recording of this sentence, we would be able to say with a fairly high degree of certainty that the speaker is a British middle-aged man with an upper-class accent, and we might also be able to guess that he is suffering from a cold and perhaps is slightly bored as he recorded the prescribed phrase. Paralinguistic

FE is supported by the Gravitation program "Language in Interaction" from the Dutch Science Foundation (NWO).

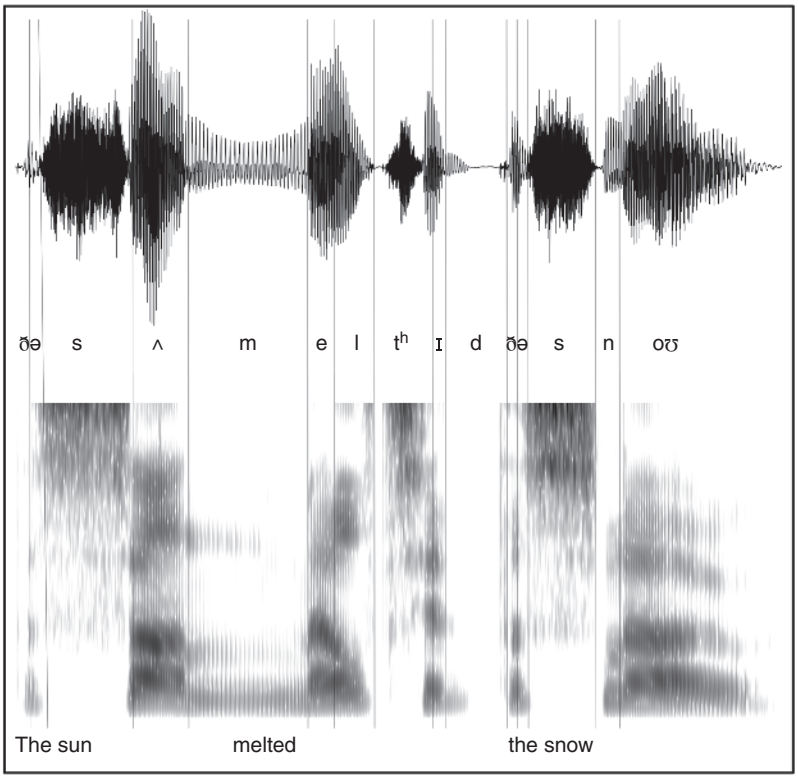

Figure 1.1 Oscillogram (top) and spectrogram (bottom) representations of the speech signal in the sentence "The sun melted the snow," spoken by a male British English speaker. The vertical lines represent approximate phoneme boundaries with phoneme transcriptions in the International Phonetic Alphabet (IPA) system. The oscillogram shows variation in amplitude (vertical axis) over time (horizontal axis). The spectrogram shows variation in the frequency spectrum (vertical axis) over time (horizontal axis); higher energy in a given part of the spectrum is represented by darker shading.

information adds to the complexity of speech, and in some cases interacts with how linguistic information is interpreted by listeners (Mullennix & Pisoni, 1990).

What Speech Perception Entails

How, then, is this complex signal perceived? In our view, speech perception is not primarily about how listeners identify individual speech segments (vowels and consonants), though of course this is an important part of the process. Speech perception is also not primarily about how listeners identify suprasegmental units such as syllables and lexical stress patterns, though this is an often overlooked part of the process, too. Ultimately, speech perception is about how listeners use combined sources of segmental and suprasegmental information to recognize spoken words. This is because the listener's goal is to grasp what a speaker means, and the only way she or he can do so is through recognizing the individual meaning units in the speaker's utterance: its morphemes and words. Perceiving segments and prosodic structures is thus at the service of word recognition.

The nature of the speech signal poses a number of computational problems that the

listener has to solve in order to be able to recognize spoken words (cf. Marr, 1982). First, listeners have to be able to recognize words in spite of considerable variability in the signal. The oscillogram and spectrogram in Figure 1.1 would look very different if the phrase had been spoken by a female adolescent speaking spontaneously in a casual conversation on a mobile phone in a noisy ski lift, and yet the same words would need to be recognized. Indeed, even if the same speaker recorded the same sentence a second time, it would be physically different (e.g., a different speaking rate, or a different fundamental frequency).

Due to coarticulation (the vocal tract changing both as a consequence of previous articulations and in preparation for upcoming articulations), the acoustic realization of any given segment can be strongly colored by its neighboring segments. There is thus no one-to-one mapping between the perception of a speech sound and its acoustics. This is one of the main factors that is still holding back automatic speech recognition systems (Benzeghiba et al., 2007). In fact, the perceptual system has to solve a many-to-many mapping problem, because not only do instances of the same speech sound have different acoustic properties, but the same acoustic pattern can result in perceiving different speech sounds, depending on the context in which the pattern occurs (Nusbaum & Magnuson, 1997; Repp & Liberman, 1987). The surrounding context of a set of acoustic cues thus has important implications on how the pattern should be interpreted by the listener.

There are also continuous speech processes through which sounds are added (a process called epenthesis), reduced, deleted, or altered, rendering a given word less like its canonical pronunciation. One example of such a process is given in Figure 1.1: The /n/ of ***sun*** is realized more like an [m], through a process called coronal place assimilation whereby the coronal /n/ approximates the labial place of articulation of the following word-initial [m].

Speech recognition needs to be robust in the face of all this variability. As we will argue, listeners appear to solve the variability problem in multiple ways, but in particular through phonological abstraction (i.e., categorizing the signal into prelexical segmental and suprasegmental units prior to lexical access) and through being flexible (i.e., through perceptual learning processes that adapt the mapping of the speech signal onto the mental lexicon in response to particular listening situations).

The listener must also solve the segmentation problem. As Figure 1.1 makes clear, the speech signal has nothing that is the equivalent of the white spaces between printed words as in a text such as this that reliably mark where words begin and end. In order to recognize speech, therefore, listeners have to segment the quasicontinuous input stream into discrete words. As with variability, there is no single solution to the segmentation problem: Listeners use multiple cues, and multiple algorithms.

A third problem derives from the fact that, across the world's languages, large lexica (on the order of perhaps 50,000 words) are built from small phonological inventories (on the order of 40 segments in a language such as English, and often much fewer than that; Ladefoged & Maddieson, 1996). Spoken words thus necessarily sound like other spoken words: They begin like other words, they end like other words, and they often have other words partially or wholly embedded within them. This means that, at any moment in the temporal unfolding of an utterance, the signal is likely to be partially or wholly consistent with many words. Once again, the listener appears to solve this "lexical embedding" problem using multiple algorithms.

We will argue that speech perception is based on several stages of processing at which a variety of perceptual operations help the listener solve these three major computational challenges—the variability problem, the segmentation problem, and the lexical embedding problem (see Box 1.1). These stages and operations have been studied over the past 70 years or so using behavioral techniques (e.g., psychophysical tasks such as identification and discrimination; psycholinguistic procedures such as lexical decision, cross-modal priming, and visual-world eye tracking); and neuroscientific techniques (especially measures using electroencephalography [EEG] and magnetoencephalography [MEG]). Neuroimaging techniques (primarily functional magnetic resonance imaging [fMRI]) and neuropsychological approaches (based on aphasic patients) have also made it possible to start to map these stages of processing onto brain regions. In the following section we will review data of all these different types. These data have made it possible to specify at least three core stages of processing involved in speech perception and the kinds of operations involved at each stage. The data also provide some suggestions about the neural instantiation of these stages.

As shown in Figure 1.2, initial operations act to distinguish incoming speech-related acoustic information from non-speech-related acoustic information. Thereafter, prelexical processes act in parallel to extract segmental and suprasegmental information from the speech signal (see Box 1.2). These processes contribute toward solving the variability and segmentation problems and serve to facilitate spoken-word recognition. Lexical processing receives input from segmental and suprasegmental prelexical processing and continues to solve the first two computational problems while also solving the lexical-embedding problem.

Box 1.1 Three Computational Challenges

1. The variability problem

The physical properties of any given segment can vary dramatically because of a variety of factors such as the talker's physiology, accent, emotional state, or speech rate. Depending on such contextual factors, the same sound can be perceived as different segments, and different sounds can be perceived as the same segment. The listener has to be able to recognize speech in spite of this variability.

2. The segmentation problem

In continuous speech there are no acoustic cues that reliably and unambiguously mark the boundaries between neighboring words or indeed segments. The boundaries are often blurred because neighboring segments tend to be coarticulated (i.e., their pronunciation overlaps in time) and because there is nothing in the speech stream that is analogous to the white spaces between printed words. The listener has to be able to segment continuous speech into discrete words.

3. The lexical-embedding problem

Spoken words tend to sound like other spoken words: They can begin in the same way (e.g., ***cap*** and ***cat***), they can end in the same way (e.g., ***cap*** and ***map***), and they can have other words embedded within them (e.g., ***cap*** in ***captain***). This means that at any point in time the speech stream is usually (at least temporarily) consistent with multiple lexical hypotheses. The listener has to be able to recognize the words the speaker intended from among those hypotheses.

Finally, processing moves beyond the realm of speech perception. Lexical processing provides input to interpretative processing, where syntactic, semantic, and pragmatic operations, based on the words that have been recognized, are used to build an interpretation of what the speaker meant.

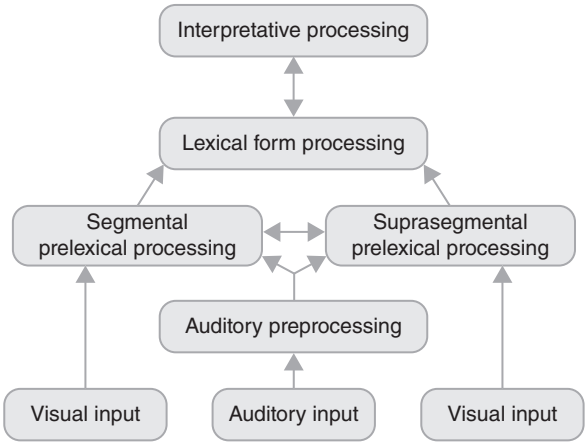

Figure 1.2 Processing stages in speech perception. Arrows represent on-line flow of information during the initial processing of an utterance.

Box 1.2 Three Processing Stages

1. Segmental prelexical processing

Phonemes are the smallest linguistic units that can indicate a difference in meaning. For example, the words ***cap*** and ***cat*** differ by one consonant, /p/ versus /t/, and ***cap*** and ***cup*** differ by one vowel, /æ/ vs. /ʌ/. Phoneme-sized segments are also perceptual categories, though it is not yet clear whether listeners recognize phonemes or some other units of perception (e.g., syllables or position-specific allophones, such as the syllable-initial [p] in ***pack*** vs. the syllable-final [p] in ***cap***). We therefore use the more neutral term *segments*. The speech signal contains acoustic cues to individual segments. Segmental prelexical processing refers to the computational processes acting on segmental information that operate prior to retrieval of words from long-term memory and that support that retrieval process.

2. Suprasegmental prelexical processing

The speech signal contains acoustic cues for a hierarchy of prosodic structures that are larger than individual segments, including syllables, prosodic words, lexical stress patterns, and intonational phrases. These structures are relevant for the perception of words. For example, the English word ***forbear*** is pronounced differently depending on whether it is a verb or a noun even though the segments are the same in both words. The difference is marked by placing stress on the first or second syllable, which can for example be signaled by an increase in loudness and/or duration. Suprasegmental prelexical processing refers to the computational processes acting on suprasegmental information that operate prior to retrieval of words from long-term memory and that support that retrieval process.

3. Lexical form processing

To understand a spoken utterance, the listener must recognize the words the speaker intended. Lexical form processing refers to the computational processes that lead to the recognition of words as phonological forms (as opposed to processes that determine the meanings associated with those forms). The listener considers multiple perceptual hypotheses about the word forms that are currently being said (e.g., ***cap***, ***cat***, ***apt***, and ***captain*** given the input ***captain***). Output from the segmental and suprasegmental prelexical stages directs retrieval of these hypotheses from long-term lexical memory. Together with contextual constraints, it also influences the selection and recognition of words from among those hypotheses.

STAGES OF PERCEPTUAL PROCESSING

Auditory Preprocessing

The sounds we encounter in our environment are converted in the inner ear from physical vibrations to electrical signals that can be interpreted by the brain. From the ear, sound representations travel along the ascending auditory pathways via several subcortical nuclei to the auditory cortex. Along the way, increasingly complex representations in the spectral and temporal domains are derived from the waveform, coding aspects of the signal such as the amplitude envelope, onsets and offsets, amplitude modulation frequencies, spectral structure, and modulations of the frequency spectrum (Theunissen & Elie, 2014). These representations are often topographically organized, for example in tonotopic "maps" that show selective sensitivity for particular frequencies along a spatial dimension (e.g., Formisano et al., 2003). There is evidence for processing hierarchies in the ascending auditory system (e.g., Eggermont, 2001). For example, whereas auditory events are represented at a very high temporal resolution subcortically, the auditory cortex appears to integrate events into longer units that are more relevant for speech perception (Harms & Melcher, 2002). Similarly, subcortical nuclei have been found to be sensitive to very fast modulations of the temporal envelope of sounds, but the auditory cortex is increasingly sensitive to the slower modulations such as the ones that correspond to prelexical segments in speech (Giraud & Poeppel, 2012; Giraud et al., 2000).

The notion of a functional hierarchy in sound processing, and speech in particular, has also been proposed for the primary auditory cortex and surrounding areas. A hierarchical division of the auditory cortex underlies the processing of simple to increasingly complex sounds both in nonhuman primates (Kaas & Hackett, 2000; Perrodin, Kayser, Logothetis, & Petkov, 2011; Petkov, Kayser, Augath, & Logothetis, 2006; Rauschecker & Tian, 2000) and in humans (e.g., Binder et al., 1997; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Obleser & Eisner, 2009; Scott & Wise, 2004). Two major cortical streams for processing speech have been proposed, extending in both antero-ventral and postero-dorsal directions from primary auditory cortex (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000; Scott & Johnsrude, 2003; Ueno, Saito, Rogers, & Lambon Ralph, 2011). The anterior stream in the left hemisphere in particular has been attributed with decoding linguistic meaning in terms of segments and words (Davis & Johnsrude, 2003; DeWitt & Rauschecker, 2012; Hickok & Poeppel, 2007; Scott, Blank, Rosen, & Wise, 2000). The anterior stream in the right hemisphere appears to be less sensitive to linguistic information (Scott et al., 2000), but more sensitive to speaker identity and voice processing (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Perrodin et al., 2011), as well as to prosodic speech cues, such as pitch (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015). The subcortical auditory system thus extracts acoustic cues from the waveform that are relevant for speech perception, whereas speech-specific processes begin to emerge in regions beyond the primary auditory cortex (Overath, McDermott, Zarate, & Poeppel, 2015).

Prelexical Segmental Processing

Neural systems that appear to be specific to speech processing relative to other types of complex sounds are mostly localized to the auditory cortex and surrounding regions in the perisylvian cortex (see Figure 1.3). Several candidate regions in the superior

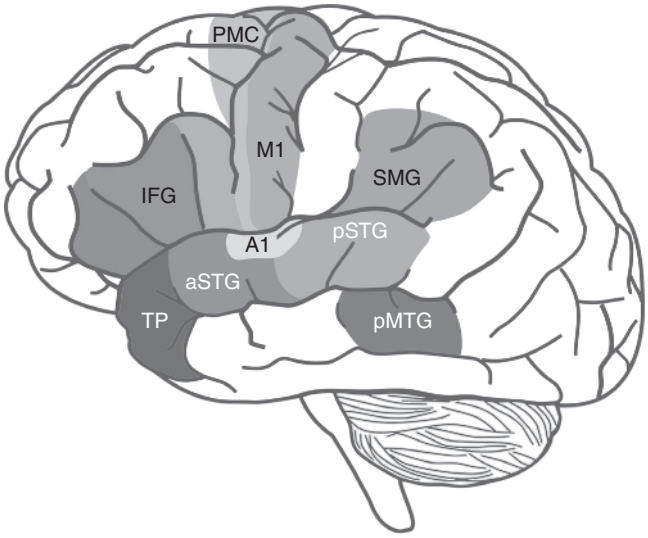

Figure 1.3 Lateral view of the left hemisphere showing the cortical regions that are central in speech perception. A1, primary auditory cortex; TP, temporal pole; aSTG, anterior superior temporal gyrus; pSTG, posterior superior temporal gyrus; pMTG, posterior middle temporal gyrus; SMG, supramarginal gyrus; M1, primary motor cortex; PMC, premotor cortex; IFG, inferior frontal gyrus. Color version of this figure is available at http://onlinelibrary.wiley.com/book/10.1002/9781119170174.

temporal cortex and the inferior parietal cortex (Chan et al., 2014; Obleser & Eisner, 2009; Turkeltaub & Coslett, 2010) have been shown to be engaged in aspects of processing speech at a prelexical level of analysis (Arsenault & Buchsbaum, 2015; Mesgarani, Cheung, Johnson, & Chang, 2014). Neural populations in these regions exhibit response properties that resemble hallmarks of speech perception, such as categorical perception of segments (Liebenthal, Sabri, Beardsley, Mangalathu-Arumana, & Desai, 2013; Myers, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009). Bilateral regions of the superior temporal sulcus have recently been shown to be selectively tuned to speech-specific spectrotemporal structure (Overath et al., 2015). Many processing stages in the ascending auditory pathways feature a topographic organization, which has led to studies probing whether a phonemic map exists in the superior temporal cortex. However, the current evidence suggests that prelexical units have complex, distributed cortical representations (Bonte, Hausfeld, Scharke, Valente, & Formisano, 2014; Formisano, De Martino, Bonte, & Goebel, 2008; Mesgarani et al., 2014).

The main computational problems to be addressed during prelexical processing are the segmentation and variability problems. The segmentation problem is not only a lexical one. There are no reliably marked boundaries between words in the incoming continuous speech stream, but there are also no consistent boundaries between individual speech sounds. Whereas some types of phonemes have a relatively clear acoustic structure (stop consonants, for instance, are signaled by a period of silence and a sudden release burst, which have a clear signature in the amplitude envelope; fricatives are characterized by high-frequency noise with a sudden onset), other types of phonemes, such as vowels, approximants, and nasals, are distinguished predominantly by their formant structure, which changes relatively slowly. The final word ***snow*** in Figure 1.1

illustrates this. There is a clear spectrotemporal signature for the initial /s/, whereas the boundaries in the following sequence /noʊ/ are much less clear. Prelexical processes segment the speech signal into individual phonological units (e.g., between the /s/ and the /n/ of ***snow***) and provide cues for lexical segmentation (e.g., the boundary between ***melted*** and ***the***).

Recent studies on neural oscillations have suggested that cortical rhythms may play an important role in segmenting the speech stream into prelexical units. Neural oscillations are important because they modulate the excitability of neural networks; the peaks and troughs in a cycle influence how likely neurons are to fire. Interestingly, oscillations in the theta range (4–8 Hz) align with the quasiperiodic amplitude envelope of an incoming speech signal. Giraud and Poeppel (2012) have suggested that this entrainment of auditory networks to speech rhythm serves to segment the speech stream into syllable-sized portions for analysis. Each theta cycle may then in turn trigger a cascade of higher-frequency oscillations, which analyze the phonetic contents of a syllable chunk on a more fine-grained time scale (Morillon, Liégeois-Chauvel, Arnal, Bénar, & Giraud, 2012).

Psycholinguistics has not yet identified one single unit of prelexical representation into which the speech stream is segmented. In addition to phonemes (McClelland & Elman, 1986), features (Lahiri & Reetz, 2002), allophones (Mitterer, Scharenborg, & McQueen, 2013), syllables (Church, 1987), and articulatory motor programs (Galantucci, Fowler, & Turvey, 2006) have all been proposed as representational units that mediate between the acoustic signal and lexical representations. There may indeed be multiple units of prelexical representation that capture regularities in the speech signal at different levels of granularity (Mitterer et al., 2013; Poellmann, Bosker, McQueen, & Mitterer, 2014; Wickelgren, 1969). The oscillations account is generally compatible with this view, since different representations of the same chunk of speech may exist simultaneously on different timescales. This line of research in speech perception is relatively new, and there are questions about whether the patterns of neural oscillations are a causal influence on or a consequence of the perceptual analysis of speech. Some evidence for a causal relationship comes from a study that showed that being able to entrain to the amplitude envelope of speech results in increased intelligibility of the signal (Doelling, Arnal, Ghitza, & Poeppel, 2014), but the mechanisms by which this occurs are still unclear.

Oscillatory entrainment may also assist listeners in solving the lexical segmentation problem, since syllable and segment boundaries tend to be aligned with word boundaries. Other prelexical segmental processes also contribute to lexical segmentation. In particular, prelexical processing appears to be sensitive to the transitional probabilities between segments (Vitevitch & Luce, 1999). These phonotactic regularities provide cues to the location of likely word boundaries. For example, a characteristic of Finnish that is known as *vowel harmony* regulates which kinds of vowels can be present within the same word. This kind of phonotactic knowledge provides useful constraints on where in the speech stream boundaries for particular words can occur, and Finnish listeners appear to be sensitive to those constraints (Suomi, McQueen, & Cutler, 1997). Regularities concerning which sequences of consonants can occur within versus between syllables (McQueen, 1998), or which sequences are more likely to be at the edge of a word (van der Lugt, 2001), also signal word boundary locations.

After segmentation, the second major computational challenge addressed at the prelexical stage is how the perception system deals with the ubiquitous variability in the speech signal. Variability is caused by a number of different sources, including speech rate, talker differences, and continuous speech processes such as assimilation and reduction.

Speech Rate

Speech rate varies considerably within as well as between talkers, and has a substantial effect on the prelexical categorization of speech sounds (e.g., Miller & Dexter, 1988). This is especially the case for categories that are marked by a temporal contrast, such as voice-onset time (VOT) for stop consonants. VOT is the most salient acoustic cue to distinguish between English voiced and unvoiced stops, and thus between words such as ***cap*** and ***gap***. However, what should be interpreted as a short VOT (consistent with ***gap***) or a long VOT (consistent with ***cap***) is not a fixed duration, but depends on the speech rate of the surrounding phonetic context (Allen & Miller, 2004; Miller & Dexter, 1988). Speech rate may even influence whether segments are perceived at all: Dilley and Pitt (2010) showed that listeners tended not to perceive the function word ***or*** in a phrase such as ***leisure or time*** when the speech was slowed down, whereas they did perceive it at a normal rate. Conversely, when the speech was speeded up, participants tended to perceive the function word when it was not actually part of the utterance.

Being able to adapt to changes in speaking rate is thus crucial for prelexical processing, and it has been known for some time that listeners are adept at doing so (Dupoux & Green, 1997), even if the underlying mechanisms are not yet clear. There is evidence that adaptability to varying speech rates is mediated not only by auditory but also by motor systems (Adank & Devlin, 2010), possibly by making use of internal forward models (e.g., Hickok, Houde, & Rong, 2011), which may help to predict the acoustic consequences of faster or slower motor sequences. There is an emerging body of research that shows that neural oscillations in the auditory cortex align to speech rate fluctuations (Ghitza, 2014; Peelle & Davis, 2012). It has yet to be established whether this neural entrainment is part of a causal mechanism that tunes in prelexical processing to the current speech rate.

Talker Differences

A second important source of variability in speech acoustics arises from physiological differences between talkers. Factors like body size, age, and vocal tract length can strongly affect acoustic parameters such as fundamental frequency and formant dispersion, which are critical parameters that encode differences between many speech sound categories. It has been known for decades that even when vowels are spoken in isolation and under laboratory conditions, there is a great amount of overlap in the formant measures (peaks in the frequency spectrum that are critical for the perception of vowel identity) for different speakers (Adank, Smits, & Hout, 2004; Peterson & Barney, 1952). In other words, formant values measured when a given speaker produces one particular vowel may be similar to when a different speaker produces a different vowel. Formant values thus need to be interpreted in the context of acoustic information that is independent of what the speaker is saying, specifically acoustic information about more general aspects of the speaker's physiology.

It has also been known for a long time that listeners do this (Ladefoged, 1989; Ladefoged & Broadbent, 1957), and the specifics of the underlying mechanisms are beginning to become clear. The perceptual system

appears to compute an average spectrum for the incoming speech stream that can be used as a model of the talker's vocal tract properties, and also can be used as a reference for interpreting the upcoming speech (Nearey, 1989; Sjerps, Mitterer, & McQueen, 2011a). Evidence from an EEG study (Sjerps, Mitterer, & McQueen, 2011b) shows that this extrinsic normalization of vowels takes place early in perceptual processing (around 120 ms after vowel onset), which is consistent with the idea that it reflects prelexical processing. Behavioral and neuroimaging evidence suggests that there are separate auditory systems that are specialized in tracking aspects of the speaker's voice (Andics et al., 2010; Belin et al., 2000; Formisano et al., 2008; Garrido et al., 2009; Kriegstein, Smith, Patterson, Ives, & Griffiths, 2007; Schall, Kiebel, Maess, & Kriegstein, 2015). These right-lateralized systems appear to be functionally connected to left-lateralized systems that are preferentially engaged in processing linguistic information, which may indicate that these bilateral systems work together in adjusting prelexical processing to speaker-specific characteristics (Kriegstein, Smith, Patterson, Kiebel, & Griffiths, 2010; Schall et al., 2015).

Listeners not only use the talker information that is present in the speech signal on-line, but also integrate adaptations to phonetic categories over longer stretches and store these adapted representations in long-term memory for later use (Norris, McQueen, & Cutler, 2003). Norris et al. demonstrated that listeners can adapt to a speaker who consistently articulates a particular speech sound in an idiosyncratic manner. The researchers did this by exposing a group of listeners to spoken Dutch words and nonwords in which an ambiguous fricative sound (/sf?/, midway between /s/ and /f/) replaced every /s/ at the end of 20 critical words (e.g., in ***radijs***, "radish"; note that ***radijf*** is not a Dutch word). A second group heard the same ambiguous sound in words ending in /f/ (e.g., ***olijf***, "olive"; ***olijs*** is not a Dutch word). Both groups could thus use lexical context to infer whether /sf?/ was meant to be an /s/ or an /f/, but that context should lead the two groups to different results. Indeed, when both groups categorized sounds on an /s/–/f/ continuum following exposure, the group in which /sf?/ had replaced /s/ categorized more ambiguous sounds as /s/, whereas the other group categorized more sounds as /f/. This finding suggests that the perceptual system can use lexical context to learn about a speaker's idiosyncratic articulation, and that this learning affects prelexical processing later on. A recent fMRI study, using a similar paradigm, provided converging evidence for an effect of learning on prelexical processing by locating perceptual learning effects to the superior temporal cortex, which is thought to be critically involved in prelexical decoding of speech (Myers & Mesite, 2014). This kind of prelexical category adjustment can be guided not only by lexical context, but also by various other kinds of language-specific information, such as phonotactic regularities (Cutler, McQueen, Butterfield, & Norris, 2008), contingencies between acoustic features that make up a phonetic category (Idemaru & Holt, 2011), or sentence context (Jesse & Laakso, 2015).

A critical feature of this type of perceptual learning is that it entails phonological abstraction. Evidence for this comes from demonstrations that learning generalizes across the lexicon, from the words heard during initial exposure to new words heard during a final test phase (Maye, Aslin, & Tanenhaus, 2008; McQueen, Cutler, & Norris, 2006; Reinisch, Weber, & Mitterer, 2013; Sjerps & McQueen, 2010). If listeners apply what they have learned about the fricative /f/, for example, to the on-line recognition of other words that have an /f/ in them, this

suggests first that listeners have abstract knowledge that /f/ is a phonological category and second that these abstract representations have a functional role to play in prelexical processing. Thus, although the nature of the unit of prelexical representation is still an open question, as discussed earlier, these data suggest that there is phonological abstraction prior to lexical access.

Several studies have investigated whether category recalibration is speaker-specific or speaker-independent by changing the speaker between the exposure and test phases. This work so far has produced mixed results, sometimes finding evidence of generalization across speakers (Kraljic & Samuel, 2006, 2007; Reinisch & Holt, 2014) and sometimes evidence of speaker specificity (Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Reinisch, Wozny, Mitterer, & Holt, 2014). The divergent findings might be partly explained by considering the perceptual similarity between tokens from the exposure and test speakers (Kraljic & Samuel, 2007; Reinisch & Holt, 2014). When there is a high degree of similarity in the acoustic-phonetic properties of the critical segment, it appears to be more common that learning transfers from one speaker to another. In sum, there is thus evidence from a variety of sources that speaker-specific information in the signal affects prelexical processing, both by using the speaker information that is available online, and by reusing speaker-specific information that was stored previously.

Accents

Everybody has experienced regional or foreign accents that alter segmental and suprasegmental information so drastically that they can make speech almost unintelligible. However, although they are a further major source of variability in the speech signal, the way in which accents deviate from standard pronunciations is regular; that is, the unusual sounds and prosody tend to occur in a consistent pattern. Listeners can exploit this regularity and often adapt to accents quite quickly. Processing gains have been shown to emerge after exposure to only a few accented sentences, as an increase in intelligibility (Clarke & Garrett, 2004) or as a decrease in reaction times in a comprehension-based task (Weber, Di Betta, & McQueen, 2014).

An important question is whether the perceptual system adapts to an accent with each individual speaker, or whether an abstract representation of that accent can be formed that might benefit comprehension of novel talkers with the same accent. Bradlow and Bent (2008) investigated this question by looking at how American listeners adapt to Chinese-accented English. Listeners were exposed to Chinese-accented speech coming either from only one speaker or from several different speakers. Following exposure, generalization was assessed in an intelligibility task with Chinese-accented speech from an unfamiliar speaker. Intelligibility increased in both conditions during training, but evidence of generalization to the novel speaker was found only after exposure to multiple speakers. This pattern suggests that the perceptual system can form an abstract representation of an accent when the accent is shared between several different speakers, which can in turn affect how speech from other speakers with the same accent is processed. Learning also generalized to different speech materials that were used in training and test, which is consistent with the notion that learned representations of speech patterns can affect perception at the prelexical level.

Continuous Speech Processes

Another aspect of variability tackled by the prelexical processor is that caused by continuous speech processes, including the coronal place assimilation process shown in Figure 1.1 (where the final segment of

sun becomes [m]-like because of the following word-initial [m] of ***melted***). Several studies have shown that listeners are able to recognize assimilated words correctly when the following context is available (Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996, 1998; Gow, 2002; Mitterer & Blomert, 2003). Different proposals have been made about how prelexical processing could act to undo the effects of assimilation, including processes of phonological inference (Gaskell & Marslen-Wilson, 1996, 1998) and feature parsing (Gow, 2002; feature parsing is based on the observation that assimilation tends to be phonetically incomplete, such that, e.g., in the sequence ***sun melted*** the final segment of ***sun*** has some features of an [m] but also some features of an [n]). The finding that Dutch listeners who speak no Hungarian show similar EEG responses (i.e., mismatch negativity responses) to assimilated Hungarian speech stimuli to those of native Hungarian listeners (Mitterer, Csépe, Honbolygo, & Blomert, 2006) suggests that at least some forms of assimilation can be dealt with by relatively low-level, language-universal perceptual processes. In other cases, however, listeners appear to use language-specific phonological knowledge to cope with assimilation (e.g., Weber, 2001).

There are other continuous speech processes, such as epenthesis (adding a sound that is not normally there, e.g., the optional insertion of the vowel /ə/ between the /l/ and /m/ of ***film*** in Scottish English), resyllabification (changing the syllabic structure; e.g., /k/ in ***look at you*** might move to the beginning of the syllable /kət/ when it would normally be the final sound of /lʊk/), and liaison (linking sounds; e.g., in some British English accents ***car*** is pronounced /ka/, but the /r/ resurfaces in a phrase like ***car alarm***). Language-specific prelexical processes help listeners cope with these phenomena. For instance, variability can arise due to reduction processes (where a segment is realized in a simplified way or may even be deleted entirely). It appears that listeners cope with reduction both by being sensitive to the fine-grained phonetic detail in the speech signal and through employing knowledge about the phonological contexts in which segments tend to be reduced (Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009b).

Multimodal Speech Input

Spoken communication takes place predominantly in face-to-face interactions, and the visible articulators convey strong visual cues to the identity of prelexical segments. The primary networks for integrating auditory and visual speech information appear to be located around the temporoparietal junction, in posterior parts of the superior temporal gyrus, and in the inferior parietal lobule (supramarginal gyrus and angular gyrus; Bernstein & Liebenthal, 2014). The well-known McGurk effect (McGurk & MacDonald, 1976) demonstrated that auditory and visual cues are immediately integrated in segmental processing, by showing that a video of a talker articulating the syllable /ba/ combined with an auditory /ga/ results in the fused percept of /da/. The influence of visual processing on speech perception is not limited to facial information; text transcriptions of speech can also affect speech perception over time (Mitterer & McQueen, 2009a).

Visual cues can also drive auditory recalibration in situations where ambiguous auditory information is disambiguated by visual information: When perceivers repeatedly heard a sound that could be either /d/ or /b/, presented together with a video of a speaker producing /d/, their phonetic category boundary shifted in a way that was consistent with the information they received through lipreading, and the ambiguous sound was

assimilated into the /d/ category. However, when the same ambiguous sound was presented with the speaker producing /b/, the boundary shift occurred in the opposite direction (Bertelson, Vroomen, & de Gelder, 2003; Vroomen & Baart, 2009). Thus, listeners can use information from the visual modality to recalibrate their perception of ambiguous speech input, in this case long-term knowledge about the co-occurrence of certain visual and acoustic cues.

Fast perceptual learning processes already modulate early stages of cortical speech processing. Kilian-Hütten et al. (Kilian-Hütten, Valente, Vroomen, & Formisano, 2011; Kilian-Hütten, Vroomen, & Formisano, 2011) have demonstrated that early acoustic-phonetic processing is already influenced by recently learned information about a speaker idiosyncrasy. Using the visually guided perceptual recalibration paradigm (Bertelson et al., 2003), regions of the primary auditory cortex (specifically, Heschl's gyrus and sulcus, extending into the planum temporale) could be identified whose activity pattern specifically reflected listeners' adjusted percepts after exposure to a speaker, rather than simply physical properties of the stimuli. This suggests not only a bottom-up mapping of acoustical cues to perceptual categories in the left auditory cortex, but also that the mapping involves the integration of previously learned knowledge within the same auditory areas—in this case, coming from the visual system. Whether linguistic processing in the left auditory cortex can be driven by other types of information, such as speaker-specific knowledge from the right anterior stream, is an interesting question for future research.

Links Between Speech Perception and Production

The motor theory of speech perception was originally proposed as a solution to the variability problem (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Given the inherent variability of the speech signal and the flexibility of perceptual categories, the source of invariance may be found in articulatory representations instead. According to this view, decoding the speech signal requires recovering articulatory gestures through mental emulation of the talker's articulatory commands to the motor system. The motor theory received support following the discovery of the mirror neuron system (Fadiga, Craighero, & D'Ausilio, 2009; Galantucci et al., 2006) and from neuroscience research that shows effects on speech processing during disruption of motor systems (e.g., Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Yuen, Davis, Brysbaert, & Rastle, 2010). However, the strong version of the theory, in which the involvement of speech motor areas in speech perception is obligatory, is not universally accepted (Hickok et al., 2011; Lotto, Hickok, & Holt, 2009; Massaro & Chen, 2008; Scott, McGettigan, & Eisner, 2009; Toni, de Lange, Noordzij, & Hagoort, 2008). The main arguments against motor theory are that lesions in the motor cortex do not result in comprehension deficits, that comprehension can occur in individuals who are unable to articulate, and that the motor cortex is not typically activated in fMRI studies using passive-listening tasks. Behavioral evidence against motor theory comes from an experiment on speech shadowing (Mitterer & Ernestus, 2008): Participants were not slower to repeat out loud a spoken stimulus when there was a gestural mismatch between the stimulus and the response than when there was a gestural match.

According to the contrasting auditory perspective, decoding the speech signal requires an analysis of acoustic cues that map onto multidimensional phonetic categories, mediated by general auditory mechanisms

(Hickok & Poeppel, 2007; Holt & Lotto, 2010; Obleser & Eisner, 2009; Rauschecker & Scott, 2009). A purely auditory perspective, however, fails to account for recent evidence from transcranial magnetic stimulation (TMS) studies showing that disruption of (pre-)motor cortex can have modulatory effects on speech perception in certain situations (D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013; Meister et al., 2007; Möttönen, Dutton, & Watkins, 2013). If motor systems are not necessary for speech perception, what might be the functionality that underlies these modulatory effects? It is noteworthy that such effects have been observed only at the phoneme or syllable level, that they appear to be restricted to situations in which the speech signal is degraded, and that they affect reaction times rather than accuracy (Hickok et al., 2011).

Although sensorimotor interactions in perception are not predicted by traditional auditory approaches, several neurobiological models of language processing have begun to account for perception–production links (Guenther, Ghosh, & Tourville, 2006; Hickok, 2012; Hickok et al., 2011; Rauschecker & Scott, 2009). From a speech production point of view, perceptual processes are necessary in order to establish internal models of articulatory sequences during language acquisition, as well as to provide sensory feedback for error monitoring. There is recent evidence from fMRI studies that the premotor cortex might facilitate perception, specifically under adverse listening conditions, because activity in motor areas has been linked to perceptual learning of different types of degraded speech (Adank & Devlin, 2010; Erb, Henry, Eisner, & Obleser, 2013; Hervais-Adelman, Carlyon, Johnsrude, & Davis, 2012). Such findings are consistent with the idea that motor regions provide an internal simulation that matches degraded speech input to articulatory templates, thereby assisting speech comprehension under difficult listening conditions (D'Ausilio et al., 2012; Hervais-Adelman et al., 2012), but direct evidence for this is lacking at present.

Summary

The prelexical segmental stage involves speech-specific processes that mediate between general auditory perception and word recognition by constructing perceptual representations that can be used during lexical access. The two main computational challenges approached at this stage are the segmentation and variability problems. We have argued that listeners use multiple prelexical mechanisms to deal with these challenges, including the detection of phonotactic constraints for lexical segmentation, processes of rate and talker normalization and of phonological inference, and engagement of speech production machinery (at least under adverse listening conditions). The two most important prelexical mechanisms, however, appear to be abstraction and adaptation. The central goal of the prelexical processor is to map from the episodic detail of the acoustic input onto abstract perceptual categories in order to be able to cope with the variability problem and hence to facilitate lexical access. This mapping process clearly seems to be adaptive: Listeners tune in to aspects of the current listening situation (e.g., who is/are talking, how fast they are talking, whether they have a foreign or regional accent). Studying perceptual learning in particular has been valuable as a window into how prelexical perceptual representations are maintained and updated.

Prelexical Suprasegmental Processing

As we have already argued, speech perception depends on the extraction of suprasegmental

as well as segmental information. Suprasegmental material is used by listeners to help them solve the lexical-embedding, variability, and segmentation problems. As with prelexical segmental processing, abstraction and adaptation are the two main mechanisms that allow listeners to solve these problems.

Words can have the same segments but differ suprasegmentally. One way in which the listener copes with the lexical-embedding problem (the fact that words sound like many other words) is thus to use these fine-grained suprasegmental differences to disambiguate between similar-sounding words. Italian listeners, for instance, can use the relative duration of segments to distinguish between alternative lexical hypotheses that have the same initial sequence of segments but different syllabification (e.g., the syllable-final /l/ of ***sil.vestre***, "sylvan," differs minimally in duration from the syllable-initial /l/ of ***si.lencio***, "silence"), and fragment priming results suggest that Italians can use this acoustic difference to disambiguate the input even without hearing the following disambiguating segments (i.e., the /v/ or /ɛ/; Tabossi, Collina, Mazzetti, & Zoppello, 2000).

English listeners use similar subtle durational cues to syllabic structure to disambiguate oronyms (***tulips*** vs. ***two lips***; Gow & Gordon, 1995); Dutch listeners use /s/ duration to distinguish between, for example, ***een spot***, "a spotlight," and ***eens pot***, "once jar" (Shatzman & McQueen, 2006b); and French listeners use small differences in the duration of consonants to distinguish between sequences with liaison (e.g., the word-final /r/ of ***dernier*** surfacing in ***dernier oignon***, "last onion") from matched sequences without liaison (e.g., ***dernier rognon***, "last kidney"; Spinelli, McQueen, & Cutler, 2003).

Durational differences across multiple segments also signal suprasegmental structure. Monosyllabic words, for example, tend to be longer than in the same segmental sequence in a polysyllabic word (e.g., ***cap*** is longer on its own than in ***captain***; Lehiste, 1972). Experiments using a variety of tasks, including cross-modal priming, eye tracking, and mouse tracking, have shown that listeners use these durational differences during word recognition, and thus avoid recognizing spurious lexical candidates (such as ***cap*** in ***captain***; Blazej & Cohen-Goldberg, 2015; Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). It appears that these effects reflect the extraction of suprasegmental structure because they are modulated by cues to other prosodic structures. Dutch listeners in an eye-tracking study looked more at a branch (a ***tak***) when hearing the longer word ***taxi*** if the cross-spliced ***tak*** came from an original context where the following syllable was stressed (e.g., /si/ in ***pak de tak sinaasappels***, "grab the branch of oranges") than if it was unstressed (/si/ in ***pak de tak citroenen***, "grab the branch of lemons"; Salverda et al., 2003).

Listeners also make use of cues to larger suprasegmental structures to disambiguate between words. The presence of the onset of a larger suprasegmental structure (e.g., an intonational phrase) affects the pronunciation of the segment that happens to be at that boundary (typically by making it longer and louder). This information can be used during lexical form processing to disambiguate between several word candidates (Keating, Cho, Fougeron, & Hsu, 2003). Cho, McQueen, and Cox (2007) examined temporarily ambiguous sequences in English such as ***bus tickets***, where words such as ***bust*** straddle the word boundary. The word ***bus*** was easier to recognize in the phrase ***bus tickets*** if it had been taken from the utterance "***When you get on the bus, tickets should be shown to the driver***" (in which the /t/ was prosodically strengthened) than if it had been taken from "***John bought several bus tickets for his family***" (in which the /t/ was

not strengthened). Christophe, Peperkamp, Pallier, Block, and Mehler (2004) found a similar effect in French. Words such as ***chat***, "cat," were harder to disambiguate from ***chagrin***, "grief," in the sequence ***chat grinchaux***, "grumpy cat," if the sequence was part of a single phrase than if a phrase boundary occurred between the two words.

Listeners also use suprasegmental cues to the lexical stress patterns of words during word recognition. These cues include pitch, amplitude, and duration differences between stressed and unstressed syllables. Dutch (Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005) and Spanish (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001) listeners are sensitive to differences between sequences that are segmentally identical but differ in stress, and use those differences to constrain lexical access (e.g., Dutch listeners can distinguish between ***voor*** taken from initially stressed ***voornaam***, "first name," and ***voor*** taken from finally stressed ***voornaam***, "respectable"; Cutler & van Donselaar, 2001). Dutch listeners use the stress information as soon as it is heard during word recognition: Eye-tracking data show disambiguation between, for example, ***oktober***, "October" (stress on the second syllable) and ***octopus***, "octopus" (stress on the first syllable) before the arrival of unambiguous segmental information (the /b/ and /p/ in this example; Reinisch, Jesse, & McQueen, 2010). Italian listeners show similar rapid use of stress information in on-line word recognition (Sulpizio & McQueen, 2012).

Interestingly, however, English listeners tend to be less sensitive to stress cues than Dutch, Spanish, and Italian listeners; across a variety of tasks, stress effects are weak and can be hard to find in English (Cooper, Cutler, & Wales, 2002; Fear, Cutler, & Butterfield, 1995; Slowiaczek, 1990). This appears to be because stress in English is primarily cued by differences between segments (the difference between full vowels and the reduced vowel schwa) rather than suprasegmental stress differences. This means that English listeners are usually able to distinguish between words using segmental information alone and hence can afford to ignore the suprasegmental information (Cooper et al., 2002; see Cutler, 2012 for further discussion). English participants (Scarborough, Keating, Mattys, Cho, & Alwan, 2009) and Dutch participants (Jesse & McQueen, 2014) are also sensitive to visual cues to lexical stress (e.g., chin or eyebrow movements).

Obviously, suprasegmental stress information can be used in speech perception only in a language that has lexical stress. Similarly, other types of suprasegmental cues can be used only in languages that make lexical distinctions based on those cues, but the cross-linguistic evidence suggests that such cues are indeed used to constrain word recognition. Speakers of languages with lexical tone, such as Mandarin and Cantonese, for example, use tone information in word recognition. Note that tone is sometimes regarded as segmental, since a vowel with one f0 pattern (e.g., a falling tone) can be considered to be a different segment from the same vowel with a different pattern (e.g., a level tone). We consider tone to be suprasegmental here, however, because it concerns an acoustic feature, pitch, which signals other suprasegmental distinctions (e.g., lexical stress). Lexical priming studies in Cantonese suggest, for example, that tonal information modulates word recognition (Cutler & Chen, 1997; Lee, 2007; Ye & Connine, 1999; Yip, 2001). Likewise, pitch-accent patterns in Japanese (based on high [H] and low [L] syllables, again cued by differences in the f0 contour) are picked up by Japanese listeners; for example, they can distinguish between /ka/ taken from ***baka*** [HL] versus ***gaka*** [LH] (Cutler & Otake, 1999), and accent patterns are used to distinguish between

words (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999).

The data previously reviewed all make the same general point about how listeners solve the lexical-embedding problem. Listeners cope with the fact that words sound like other words in part by using suprasegmental disambiguating information. Suprasegmental prelexical processing thus entails the extraction of this information so that it can be used in lexical processing. This can be also be considered to be a way in which listeners solve the variability problem. Segments have different physical realizations in different prosodic and intonational contexts (e.g., they are longer, or louder, or have higher pitch). The suggestion here is that this kind of variability is dealt with by suprasegmental prelexical processes, which use this information to build phonologically abstract prosodic structures that are then used to constrain word recognition.

As with segmental prelexical processing, therefore, abstraction is a key mechanism that allows listeners to cope with variability. Word-learning experiments provide evidence for suprasegmental abstraction. In Shatzman and McQueen (2006a), Dutch listeners were taught pairs of novel words, such as ***bap*** and ***baptoe***, that were analogues of real pairs such as ***cap*** and ***captain***. The listeners had to learn to associate the new words with nonsense shapes. Critically, during learning, the durational difference between the monosyllabic novel words and the same syllable in the longer words was neutralized. In an eye-tracking test phase, however, the syllables had their normal duration (***bap*** was longer than the ***bap*** in ***baptoe***). Even though the listeners had never heard these forms before, effects of the durational differences (analogous to those found in eye tracking with real words) were observed (e.g., listeners made more fixations to the ***bap*** nonsense shape when the input syllable was longer than when it was shorter). This suggests that the listeners had abstract knowledge about the durational properties of monosyllabic and polysyllabic words and could bring that knowledge to bear during word recognition the first time they heard the novel words with those properties. A word-learning experiment with a similar design (Sulpizio & McQueen, 2012) suggests that Italian listeners have abstract suprasegmental knowledge about lexical stress (about the distribution of lexical stress patterns in Italian, and about the acoustic-phonetic cues that signal stress), and that they too can use that knowledge during online recognition of novel words, in spite of never having heard those words with those stress cues ever before.

A perceptual learning experiment using the lexically guided retuning paradigm of Norris et al. (2003) also provides evidence for suprasegmental abstraction. Mandarin listeners exposed to syllables with ambiguous pitch contours in contexts that biased the interpretation of the ambiguous syllables toward either tone 1 or tone 2 subsequently categorized more stimuli on tone 1–tone 2 test continua in a way that was consistent with the exposure bias (Mitterer, Chen, & Zhou, 2011). This tendency was almost as large for new test words as for words that had been heard during exposure. This generalization of learning indicates that the listeners had adjusted phonologically abstract knowledge about lexical tone. Generalization of perceptual learning across the lexicon about the pronunciation of syllables also indicates that listeners have abstract knowledge about suprasegmental structure (Poellmann et al., 2014).

Suprasegmental information also has a role to play in solving the segmentation problem. The studies previously reviewed on uptake of fine-grained suprasegmental cues (Blazej & Cohen-Goldberg, 2015; Cho et al., 2007; Christophe et al., 2004; Davis et al.,

2002; Gow & Gordon, 1995; Salverda et al., 2003; Spinelli et al., 2003) can all also be considered as evidence for the role of these cues in segmentation. The fine-grained detail is extracted prelexically and signals word boundaries.

But there is also another important way in which suprasegmental prelexical processing supports lexical segmentation. The rhythmic structure of speech can signal the location of word boundaries (Cutler, 1994). Languages differ rhythmically, and the segmentation procedures vary across languages accordingly. In languages such as English and Dutch, rhythm is stress-based, and strong syllables (i.e., those with full vowels, which are distinct from the reduced vowels in weak syllables) tend to mark the locations of the onsets of new words in the continuous speech stream (Cutler & Carter, 1987; Schreuder & Baayen, 1994). Listeners of such languages are sensitive to the distinction between strong and weak syllables (Fear et al., 1995), and use this distinction to constrain spoken-word recognition, as measured by studies examining word-boundary misperceptions (Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2013; Cutler & Butterfield, 1992; Vroomen, van Zon, & de Gelder, 1996) and in word-spotting tasks (Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Vroomen et al., 1996; Vroomen & de Gelder, 1995). Cutler and Norris (1988), for example, compared word-spotting performance for target words such as ***mint*** in ***mintayf*** (where the second syllable was strong) and ***mintef*** (where the second syllable was weak). They found poorer performance in sequences such as ***mintayf***, and argued that this was because the strong syllable—***tayf***—indicated that there was likely to be a new word starting at the /t/, which then made it harder to spot ***mint***.

Languages with different rhythms are segmented in different ways. Languages such as French, Catalan, and Korean have rhythm based on the syllable, and speakers of these languages appear to use syllable-based segmentation procedures (Content, Meunier, Kearns, & Frauenfelder, 2001; Cutler, Mehler, Norris, & Segui, 1986, 1992; Kim, Davis, & Cutler, 2008; Kolinsky, Morais, & Cluytens, 1995; Sebastián-Gallés, Dupoux, Segui, & Mehler, 1992). Likewise, languages such as Japanese and Telugu have rhythm based on the mora, and speakers of these languages appear to use mora-based segmentation procedures (Cutler & Otake, 1994; Murty, Otake, & Cutler, 2007; Otake, Hatano, Cutler, & Mehler, 1993). In spite of these differences across languages, what appears to be common is that segmentation uses rhythm.

Summary

The prelexical suprasegmental stage acts in parallel with the prelexical segmental stage to construct speech-specific representations of suprasegmental structures that can be used to constrain and assist lexical access. Multiple mechanisms at this stage of processing help the listener to solve all three major computational problems. As with prelexical segmental processing, the key mechanisms in suprasegmental processing are abstraction and adaptation. There has been relatively little work using neuroscientific methods to address the nature of prelexical suprasegmental processing.

Lexical Form Processing

Although it is broadly established that prelexical processes and representations are instantiated in the superior temporal lobes, there is less consensus about the localization of lexical processing (see, e.g., Price, 2012). In some neurobiological models, the primary pathway from prelexical processes to word forms and meaning is along the antero-ventral stream (DeWitt &

Rauschecker, 2012; Rauschecker & Scott, 2009; Scott et al., 2000; Ueno et al., 2011), interfacing with semantic and conceptual representations in the temporal poles (e.g., Rice, Lambon Ralph, & Hoffman, 2015). Several other neurobiological models postulate that the lexicon consists of interconnected networks containing different types of representation such as surface forms, abstract phonological forms, an auditory–motor interface, or a semantic interface, and which are spatially distributed across the temporal and inferior parietal lobes (Davis, 2016; Gow, 2012; Hickok & Poeppel, 2007).

Three major lexical processing streams have been proposed in the literature. Starting in the mid-superior temporal cortex, one stream runs in an antero-ventral direction along the superior temporal gyrus, one in a posteriodorsal direction along the temporoparietal junction to the supramarginal gyrus, and one in a posterioventral direction via pSTG and pMTG to the posterior inferior temporal gyrus. Whether all of these streams are essential for lexical processing in speech recognition and how they might work together in binding different types of lexical representations remain open questions. We suggest that studying learning processes may provide an opportunity to move forward in localizing lexical processes.

Spoken-word recognition is characterized by two key processes: the parallel evaluation of multiple lexical hypotheses, and competition among those hypotheses. Together, these two processes allow the listener to solve the lexical-embedding problem. There is a substantial body of converging evidence for both processes.

Evidence for the simultaneous evaluation of multiple word hypotheses comes, for example, from cross-modal priming (Zwitserlood, 1989; Zwitserlood & Schriefers, 1995), eye-tracking (Allopenna, Magnuson, & Tanenhaus, 1998; Huettig & McQueen, 2007; Yee & Sedivy, 2006), and EEG experiments (van Alphen & Van Berkum, 2010, 2012). Because words sound like other words (i.e., because of the lexical-embedding problem), listeners need to consider overlapping hypotheses of many different types. Words beginning like other words are considered in parallel (e.g., in Dutch, ***kapitaal***, "capital," when the onset of ***kapitein***, "captain," is heard; Zwitserlood, 1989; see also Allopenna et al., 1998; Huettig & McQueen, 2007), as are words embedded in the onset of longer words (e.g., ***cap*** in ***captain***; Davis et al., 2002; Salverda et al., 2003; van Alphen & Van Berkum, 2010, 2012). Words embedded in the offset of longer words are also considered when the longer word is heard (e.g., ***bone*** in ***trombone***; Isel & Bacri, 1999; Luce & Cluff, 1998; Shillcock, 1990; van Alphen & Van Berkum, 2010, 2012; Vroomen & de Gelder, 1997). The evidence is weaker for offset embeddings than for onset embeddings (see, e.g., Luce & Lyons, 1999), presumably because of the temporal nature of the speech signal (there is already strong support for the longer word before there is any support for the offset embedding). Embedded words may be stronger candidates for recognition when the speech signal is higher in quality (Zhang & Samuel, 2015). Lexical hypotheses are also considered that span word boundaries in the input (e.g., ***visite***, "visits," given the input ***visi tediati***, "faces bored"; Tabossi, Burani, & Scott, 1995; see also Cho et al., 2007; Gow & Gordon, 1995).

The strength of different hypotheses is determined in part by their goodness of fit to the available speech input. The phonetic similarity between an intended word (e.g., ***cabinet***; Connine, Titone, Deelman, & Blasko, 1997) and a mispronounced nonword (e.g., ***gabinet*** vs. ***mabinet*** vs. ***shuffinet***) influences how much the mispronunciation disrupts lexical access. The more similar the

mismatching sound and the intended sound are, the greater the support for the intended word (Connine et al., 1997). Once again, there is converging evidence of this across tasks: phoneme monitoring (Connine et al., 1997) and cross-modal priming (Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & van Halen, 1996). The relative intolerance of the recognition system to mismatching segmental information is one way in which it deals with the lexical-embedding problem. Words that do not fit the input very well are not considered as serious lexical hypotheses. This assumption is central to the Shortlist model (Norris, 1994) and gives it its name: Only the best matching candidates enter the shortlist for recognition.

Selection among hypotheses appears to be based not only on goodness of fit. Lexical hypotheses compete with each other, as shown by increasing response latencies in word-recognition tasks as competition intensifies. As the number and frequency of similar-sounding words in the lexical neighborhood of a word increase, it becomes harder to recognize that word (Cluff & Luce, 1990; Luce & Large, 2001; Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce, 1998, 1999). Gaskell and Marslen-Wilson (2002) showed, in a priming experiment, that the number of words beginning in the same way as a prime word (or word fragment) influenced the size of the resulting priming effect. There thus appears to be competition among words that begin in the same way. There is also competition among words starting at different points in the speech input. In a word-spotting task, listeners find it harder to spot a word in a nonsense sequence that is the onset of a real word (e.g., ***mess*** in ***domes***) and hence where there is competition with that real word (***domestic***) than in a matched nonsense sequence that is not the onset of a real word (e.g., ***mess*** in ***nemess***; McQueen et al., 1994). The number of words beginning later in the speech signal than the target word also influences word-spotting performance (Norris et al., 1995; Vroomen & de Gelder, 1995).

Lexical competition plays a key role not only in solving the lexical-embedding problem but also in solving the segmentation problem. In the absence of any signal-based cues to word boundaries, competition can nevertheless produce a lexical parse of continuous speech: The best-matching words (wherever they may begin or end) win the competition, and hence the input is segmented (McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997). As previously reviewed, however, there are multiple segmental and suprasegmental cues to possible word boundaries in the continuous speech stream, and these appear to be extracted during prelexical processing. Possible electrophysiological markers of lexical segmentation have been documented (Sanders, Newport, & Neville, 2002). It appears that the relative roles of lexical and signal-based factors in segmentation change in different listening situations (e.g., in the context of different amounts of background noise; Mattys, White, & Melhorn, 2005; Newman, Sawusch, & Wunnenberg, 2011). Mattys et al. (2005) have suggested, for example, that lexical knowledge (e.g., whether the context of a target word was another word or a nonword) tends to matter more in segmentation than do signal-based segmental cues (e.g., whether segments and their contexts were coarticulated), and these cues in turn tend to be more important than signal-based suprasegmental cues (e.g., whether stimuli began with strong or weak syllables).

How, then, might these different types of cues and lexical constraints jointly determine segmentation? Norris et al. (1997) proposed the possible-word constraint (PWC) as a unifying segmentation algorithm.

According to the PWC account, lexical hypotheses are evaluated as to whether they are aligned with likely word boundaries, as cued by any of the signal-based cues. If not, those hypotheses are disfavored. A word is considered to be misaligned if there is no vocalic portion between the word's edge (its beginning or its end) and the location of the likely word boundary. Cross-linguistically, a residue of speech without a vowel cannot itself be a possible word, and so a parse involving that residue and a lexical hypothesis is very unlikely to be what the speaker intended (e.g., if the input is ***clamp***, it is improbable than the speaker intended ***c lamp*** because [k] on its own is not a possible word of English). Evidence for the PWC has now been found in many languages, including English (Newman et al., 2011; Norris et al., 1997; Norris, McQueen, Cutler, Butterfield, & Kearns, 2001), Dutch (McQueen, 1998), Japanese (McQueen, Otake, & Cutler, 2001), Sesotho (Cutler, Demuth, & McQueen, 2002), Cantonese (Yip, 2004), and German (Hanulíková, Mitterer, & McQueen, 2011). Evidence for the PWC has also been found in Slovak (Hanulíková, McQueen, & Mitterer, 2010), in spite of the fact that Slovak (like other Slavic languages) permits words without vowels (but only for a small number of consonants, those functioning as closed-class words; these consonants are treated as a special case in Slovak segmentation). The only language tested to date for which no evidence for the PWC has been found is Berber (El Aissati, McQueen, & Cutler, 2012), a language that has many words without vowels. Although speakers of Berber appear not to use the PWC (it would be disadvantageous for them to do so), speakers of all other languages seem to benefit from this segmentation algorithm.

Lexical processing also has a role to play in solving the variability problem. Evidence previously reviewed suggests that prelexical processes of abstraction (about segments and about suprasegmental structures), and perceptual learning mechanisms acting on those abstractions, have a major role to play in dealing with speech variability. But especially when the listener has to deal with extreme forms of variability, as when the pronunciation of a word deviates substantially from its canonical form, lexical processing can step in. More specifically, it appears that some pronunciation variants of words are stored in the mental lexicon. When Dutch listeners have to recognize that [tyk] is a form of the word ***natuurlijk***, "of course," for example, it appears they do so by storing that form rather than through prelexical processes that reconstruct the canonical form (Ernestus, 2014; Ernestus, Baayen, & Schreuder, 2002). Support for the view that lexical storage can help deal with pronunciation variability concerning not only extreme forms of reduction, but also other forms of variability (e.g., that ***gentle*** in American English can be produced either with a medial [nt] or with a medial nasal flap; Ranbom & Connine, 2007), comes from evidence of effects on word recognition of the frequency of occurrence of particular pronunciation variants (Connine, 2004; Connine, Ranbom, & Patterson, 2008; Pitt, Dilley, & Tat, 2011; Ranbom & Connine, 2007).

If, as discussed earlier, prelexical segmental and suprasegmental processing entails phonological abstraction, then lexical form representations must be abstract, too, rather than episodic in nature. Experiments on novel-word learning also support this view. Lexical competition between a newly learned word and its existing phonological neighbors can be used as an index that it has been integrated into the mental lexicon (as, e.g., when the new word ***cathedruke*** starts to compete with ***cathedral***, slowing responses to ***cathedral***; Gaskell & Dumay, 2003). Different behavioral measures of

competition (and other measures that new words have been added to the lexicon; Leach & Samuel, 2007) have indicated that lexical integration tends to be a gradual process that is enhanced by sleep (Dumay & Gaskell, 2007, 2012) and can take several days to complete, though some data suggest it can occur without sleep (Kapnoula & McMurray, 2015; Lindsay & Gaskell, 2013; Szmalec, Page, & Duyck, 2012).

Integration of a new words into the lexicon appears to reflect transfer from initially episodic representations to phonologically abstract representations, as shown, for instance, by evidence that new words learned only in printed form (i.e., as print episodes that have never been heard) nonetheless start to compete with spoken words (Bakker, Takashima, van Hell, Janzen, & McQueen, 2014). This transfer process is consistent with the complementary learning systems account of memory consolidation (Davis & Gaskell, 2009; McClelland, McNaughton, & O'Reilly, 1995). In line with that account, the emergence of lexical competition appears to parallel a shift from episodic memory in medial temporal lobe structures (the hippocampus in particular) to lexical memory in neocortical structures, including the pMTG (Davis & Gaskell, 2009; Takashima, Bakker, van Hell, Janzen, & McQueen, 2014; see also Breitenstein et al., 2005). Lexicalization can also be tracked by measuring EEG oscillatory activity: There are differences in theta band (4–8 Hz) power between existing words and novel words that have not been learned, but no such differences for novel words that have been learned the previous day (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015).

Summary

The three major computational challenges faced by the listener—the variability problem, the segmentation problem, and the lexical-embedding problem—must all ultimately be resolved at the stage of lexical form processing. These problems appear to be solved through parallel evaluation of multiple lexical hypotheses, the use of segmental and suprasegmental information that constrains the lexical search to only the most likely hypotheses and that indicates the location of likely word boundaries, and competition among those hypotheses. The PWC is a segmentation algorithm that appears to further modulate this competition process. It appears that lexical form representations are phonologically abstract rather than episodic, and that multiple (abstract) pronunciation variants of the same word can be stored. In keeping with research on prelexical processing, behavioral learning studies have been especially valuable as a window into the nature of lexical representations. Neuroscientific studies on word learning have the potential to help in localizing lexical processing and representations.

FLOW OF INFORMATION: HOW DO THE DIFFERENT STAGES TALK TO EACH OTHER?

Thus far, we have reviewed evidence suggesting that the four stages of processing of speech perception are distinct from each other, functionally and in terms of neural implementation, and have discussed data that has constrained accounts of the operations within each of those stages. The next issue to address is how the stages talk to each other. There are actually several questions here. First, we can ask whether bottom-up information processing is serial or cascaded. That is, do the prelexical stages complete their work before passing information on to lexical processing in a serial manner, or is there continuous, cascaded information flow? Second, are segmental processing and

suprasegmental processing fully independent of each other, or is there cross-talk at the prelexical level? Third, is there feedback of information from lexical processing to prelexical processing?

Cascaded Processing

As already discussed, multiple lexical hypotheses are considered in parallel during the word-recognition process. We can therefore address whether segmental information is passed serially or in cascade to lexical processing by asking whether lexical processing changes as a function of subsegmental differences in the speech input. Such differences entail fine-grained acoustic-phonetic distinctions that are perceived as falling within segmental categories rather than those that signal differences between categories. If processing is serial, these subsegmental differences should be resolved prelexically (e.g., the perceptual decision should be taken that a /k/ has been heard irrespective of how prototypical a /k/ it is). In contrast, if processing is cascaded, subsegmental differences should be passed forward to lexical processing (e.g., differences in the goodness of a /k/ should influence the relative strength of different word hypotheses at the lexical stage).

In English, voice-onset time (VOT) is a major acoustic-phonetic cue to the distinction between voiceless stop consonants (e.g., /k/, with longer VOTs) and voiced stops (e.g., /g/, with shorter VOTs). In a priming task, Andruski, Blumstein, and Burton (1994) observed that responses to target words such as ***queen*** were faster after semantically related prime words (e.g., ***king***) than after unrelated words. Importantly, this priming effect became smaller as VOT was reduced (i.e., as the /k/ became more like a /g/, but was still identified as a /k/). This suggests that subsegmental detail influences lexical processing (the degree of support for ***king***, and hence its efficacy as a prime, was reduced as the /k/ was shortened). Converging evidence for cascaded processing is provided by eye-tracking data (McMurray, Tanenhaus, & Aslin, 2002), by other priming data (van Alphen & McQueen, 2006), and by EEG data (Toscano, McMurray, Dennhardt, & Luck, 2010). Toscano et al. showed, for example, that the amplitude of early EEG components was modulated by changes in VOT in word-initial stops (e.g., /b/ and /p/ in ***beach–peach***). Although the amplitude of a frontal negativity at around 100 ms after stimulus onset (N1) was modulated by VOT but was unaffected by the category distinction between /b/ and /p/, the amplitude of a parietal positivity at around 300 ms (P3) was modulated by both factors, suggesting that the fine-grained VOT information was being passed forward at least to the categorical level.

Effects of phonetic similarity on the strength of lexical hypotheses (Connine et al., 1993, 1997; Marslen-Wilson et al., 1996) are also consistent with the idea that prelexical processing is cascaded. Further evidence comes from research showing that fine-grained acoustic-phonetic detail can modulate word recognition in continuous speech, and hence help the listener deal with the consequences of continuous speech production processes. Across processes and languages, fine-grained detail about the duration or spectral structure of segments helps listeners cope with variable realizations of those segments in particular phonological contexts (e.g., place assimilation in English; Gow, 2002; /t/ reduction in Dutch; Mitterer & Ernestus, 2006; liaison in French; Spinelli et al., 2003).

Cascaded processing can in addition be tested by asking whether subsegmental information interacts with lexical competition. If fine-grained phonetic information

(subsegmental details and cues for the resolution of the effects of continuous speech processes) modulates the competition process, then that information must have been passed forward to lexical processing. There have been several demonstrations of such interactions. Marslen-Wilson et al. (1996), for example, found that the effects of a perceptual ambiguity at the segmental level could be detected at the lexical level. Specifically, word recognition was delayed when the ambiguity was potentially consistent with other lexical hypotheses. Van Alphen and McQueen (2006) showed, similarly, that the effect of VOT variability on word recognition depended on the lexical competitor environment (i.e., whether the voiced and voiceless interpretations of a stop consonant were both words, were both nonwords, or were one word and one nonword).

The interaction of subsegmental and lexical information has been studied most extensively in a series of experiments with stimuli in which subsegmental cues are mismatched by cross-splicing different parts of spoken words (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Streeter & Nigro, 1979; Whalen, 1984, 1991). Cross-splicing the initial consonant and vowel of ***jog*** with the final consonantal release of ***job***, for example, produces a stimulus that sounds like ***job***, but that contains a vowel with acoustic evidence for an upcoming /g/. The degree to which such cross-splicing disrupts word recognition (as measured across a range of tasks including lexical decision, phoneme decision, and eye tracking) depends not only on whether the resulting sequence is a word (e.g., ***job*** vs. ***shob***) but also on whether the parts used in the cross-splicing originate from words (e.g., ***jog***) or nonwords (e.g., ***jod***).

There is also cascade of suprasegmental information up to the lexical stage. Many of the studies on suprasegmental prelexical processing reviewed earlier provide evidence of this. The suprasegmental information that listeners use to distinguish between words and to segment the speech stream appears to modulate the lexical competition process (Cho et al., 2007; Davis et al., 2002; Reinisch et al., 2010; Salverda et al., 2003; Shatzman & McQueen, 2006a). As with the evidence on segmental cascade, these interactions with lexical competition suggest that the suprasegmental information is being passed continuously forward to lexical processing.

Segmental–Suprasegmental Cross-Talk

It is important to emphasize that, although this review has so far considered that segmental and suprasegmental prelexical processing are distinct, there must be substantial interaction between the two processes (as shown by the bidirectional arrow in Figure 1.2). As argued by Cho et al. (2007), for example, this is because the cues used by the two processors can be the same. For instance, the duration of a segment can simultaneously signal a segmental contrast, since some segments are longer than others, and a suprasegmental contrast (e.g., the location of a word or phrase boundary). It can also be the case that there are interdependencies between the two processors; for example, determination of a prosodic structure based on a durational cue could depend on knowledge about the intrinsic duration of the segments involved (again because some segments tend to be longer than others). Tagliapietra and McQueen (2010), in a cross-modal priming study on the recognition of geminate consonants in Italian, present evidence that the same information (the duration of the geminate consonant) is used for both segmental analysis ("what" decisions about whether the consonant is a singleton or a geminate) and suprasegmental analysis

("where" decisions about the location of the segment within the word). Recent research has examined how computation of prosodic structure may modulate perceptual decisions about segments (Mitterer, Cho, & Kim, 2016), indicating once again the need for segmental–suprasegmental cross-talk.

No Online Top-Down Informational Feedback

For spoken-word recognition to succeed, information in the incoming speech signal must be fed forward to lexical processing. As we have just seen, the bottom-up flow of information is cascaded, and it entails interactions between segmental and suprasegmental prelexical processing. But is there top-down feedback from lexical to prelexical processing?

Demonstrations of lexical involvement in phonemic decision making might appear to show that this is the case. There are, at least under certain experimental conditions, lexical effects in phoneme monitoring, including faster responses to target phonemes in words, such as /b/ in ***bat***, than in nonwords, such as /b/ in ***bal*** (Cutler & Carter, 1987; Rubin, Turvey, & Van Gelder, 1976) and faster responses to targets in high-frequency words than in low-frequency words (Segui & Frauenfelder, 1986). There are also lexical effects in phonetic categorization, again under at least some conditions (Burton, Baum, & Blumstein, 1989; Connine & Clifton, 1987; Fox, 1984; Ganong, 1980; McQueen, 1991; Miller & Dexter, 1988; Pitt & Samuel, 1993). Listeners who are asked to categorize ambiguous sounds (on an artificially constructed continuum between, e.g., /d/ and /t/) are more likely to label the sound in a lexically consistent way (e.g., more /d/ decisions to stimuli from a ***deep–teep*** continuum and more /t/ decisions to stimuli from a ***deach–teach*** continuum; Ganong, 1980). In addition to this Ganong effect, there are lexical effects in rhyme monitoring (McQueen, 1993) and there is lexical involvement in the phonemic restoration illusion (the tendency for listeners to hear an illusory phoneme in a sequence where that phoneme has been replaced by noise; Warren, 1970). The illusion, for example, is stronger in real-word sequences than in nonsense-word sequences (Samuel, 1981, 1987, 1996). Once again, lexical involvement in phonemic restoration does not appear under all conditions (Samuel, 1996).

All of these demonstrations are consistent with the claim that there is top-down feedback from lexical to prelexical processing, and they have indeed been used to support this claim (for phonetic categorization, Ganong, 1980; for phonemic restoration, Samuel, 1981; for phoneme monitoring, Stemberger, Elman, & Haden, 1985). Lexical feedback could modulate prelexical phonemic processing, leading to the lexical biases and reaction time advantages observed across tasks requiring phonemic decisions. But simple demonstrations of lexical involvement in such tasks are equally compatible with the view that there is no feedback (Cutler, Mehler, Norris, & Segui, 1987; McQueen, 1991; Norris, McQueen, & Cutler, 2000). If phonemic decisions are made postlexically (as, e.g., in the Merge model; Norris et al., 2000), then lexical involvement in these tasks would also be expected.

The question about whether there is feedback from lexical to prelexical processing therefore cannot be settled using metalinguistic tasks that test for lexical effects on phonemic decisions. Measures are required that test more specific predictions about feedback. Here we discuss three lines of research that have attempted to do this.

First, studies have asked whether there are not only facilitatory lexical effects in phoneme monitoring but also inhibitory

effects (i.e., slower responses when the lexicon supports a different phoneme from that in the input). If there is feedback, both kinds of effects should be found. Frauenfelder, Segui, and Dijkstra (1990) found facilitatory effects in an experiment in French (e.g., faster responses to /t/ in ***gladiateur*** than in the matched nonword ***bladiateur***) but no inhibitory effects (e.g., responses to /t/ in ***vocabutaire*** were no slower than in ***socabutaire***, in spite of the fact that top-down feedback from the word ***vocabulaire*** ought to have been supporting /l/). Mirman, McClelland, and Holt (2005) showed, however, that lexically induced inhibition can be found, but only when the two phonemes are more phonetically similar than /t/ and /l/. For example, in an experiment in English, responses to /t/ in ***arsenit*** were delayed, but those to /t/ in ***abolit*** were not, presumably because /t/ is more similar to the lexically consistent, word-final /k/ in ***arsenic*** than to the word-final /ʃ/ in ***abolish***. Accounts with feedback (McClelland & Elman, 1986) and without feedback (Norris et al., 2000) agree that phoneme monitoring latency should be modulated by phonetic similarity, and both can explain how facilitatory and inhibitory lexical effects depend on similarity. Once again, therefore, these data do not determine whether there is or is not feedback.

The second approach is based on the logic that, if there is feedback from lexical to prelexical processing, lexical factors should modulate the inner workings of the prelexical processor (Elman & McClelland, 1988). Perceptual compensation for fricative-stop coarticulation is the tendency for listeners to perceive ambiguous stops on a continuum between /t/ and /k/ as /k/ after the fricative /s/ but as /t/ after the fricative /ʃ/ (Mann & Repp, 1981). This prelexical process reflects compensation for the acoustic consequences of fricative-stop coarticulation. Elman and McClelland showed that compensation for coarticulation appeared to be lexically mediated. Listeners made more /k/ responses in a sequence such as to ***christma[s/ʃ] [t/k]apes*** (with an ambiguous fricative and ambiguous stops) than in a sequence such as ***fooli[s/ʃ] [t/k]apes***. Feedback from lexical to prelexical processing would appear to be filling in the lexically consistent fricative (as in the Ganong effect), but crucially this fricative then appears to have a similar effect on the prelexical compensation process as an unambiguous fricative. These findings appear to show that there is feedback from lexical to prelexical processing.

Given the theoretical importance of the seminal work of Elman and McClelland (1988), it should come as no surprise that there have been a substantial number of follow-up studies. Some of these studies call into question the conclusion that there is feedback. Transitional probabilities between word-final fricatives and their preceding segments may provide an alternative explanation for apparent lexical effects (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Pitt & McQueen, 1998). If these probabilities are coded at the prelexical level, no feedback is required to explain the mediated compensation effect. Experiment-induced biases may also provide an alternative reason for the effects that again does not require feedback (McQueen, 2003; McQueen, Jesse, & Norris, 2009). Effects of word length and of perceptual grouping (Samuel & Pitt, 2003) make it more difficult to interpret results from this paradigm, and there are problems with the replicability of the original effect (McQueen et al., 2009; Samuel & Pitt, 2003).

McQueen et al. (2009) reviewed this literature and argued that there was no convincing data for lexical–prelexical feedback from the compensation for coarticulation paradigm. In fact, there is data from the paradigm that suggest that there is no such feedback. Lexical

effects in decisions about the fricatives (e.g., more /s/ responses to ***christma[s/ʃ]*** than to ***fooli[s/ʃ]***) can be found without lexical effects on the stops (i.e., no lexically mediated compensatory shift in /t/–/k/ decisions) or even in the presence of effects on the stops opposite to those predicted by the lexical bias (McQueen et al., 2009; Pitt & McQueen, 1998). These dissociations between fricative and stop decisions are inconsistent with feedback (if feedback is operating, it should produce consistent lexical biases on both the fricatives and the stops). The dissociations support feedforward accounts in which the prelexical compensation process is immune to lexical effects, but in which lexical processing can still influence postlexical fricative decisions, as in Merge (Norris et al., 2000).

The third line of research on feedback has combined the behavioral Ganong effect with neuroimaging techniques. The logic here is that if lexical variables can be shown to modulate activity in prelexical processing regions, then that modulation must be the result of top-down feedback. Participants in an fMRI study (Myers & Blumstein, 2008) demonstrated a lexical bias in phonetic categorization (e.g., more /k/ responses in a ***kiss–giss*** context than in a ***kift–gift*** context) and a parallel effect in brain activity (the blood-oxygenation-level dependent [BOLD] signal in the bilateral superior temporal gyri [STGs] varied as a function not only of the acoustic-phonetic ambiguity of the stop consonant but also of the lexical context). In a similar study, Gow, Segawa, Ahlfors, and Lin (2008) also found a behavioral Ganong effect, and related that to the results of a Granger causality analysis using a combination of MEG, EEG, and structural MRI data. Time-varying activity in the supramarginal gyrus (SMG) Granger-caused time-varying activity in the posterior STG, 280–480 ms after stimulus onset.

Although these findings can be taken as evidence for feedback, this conclusion rests on a number of assumptions. First, it rests on the claim that the STG supports prelexical processing but not lexical processing. As discussed earlier, the STG certainly appears to be involved in prelexical processing. But it is not yet known whether this is all that the STG does (DeWitt & Rauschecker, 2012; Price, 2012; Ueno et al., 2011). Second, a problem with the Myers and Blumstein (2008) findings (but not those of Gow et al. 2008) is that they are based on the BOLD signal, which reflects processes occurring over time (until 1,200 ms after stimulus offset in this case). The effect may therefore not reflect online perceptual processing. Third, and relatedly, it is not clear whether the effects reported in both studies, even if they do show evidence of higher-level influence on prelexical processing, reflect online top-down transmission of information. That is, they may not reflect the type of feedback that we have been discussing thus far. It is possible that modulation of activity in the STG could reflect other kinds of computation than online information transmission, including feedback for learning, feedback for attentional control, or feedback for perceptual binding.

It would thus be premature to conclude, on the basis of these neuroimaging studies, in favor of feedback. The available studies on compensation for coarticulation offer no unambiguous support for feedback either, and indeed provide evidence against it. Evidence for lexical involvement in selective adaptation effects (Samuel, 1997, 2001) may, like that from the neuroimaging studies, reflect perceptual learning processes rather than online information feedback (see McQueen et al., 2009, for further discussion). Although older findings on the absence of inhibitory effects in phoneme monitoring once challenged the feedback view, more recent studies show that such effects can be

found. On balance, then, there is no strong empirical support for online feedback of information from lexical to prelexical processing (see also Kingston, Levy, Rysling, & Staub, 2016).

There are also theoretical arguments against this kind of feedback. As argued in more detail by Norris et al. (2000) and Norris, McQueen, and Cutler (2015), informational feedback cannot benefit word recognition, and can harm phoneme recognition. The best a word-recognition system can do is recognize the words that are most probable given the input (Norris & McQueen, 2008). If processing is optimal in this way, feedback simply cannot improve on this.

In contrast, lexical retuning of speech perception, one of the types of perceptual learning discussed earlier, is beneficial for speech perception. The adjustments listeners make, using their lexical knowledge to retune perceptual categories, help them understand the speaker the next time that speaker is encountered. It is thus important to distinguish between feedback for learning, which can enhance speech perception over time and for which there is strong empirical support, and online informational feedback, which cannot enhance speech perception and which lacks empirical support. It remains possible that more convincing evidence of online feedback will be found in the future, but it appears more likely that evidence will be found for other ways in which higher-level processing influences prelexical processing. These are beneficial (and indeed necessary) for speech perception, and include feedback for perceptual learning, processes of perceptual binding (lining up words to their constituent sounds), and feedback for attentional control.

Summary

The available evidence suggests that there are constraints on flow of information in the speech-recognition system. Bottom-up flow of information is cascaded with respect to both segmental and suprasegmental properties of the speech signal, and there appears to be cross-talk between segmental and suprasegmental processing. But although there is evidence for top-down feedback for perceptual learning and there is a need for top-down feedback for binding and attentional control, there appears not to be online top-down feedback of information. That is, the lexicon appears not to influence the prelexical evaluation of the evidence in the speech signal as that input is being heard.

CONCLUSION

The aim of this chapter has been to give an account of how listeners extract words from the speech signal, up to the point where the system has recognized a particular word form. We have argued that, for this to occur, listeners need to solve three major computational problems: the variability problem (spoken sounds and words are not acoustically invariant), the segmentation problem (discrete words need to be extracted from a quasicontinuous speech stream) and the lexical-embedding problem (words sound like other words). We have presented evidence on how multiple mechanisms, at different stages in the speech-processing hierarchy, process segmental and suprasegmental information in parallel in order to solve these three problems. We have made the case that abstraction and adaptation are particularly important mechanisms. Listeners build phonologically abstract representations of the incoming speech signal, and speech perception is adaptive (i.e., it is flexible in response to different listening situations).

Spoken-language comprehension certainly does not end with the recognition of word forms—the box in Figure 1.2

labeled "interpretative processing" is a placeholder for a range of processes that we have not discussed, such as syntactic processing and retrieval of concepts. Although these processes are of course central to speech comprehension, they are beyond the scope of this review. It is important to note, however, that syntactic and semantic processing can influence word-form processing. For example, contextual information plays a key role in helping listeners recognize reduced words in continuous speech (Ernestus, 2014). Information in the sentence context is used rapidly in the competition process to select among possible candidate word forms, but the bottom-up signal has priority in determining which candidates are considered (see, e.g., Dahan & Tanenhaus, 2004; Marslen-Wilson, 1987; Nygaard & Queen, 2008; van den Brink, Brown, & Hagoort, 2001; Zwitserlood, 1989). An important objective for future research is to establish (and computationally implement) how contextual constraints are combined with signal-driven constraints on spoken-word recognition, and in particular to specify how form-based processes ("lexical form processing" in Figure 1.2) interface with syntactic, semantic, and pragmatic processes ("interpretative processing" in Figure 1.2). How, for example, are the representations of the phonological forms of words bound to representations of their meanings?

This review is limited in scope in a number of other ways. We have not considered in detail the time course of speech processing (e.g., the speed with which prelexical and lexical processing must operate such that the listener can keep up with the speaker's average of four syllables per second). In keeping with the evidence on cascaded processing reviewed earlier, speech perception appears to be fully incremental, with information passed rapidly and continuously forward to interpretative processing. We have also not considered the recognition of morphologically complex words (see, e.g., Balling & Baayen, 2012, for a discussion) and many of the ways in which speech perception is tuned to the phonological properties of the native language (see, e.g., Cutler, 2012, for an overview of the differences between native and nonnative listening). Clearly, a full account of speech perception would include both of these dimensions.

Studying the mental processes that transform acoustic information in the speech signal into linguistic meaning has for a long time been in the domain of psycholinguistics. As in other areas of psychology, cognitive neuroscience has had an increasing impact on the field over the past two decades. The hope is of course that combining insights from the two fields will ultimately result in an account of how language comprehension is instantiated in the brain. The impact of cognitive neuroscience has varied across the different stages of processing in speech perception. Although there is important evidence coming from neuroimaging in the domains of auditory and prelexical processing, there are relatively few neuroscientific studies on suprasegmental processing.

Psychophysics and animal models have provided a good account of how general auditory processing extracts features from the signal, which then form the starting point for speech-specific computations. From psycholinguistics we have a fair amount of knowledge about the computations, information flow, processing stages, and representations that are involved in speech perception. PET and fMRI studies have, on a macroanatomical scale, described how the key stages of processing map onto brain structures, and to some extent the functional connections between these distributed structures. But these techniques have not yet made a large contribution to our understanding of the nature of prelexical and lexical

representations. These techniques are also quite limited in their ability to investigate processes, partly because of low temporal resolution, but also because of conceptual problems in distinguishing processes from representations in experimental designs that rely on cognitive subtraction. When comparing two conditions in a subtraction-based design, it is often unclear whether the activations reflect a difference in processing, or in the outcome of that processing, or both (Obleser & Eisner, 2009).

Electrophysiological methods (EEG, MEG) are beginning to the uncover neural mechanisms that are necessary for decoding the temporal dynamics of speech. We have mentioned a few examples of how neural oscillations have been linked to computational processes. This line of research promises to be able to track the processing of linguistic structures on different timescales (Ding, Melloni, Zhang, Tian, & Poeppel, 2016), and to track the flow of information between key cortical regions with high temporal resolution (Park, Ince, Schyns, Thut, & Gross, 2015). Although it seems clear that speech-processing networks entrain to the rhythm of continuous speech (Ding & Simon, 2014), it remains to be established to what extent this reflects a causal role in computational processes such as segmentation. Through its very high spatial and temporal resolution, electrocorticography (ECoG) offers the ability to study representations as well as information flow in the speech-perception network, and although it can only be used in specific patient populations, there are interesting new studies coming out that confirm the complex and distributed nature of the speech-processing architecture in the brain (Mesgarani et al., 2014).

Joining the concepts and models from psycholinguistics with those from neuroscience is a big challenge, not only on a technical level, but also because they are often concerned with different levels of explanation. For example, in psycholinguistics there are several models of spoken word recognition that can account for a wealth of behavioral data and have a computational implementation. A current challenge is to design a next-generation model that combines the best features of the existing models. There are two broad classes of implemented models, abstractionist and episodic. Abstractionist models such as TRACE (McClelland & Elman, 1986), the Distributed Cohort Model (Gaskell & Marslen-Wilson, 1997), or Shortlist (Norris, 1994), work with abstract units of representation (e.g., phonemes or phonological word forms), which do not contain acoustic-phonetic detail. In contrast, episodic models such as MINERVA (Goldinger, 1998) encode detailed memory traces about every spoken word they encounter, but do not include phonologically abstract prelexical units. Although episodic models can account, for example, for evidence that differences in the way talkers pronounce words can influence word recognition (Goldinger, 1998; McLennan & Luce, 2005; Mullennix, Pisoni, & Martin, 1989; Nygaard, Sommers, & Pisoni, 1994) and talker-specific learning effects in speech perception, they cannot explain how that learning then generalizes across the mental lexicon (McQueen et al., 2006), even to words of other languages (Reinisch et al., 2013). This generalization is difficult to explain without abstract prelexical representations that connect to all entries in the lexicon (Cutler, Eisner, McQueen, & Norris, 2010).

Abstractionist models have the reverse limitation: They cannot account for the talker-specific effects but are able to explain generalization. The next generation of models will likely be kinds of hybrid models that can account for abstraction (McClelland & Elman, 1986; Norris & McQueen, 2008),

adaptability (Kleinschmidt & Jaeger, 2015; Yildiz, Kriegstein, & Kiebel, 2013), and talker-specific representations (Goldinger, 1998). They would also need to incorporate an account of how expectations about the incoming signal are updated continuously (Astheimer & Sanders, 2011; Gagnepain, Henson, & Davis, 2012; Sohoglu, Peelle, Carlyon, & Davis, 2012). These last papers exemplify another recent trend in cognitive neuroscience: the question about whether perception is predictive. There is considerable behavioral evidence that listeners use multiple sources of information to make predictions about upcoming speech; as Norris et al. (2015) argue, a goal for future research will be to specify the mechanisms that underlie this predictive behavior.

This chapter illustrates the current mismatch in level of analysis between the experimental psychology of speech perception and the neuroscience of speech perception. Neurobiological models are still largely concerned with mapping key cortical areas and connections on a macroanatomical scale (e.g., Hickok & Poeppel, 2007), and it is not easy at present to study the neural implementation of psycholinguistic concepts such as a phoneme or a lexical representation. If current trends continue, however, there is reason to be optimistic that the interaction between experimental psychology and neuroscience will increase. For this interaction to be a true two-way street, the computational and implementational levels of speech perception will need to be linked through functional models of speech perception.

LIST OF ABBREVIATIONS

BOLD	blood-oxygenation-level dependent
ECoG	electrocorticography
EEG	electroencephalography
fMRI	functional magnetic resonance imaging
MEG	magnetoencephalography
MTG	middle temporal gyrus
PET	positron emission tomography
PWC	possible-word constraint
SMG	supramarginal gyrus
STG	superior temporal gyrus
VOT	voice-onset time

REFERENCES

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage, 49*(1), 1124–1132. doi:10.1016/j.neuroimage.2009.07.032

Adank, P., Smits, R., & Hout, R. V. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America, 116*(5), 3099–3107.

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 115*(6), 3171–3183.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. doi:10.1006/jmla.1997.2558

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage, 52*(4), 1528–1540. doi:10.1016/j.neuroimage.2010.05.048

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163–187.

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *Journal of Neuroscience*, *35*(2), 634–642. doi:10.1523/JNEUROSCI.2454-14.2015

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, *49*(12), 3512–3516. doi:10.1016/j.neuropsychologia.2011.08.014

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. doi:10.1016/j.jml.2014.03.002

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience*, *27*(7), 1286–1297. doi:10.1162/jocn_a_00801

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, *125*(1), 80–106. doi:10.1016/j.cognition.2012.06.003

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., … Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, *49*(10–11), 763–786. doi:10.1016/j.specom.2007.02.006

Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience*, *8*, 386. doi:10.3389/fnins.2014.00386

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*, 592–597.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, *17*, 353–362.

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 50–68. doi:10.1037/a0038509

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience*, *34*(13), 4548–4557. doi:10.1523/JNEUROSCI.4339-13.2014

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. J. (2013). The role of linguistic and indexical information in improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America*, *133*(1), 474–482. doi:10.1121/1.4770239

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*(3), 958–968. doi:10.1016/j.neuroimage.2004.12.019

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 567–575.

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., … Cash, S. S. (2014). Speech-specific tuning of neurons in human superior temporal gyrus. *Cerebral Cortex*, *24*(10), 2679–2693. doi:10.1093/cercor/bht127

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*(2), 210–243. doi:10.1016/j.wocn.2006.03.003

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I Adult data. *Journal of Memory and Language*, *51*(4), 523–547. doi:10.1016/j.jml.2004.07.001

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*(1–2), 53–69.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 551–563.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes*, *16*(5–6), 535–564. doi:10.1080/01690960143000155

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, *11*(6), 1084–1089. doi:10.3758/BF03196741

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*(2), 193–210. doi:10.1006/jmla.1993.1011

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 291–299.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, *70*(3), 403–411.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*(4), 463–480. doi:10.1006/jmla.1997.2535

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes*, *16*(5–6), 609–636. doi:10.1080/01690960143000083

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*(3), 207–228. doi:10.1177/00238309020450030101

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, *92*, 81–104.

Cutler, A. (2012). *Native listening*. Cambridge, MA: MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. doi:10.1016/0749-596x(92)90012-m

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3–4), 133–142. doi:10.1016/0885-2308(87)90004-0

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165–179.

Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science*, *13*(3), 258–262.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology* (Vol. 10, pp. 91–111). Berlin, Germany: Mouton de Gruyter.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH 2008)* (p. 2056). Red Hook, NY: Interspeech.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400. doi:10.1016/0749-596x(86)90033-1

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*(2), 141–177.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech

segmentation by bilinguals. *Cognitive Psychology*, *24*(3), 381–410.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121. doi:10.1037//0096-1523.14.1.113

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*(6), 824–844. doi:10.1006/jmla.1994.1039

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*, *105*(3), 1877–1888. doi:10.1121/1.426724

Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, *44*(2), 171–195. doi:10.1177/00238309010440020301

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5–6), 507–534. doi:10.1080/01690960143000074

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513. doi:10.1037/0278-7393.30.2.498

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, *48*(7), 882–887. doi:10.1016/j.cortex.2011.05.017

Davis, M. H. (2016). The neurobiology of lexical access. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 541–555). London, United Kingdom: Academic Press/Elsevier. doi:10.1016/B978-0-12-407794-2.00044-4

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3773–3800. doi:10.1098/rstb.2009.0111

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423–3431.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. doi:10.1037/0096-1523.28.1.218

DeWitt, I., & Rauschecker, J. P. (2012). PNAS Plus: Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, USA*, *109*(8), E505–E514. doi:10.1073/pnas.1113427109

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. doi:10.1177/0956797610384743

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. doi:10.1038/nn.4186

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. doi:10.3389/fnhum.2014.00311

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85* Pt. 2, 761–768. doi:10.1016/j.neuroimage.2013.06.035

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Dumay, N., & Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119–132. doi:10.1016/j.cognition.2011.12.009

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 914–927.

Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, *157*, 1–42.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238.

El Aissati, A. McQueen, J. M., & Cutler, A. (2012). Finding words in a language that allows words without vowels. *Cognition*, *124*(1), 79–84. doi:10.1016/j.cognition.2012.03.006

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*(2), 143–165. doi:10.1016/0749-596X(88)90071-X

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, *33*(26), 10688–10697. doi:10.1523/JNEUROSCI.4596-12.2013

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. doi:10.1016/j.lingua.2012.12.006

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1–3), 162–173. doi:10.1006/brln.2001.2514

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–458. doi:10.1111/j.1749-6632.2009.04582.x

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893. doi:10.1121/1.412063

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. doi:10.1126/science.1164318

Formisano, E., Kim, D.-S., Di Salle, F., Moortele, P.-F. V. de, Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*, 859–869.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526–540.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 77–91.

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*(7), 615–621. doi:10.1016/j.cub.2012.02.015

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*(3), 361–377.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J., . . . Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia*, *47*, 123–131.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144–158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613–656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental*

Psychology: Human Perception and Performance, *24*(2), 380–396.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266.

Ghitza, O. (2014). Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, *5*, 652. doi:10.3389/fpsyg.2014.00652

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmitt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*, 1588–1598.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. doi:10.1038/nn.3063

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 163–179. doi:10.1037/0096-1523.28.1.163

Gow, D. W., Jr. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language*, *121*(3), 273–288. doi:10.1016/j.bandl.2012.03.005

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359.

Gow, D. W., Segawa, J. A., Ahlfors, S. P., & Lin, F.-H. (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, *43*(3), 614–623. doi:10.1016/j.neuroimage.2008.07.027

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. doi:10.1016/j.bandl.2005.06.001

Hanulíková, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology*, *63*(3), 555–579. doi:10.1080/17470210903038958

Hanulíková, A., Mitterer, H., & McQueen, J. M. (2011). Effects of first and second language on segmentation of non-native speech. *Bilingualism: Language and Cognition*, *14*(04), 506–521. doi:10.1017/S1366728910000428

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology*, *88*, 1433–1450.

Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7–8), 1145–1166. doi:10.1080/01690965.2012.662280

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. doi:10.1038/nrn3158

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422. doi:10.1016/j.neuron.2011.01.019

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. doi:10.1038/nrn2113

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics*, *72*(5), 1218–1227. doi:10.3758/APP.72.5.1218

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. doi:10.1016/j.jml.2007.02.001

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical

learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939–1956. doi:10.1037/a0025641

Isel, F., & Bacri, N. (1999). Spoken-word recognition: The access to embedded words. *Brain and Language*, *68*(1–2), 61–67. doi:10.1006/brln.1999.2087

Jesse, A., & Laakso, S. (2015). *Sentence context can guide the retuning of phonetic categories to speakers* (pp. 274–275). Presented at the Psychonomic Society Annual Meeting.

Jesse, A., & McQueen, J. M. (2014). Suprasegmental lexical stress cues in visual speech can guide spoken-word recognition. *Quarterly Journal of Experimental Psychology*, *67*(4), 793–808. doi:10.1080/17470218.2013.834371

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences, USA*, *97*(22), 11793–11799.

Kapnoula, E. C., & McMurray, B. (2015). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic Bulletin and Review*, *23*(2), 491–499. doi:10.3758/s13423-015-0897-1

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology* (Vol. *6*, pp. 143–161). Cambridge, United Kingdom: Cambridge University Press.

Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, *31*(5), 1715–1720. doi:10.1523/JNEUROSCI.4572-10.2011

Kilian-Hütten, N., Vroomen, J., & Formisano, E. (2011). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *NeuroImage*, *57*(4), 1601–1607. doi:10.1016/j.neuroimage.2011.05.043

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, *51*(4), 343–359. doi:10.1177/0023830908099069

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception & Performance*, *42*(12), 1969–1988. http://psycnet.apa.org/doi/10.1037/xhp0000269

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi:10.1037/a0038695

Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language*, *34*(1), 19–40. doi:10.1006/jmla.1995.1002

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*(2), 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1–15.

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, *25*(12), 2179–2188. doi:10.1162/jocn_a_00463

Kriegstein, K. von, Smith, D.R.R., Patterson, R. D., Ives, D. T., & Griffiths, T. D. (2007). Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Current Biology*, *17*(13), 1123–1128. doi:10.1016/j.cub.2007.05.061

Kriegstein, K. von, Smith, D.R.R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *Journal of Neuroscience*, *30*(2), 629–638. doi:10.1523/JNEUROSCI.2742-09.2010

Ladefoged, P. (1989). A note on "Information conveyed by vowels." *The Journal of the Acoustical Society of America*, *85*, 2223–2224.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*, 98–104.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, United Kingdom: Blackwell.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. *7*, pp. 637–676). Berlin, Germany: Mouton de Gruyter.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. doi:10.1016/j.cogpsych.2007.01.001

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*(1), 101–123. doi:10.1177/00238309070500010501

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America*, *51*(6B), 2018–2024. doi:10.1121/1.1913062

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*, 1621–1631.

Liebenthal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., & Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *Journal of Neuroscience*, *33*(39), 15414–15424. doi:10.1523/JNEUROSCI.1511-13.2013

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608–622. doi:10.1037/a0029243

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114. doi:10.1016/j.tics.2008.11.008

Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, *60*(3), 484–490.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5–6), 565–581. doi:10.1080/01690960143000137

Luce, P. A., & Lyons, E. A. (1999). Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 174–183.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science*, *27*, 285–298.

Mann, V. A., & Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, *69*(2), 548–558.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. doi:10.1016/0010-0277(87)90005-9

Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1376–1392.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675.

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin and Review*, *15*(2), 453–457.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of*

Experimental Psychology: General, *134*(4), 477–500. doi:10.1037/0096-3445.134.4.477

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*, 543–562.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

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

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321. doi:10.1037/0278-7393.31.2.306

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 433–443.

McQueen, J. M. (1993). Rhyme decisions to spoken words and nonwords. *Memory & Cognition*, *21*(2), 210–222.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46.

McQueen, J. M. (2003). The ghost of Christmas future: Didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science*, *27*(5), 795–799. doi:10.1016/S0364-0213(03)00069-7

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, *61*(1), 1–18.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 621–638. doi:10.1037/0278-7393.20.3.621

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1363–1389. doi:10.1037/0096-1523.25.5.1363

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, *45*(1), 103–132.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692–1696. doi:10.1016/j.cub.2007.08.064

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010. doi:10.1126/science.1245994

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369–378.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, *52*, 416–435.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, *65*(6), 956–969.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone

variation: Evidence from a learning paradigm. *Cognitive Science*, *35*(1), 184–197. doi:10.1111/j.1551-6709.2010.01140.x

Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics*, *54*, 68–79. doi:10.1016/j.wocn.2015.09.002

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, *30*(3), 451–479. doi:10.1207/s15516709cog0000_57

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*(1), 73–103. doi:10.1016/j.wocn.2005.03.003

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. doi:10.1016/j.cognition.2008.08.002

Mitterer, H., & McQueen, J. M. (2009a). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLOS ONE*, *4*(11), e7785. doi:10.1371/journal.pone.0007785

Mitterer, H., & McQueen, J. M. (2009b). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 244–263. doi:10.1037/a0012730

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361. doi:10.1016/j.cognition.2013.07.011

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C.-G., & Giraud, A.-L. (2012). Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, *3*, 248. doi:10.3389/fpsyg.2012.00248

Möttönen, R., Dutton, R., & Watkins, K. E. (2013). Auditory-motor processing of speech sounds. *Cerebral Cortex*, *23*(5), 1190–1197. doi:10.1093/cercor/bhs110

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379–390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Murty, L., Otake, T., & Cutler, A. (2007). Perceptual tests of rhythmic similarity: I. *Mora rhythm. Language and Speech*, *50*(Pt. 1), 77–99.

Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fMRI investigation. *Neuropsychologia*, *45*(7), 1463–1473. doi:10.1016/j.neuropsychologia.2006.11.005

Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: an fMRI investigation. *Cerebral Cortex*, *18*(2), 278–288. doi:10.1093/cercor/bhm053

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895–903. doi:10.1111/j.1467-9280.2009.02380.x

Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language*, *76*, 80–93. doi:10.1016/j.jml.2014.06.007

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, *85*(5), 2088–2113.

Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language*, *64*(4), 460–476.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. doi:10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209–1228.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

Norris, D., McQueen, J. M., & Cutler, A. (2015). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. doi:10.1080/23273798.2015.1081703

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*(3), 191–243. doi:10.1006/cogp.1997.0671

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes*, *16*(5–6), 637–660. doi:10.1080/01690960143000119

Nusbaum, H., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. A. Johnson & J. W. Mullennix (Eds.), *Talker variability and speech processing* (pp. 109–132). San Diego, CA: Collection.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030. doi:10.1037/0096-1523.34.4.1017

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, *13*(1), 14–19. doi:10.1016/j.tics.2008.09.005

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*(2), 258–278. doi:10.1006/jmla.1993.1014

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*. doi:10.1038/nn.4021

Park, H., Ince, R.A.A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649–1653. doi:10.1016/j.cub.2015.04.049

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. doi:10.3389/fpsyg.2012.00320

Perrodin, C., Kayser, C., Logothetis, N. K., & Petkov, C. I. (2011). Voice cells in the primate temporal lobe. *Current Biology*, 1–8. doi:10.1016/j.cub.2011.07.028

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184. doi:10.1121/1.1906875

Petkov, C. I., Kayser, C., Augath, M., & Logothetis, N. K. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLOS Biology*, *4*, 1–14.

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, *39*(3), 304–311. doi:10.1016/j.wocn.2010.07.004

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*(3), 347–370.

Pitt, M. A., & 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.

Poellmann, K., Bosker, H. R., McQueen, J. M., & Mitterer, H. (2014). Perceptual adaptation to segmental and syllabic reductions in

continuous spoken Dutch. *Journal of Phonetics*, *46*, 101–127. doi:10.1016/j.wocn.2014.06.004

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. doi:10.1016/j.neuroimage.2012.04.062

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. doi:10.1016/j.jml.2007.04.001

Rauschecker, J., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. doi:10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences, USA*, *97*(22), 11800–11806. doi:10.1073/pnas.97.22.11800

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 539–555. doi:10.1037/a0034409

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, *63*(4), 772–783. doi:10.1080/17470210903104412

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75–86. doi:10.1037/a0027979

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105. doi:10.1016/j.wocn.2014.04.002

Repp, B., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). New York, NY: Collection.

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The roles of left versus right anterior temporal lobes in conceptual knowledge: An ALE meta-analysis of 97 functional neuroimaging studies. *Cerebral Cortex*, *25*(11), 4374–4391. doi:10.1093/cercor/bhv024

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, *19*(5), 394–398. doi:10.3758/BF03199398

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89.

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*(23), 3079–3085. doi:10.1016/j.cub.2015.10.009

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–494.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, *26*(1), 36–56.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*(1), 28–51. doi:10.1037/0096-3445.125.1.28

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*(2), 97–127. doi:10.1006/cogp.1997.0646

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*(4), 348–351.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, *48*, 416–434.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related

potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703. doi:10.1038/nn873

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, *52*(Pt. 2–3), 135–175.

Schall, S., Kiebel, S. J., Maess, B., & Kriegstein, K. von. (2015). Voice identity recognition: Functional division of the right STS and its behavioral relevance. *Journal of Cognitive Neuroscience*, *27*(2), 280–291. doi:10.1093/cercor/11.10.946

Schreuder, R., & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, *33*(3), 357–375. doi:10.1006/jmla.1994.1017

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(Pt. 12), 2400–2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100–107. doi:10.1016/S0166-2236(02)00037-1

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302. doi:10.1038/nrn2603

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*(1–2), 13–45.

Sebastián-Gallés, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, *31*(1), 18–32. doi:10.1016/0749-596x(92)90003-g

Segui, J., & Frauenfelder, U. H. (1986). The effect of lexical constraints upon speech perception. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 795–808). Amsterdam, Netherlands: North Holland.

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, *28*(4), 439–454. doi:10.1023/A:1023245216726

Shatzman, K. B., & McQueen, J. M. (2006a). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, *17*(5), 372–377. doi:10.1111/j.1467-9280.2006.01714.x

Shatzman, K. B., & McQueen, J. M. (2006b). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, *68*(1), 1–16.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211. doi:10.1037/a0016803

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011a). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception & Psychophysics*, *73*(4), 1195–1215. doi:10.3758/s13414-011-0096-8

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011b). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia*, *49*(14), 3831–3846. doi:10.1016/j.neuropsychologia.2011.09.044

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, *33*(Pt. 1), 47–68.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. doi:10.1523/JNEUROSCI.5069-11.2012

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and*

Language, *45*(3), 412–432. doi:10.1006/jmla.2000.2783

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, *48*(2), 233–254. doi:10.1016/S0749-596X(02)00513-2

Stemberger, J. P., Elman, J. L., & Haden, P. (1985). Interference between phonemes during phoneme monitoring: Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4), 475–489.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *The Journal of the Acoustical Society of America*, *65*(6), 1533–1541.

Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*(1), 177–193. doi:10.1016/j.jml.2011.08.001

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, *36*(3), 422–444. doi:10.1006/jmla.1996.2495

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342–354. doi:10.1016/j.jml.2012.07.001

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, *34*(4), 440–467. doi:10.1006/jmla.1995.1020

Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 758–775.

Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminates and singletons in spoken Italian. *Journal of Memory and Language*, *63*(3), 306–323.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. doi:10.1016/j.neuroimage.2013.08.023

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, *15*(6), 355–366. doi:10.1038/nrn3731

Toni, I., de Lange, F. P., Noordzij, M. L., & Hagoort, P. (2008). Language beyond action. *Journal of Physiology—Paris*, *102*(1–3), 71–79. doi:10.1016/j.jphysparis.2008.03.005

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*(10), 1532–1540. doi:10.1177/0956797610384142

Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*(1), 1–15. doi:10.1016/j.bandl.2010.03.008

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2), 385–396. doi:10.1016/j.neuron.2011.09.013

van Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 178–196.

van Alphen, P. M., & Van Berkum, J.J.A. (2010). Is there pain in champagne? Semantic involvement of words within words during sense-making. *Journal of Cognitive Neuroscience*, *22*(11), 2618–2626. doi:10.1162/jocn.2009.21336

van Alphen, P. M., & Van Berkum, J.J.A. (2012). Semantic involvement of initial and final lexical embeddings during sense-making: The advantage of starting late. *Frontiers in Psychology*, *3*, 190. doi:10.3389/fpsyg.2012.00190

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, *63*(5), 811–823.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, *58*(2), 251–273. doi:10.1080/02724980343000927

Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 270–278.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329. doi:10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vroomen, J., & Baart, M. (2009). Phonetic recalibration only occurs in speech mode. *Cognition*, *110*(2), 254–259. doi:10.1016/j.cognition.2008.10.015

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 98–108. doi:10.1037/0096-1523.21.1.98

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 710–720. doi:10.1037/0096-1523.23.3.710

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, *24*(6), 744–755.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, *44*(Pt. 1), 95–118.

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34–51. doi:10.1016/j.wocn.2014.05.002

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, *35*(1), 49–64.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, *50*(4), 351–360.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*(1), 1–15. doi:10.1037/h0026823

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5–6), 609–630. doi:10.1080/016909699386202

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14. doi:10.1037/0278-7393.32.1.1

Yildiz, I. B., Kriegstein, K. von, & Kiebel, S. J. (2013). From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLOS Computational Biology*, *9*(9), e1003219. doi:10.1371/journal.pcbi.1003219

Yip, M.C.W. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia*, *44*(3), 223–229. doi:10.2117/psysoc.2001.223

Yip, M.C.W. (2004). Possible-word constraints in Cantonese speech segmentation. *Journal of Psycholinguistic Research*, *33*(2), 165–173.

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences, USA*, *107*(2), 592–597. doi:10.1073/pnas.0904774107

Zhang, X., & Samuel, A. G. (2015). The activation of embedded words in spoken word recognition. *Journal of Memory and Language*, *79–80*, 53–75. doi:10.1016/j.jml.2014.12.001

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*(1), 25–64.

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, *10*(2), 121–136. doi:10.1080/01690969508407090