Native listening

The flexibility dimension

Anne Cutler
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; MARCS Institute, University of Western Sydney, Australia; Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

The way we listen to spoken language is tailored to the specific benefit of native-language speech input. Listening to speech in non-native languages can be significantly hindered by this native bias. Is it possible to determine the degree to which a listener is listening in a native-like manner? Promising indications of how this question may be tackled are provided by new research findings concerning the great flexibility that characterises listening to the L1, in online adjustment of phonetic category boundaries for adaptation across talkers, and in modulation of lexical dynamics for adjustment across listening conditions. This flexibility pays off in many dimensions, including listening in noise, adaptation across dialects, and identification of voices. These findings further illuminate the robustness and flexibility of native listening, and potentially point to ways in which we might begin to assess degrees of 'native-likeness' in this skill.

Keywords: speech, native language, non-native language, phonetic processing, lexical processing, perceptual learning, talker identification, dialects, listening in noise

1. Introduction

Just as we are all native speakers of a given language, we are also 'native listeners'; experience with listening to a particular language encourages us to listen in the way that is most efficient for that language (Cutler, 2012).

The native influence in listening takes many forms. Phoneme repertoires vary in size and makeup: in Spanish, listeners have to distinguish 20 consonants but only five vowels, while listeners to British English need to distinguish 24 consonants and no fewer than 26 vowels (Maddieson, 1984; Wells, 1982). More

Dutch Journal of Applied Linguistics 1:2 (2012), **169–187**. DOI 10.1075/dujal.1.2.02cut
ISSN 2211–7245 / E-ISSN 2211–7253

phonemes means more contrasts to distinguish, and also more potential phonetic environments in which each sound can occur. Listeners are sensitive to this type of variance; thus Spanish listeners, with their highly skewed consonant-vowel ratio, expect vowels to be more variably affected by consonant context than consonants are affected by vowel context, while Dutch listeners, whose language has a much more balanced ratio, expect equivalent phonetic context effects in either direction (Costa, Cutler, & Sebastián-Gallés, 1998).

Not only the number of phonetic contrasts varies across languages; their nature notoriously varies too, with the well-known results of perceptual difficulty for mismatching contrasts, especially when two separate categories of a non-native category map to a single category of the native tongue (Best & Tyler, 2007). Still at the word level, languages vary also in the dimensions that they use to encode inter-word contrasts (e.g., tone languages have pitch distinctions that are irrelevant in other languages), as well as in word class and morphology (e.g., use of articles, or of affixes), in systems for classification (e.g., noun gender), and in many more ways.

Further, words arrive at the listener's ear not as clearly separated units, but embedded in a continuous stream of speech without robust or reliable boundary signals; to understand messages, listeners must parse the stream into individual words. Here too the native tongue helps: listeners develop segmentation procedures based on phonological likelihood. In English, for instance, stressed syllables are assumed to be word-initial (Cutler & Butterfield, 1992; Cutler & Norris, 1988), a highly efficient strategy given the lexical statistics of the language (less than 10% of lexical words in speech begin with unstressed syllables; Cutler & Carter, 1987). But as most languages of the world do not have English-like stress (Van der Hulst, 1999), this procedure must be language-specific, and it is: while the stress-based segmentation procedure of English also works for the similarly biased stress language Dutch (Vroomen, Van Zon, & De Gelder, 1996), syllabic segmentation proves more useful for French, Spanish and Korean (Cutler, Mehler, Norris, & Segui, 1986; Kim, Davis, & Cutler, 2008; Kolinsky, Morais, & Cluytens, 1995; Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993), and a mora-based procedure is used in Japanese and in Telugu (Cutler & Otake, 1994; Murty, Otake, & Cutler, 2007; Otake, Hatano, Cutler, & Mehler, 1993). These different procedures develop from language experience; whatever our language, we listen to it in the way that best suits its structure.

What is efficient for one language is not always efficient for another, however. If listening is language-specific, then it follows that someone who has grown up with one language will listen with less than maximal efficiency to a language with different structure. Listening to non-native languages will in effect be hampered by an accent for the native tongue, in much the same way that non-native languages

are spoken with the accent of the native tongue. The crucial role of language experience means that second languages learned late will be harder to process than a language experienced since birth, as indeed research on listening to speech has abundantly shown (Cutler, 2012).

Is it feasible to investigate degrees of nativeness in the way we listen to speech? Of course, listening proficiency in a particular language is quantifiable on many dimensions: phonetic identification, word recognition, prosodic sensitivity, knowledge of syntactic structures, of idioms, of semantic implicatures. But consider that a highly proficient L2 user may score at ceiling on such tests but still find listening skills breaking down whenever listening becomes in some way difficult. There is, it is argued here, a set of dimensions that particularly characterise native listening, and that can be grouped together under the heading of listening flexibility. They include the ability to understand, in most cases immediately, newly encountered talkers whom we have never heard before, to adjust rapidly to speech that is atypical (e.g., accented), and to succeed in listening despite background noise or reverberation, or interruptions of the signal in a radio broadcast or phone conversation. In principle all these skills are also available to L2 listeners, but in practice they are often less in evidence in the L2 than in the L1 case. As this makes clear, these are gradable dimensions, on which L1 and L2 listening can be compared. Sections 2 and 3 will summarise relevant new findings concerning flexibility of listening, respectively at the phonetic and the lexical level.

2. Flexibility at the phonetic level

Listeners adapt rapidly to new talkers, certainly if they are speaking the native variety of the native language. It is simply no problem to talk to a shopkeeper or new neighbour with whom we have never spoken before, even though every talker's vocal tract produces speech with talker-specific characteristics. This adaptation is accomplished by adjusting the placement of phoneme category boundaries (Cutler, Eisner, McQueen, & Norris, 2010; Norris, McQueen, & Cutler, 2003). A process of perceptual learning draws on existing knowledge to resolve ambiguity, then applies the learning to adjust future decisions. The experiments that show this have two parts: an exposure phase in which listeners can use existing knowledge to interpret ambiguous sounds, followed by a test phase to assess the effect of the exposure. In the exposure phase a phoneme, ambiguous between two interpretations, appears in a context in which one interpretation is favoured. In the initial demonstration of this perceptual learning (Norris et al., 2003), the exposure involved a lexical decision task, in which a phoneme halfway between /f/ and /s/ is likely to be interpreted as /f/ in the context *gira-* (because *giraffe* is a word but

girasse is not), but as /s/ in *hor-* (because *horse* is a word but *horf* is not; Dutch example contexts are *olij-* to induce /f/ and *radij-* to induce /s/).

In the test phase listeners categorise ambiguous phonemes (in this case as /f/ or /s/). The tokens for categorisation vary along a continuum from more /f/-like to more /s/-like, with the exposure sound as the continuum centre point. The results reveal that not only the exposure sound shows effects of the earlier exposure; rather, the whole category boundary shifts. The /f/ category expands for those listeners who were trained to interpret the ambiguous sound as /f/, while the /s/ category expands for those who were led to interpret it as /s/. The listeners thus learn to retune their phoneme categories for this speaker, allowing the category to include a wider variety of realisations of the phoneme. Importantly, the retuning is speaker-specific, in that training with one person's /s/ or /f/ does not alter decisions about another person's utterances of the same sounds (Eisner & McQueen, 2005). Thus it works in just the right way to enable adjustment to a new talker.

Furthermore, the learning is long-lasting. It is as strong after a short interval, after a day of normal life, or after a night's sleep, as it was right after the initial exposure (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). This likewise is consistent with a function serving adaptation to newly encountered talkers; next time we meet the same talker we want to reap the benefit of the learning built up in previous interactions. Learning is inhibited, however, by evidence that the unusual pronunciation has some extraneous cause such as a dialectal feature (Kraljic, Brennan, & Samuel, 2008) or a temporary interruption of the speech signal (e.g., the speaker has a pencil in her mouth; Kraljic, Samuel, & Brennan, 2008).

Lexical access and lexical decision (involving awareness of the source providing interpretation of the ambiguity) are not necessary conditions for this type of perceptual learning; all sorts of exposure prove effective — tallying a list of words (McQueen, Norris, & Cutler, 2006), listening to a story (Eisner & McQueen, 2006), picture-name matching (McQueen, Tyler, & Cutler, 2012), or even hearing phonotactically constrained nonsense words such as *smuter* or *frulic* (Cutler, McQueen, Butterfield, & Norris, 2008). Similarly, the effects of exposure appear not only in phonetic tasks but also in name-picture matching (McQueen et al., 2012) and in deciding between minimal word pairs such as *knife*/*nice* or *doof*/*doos* (McQueen, Cutler, & Norris, 2006; Sjerps & McQueen, 2010).

Adjustment of the interpretation of ambiguous percepts by reference to existing knowledge is not speech-specific; such learning is a powerful mechanism for perceptual tasks. In our world, where complex signals arrive at speed, vary in form, and overlap or occur simultaneously, ambiguity is all around us; perceptual learning uses existing knowledge to retune decision-making for future encounters with such ambiguity. This type of learning also underlies visual interpretation of colour in context, for example, and learning about letters in print and in handwriting.

Within speech it appears with lexical tones (Mitterer, Chen, & Zhou, 2011) as well as with phonemes, and among phonemes with fricatives, stops (Kraljic & Samuel, 2006) and liquids (Scharenborg, Mitterer, & McQueen, 2011), and in a similar task with vowels (Maye, Aslin, & Tanenhaus, 2008). The source of knowledge in the speech case does not have to be lexical but can also be phonotactic (Cutler et al., 2008), with analogous learning about phoneme mapping from text (Escudero, Hayes-Harb, & Mitterer, 2008; Mitterer & McQueen, 2009) and from visual cues to articulation (Van Linden & Vroomen, 2007). It does not depend on a large vocabulary, or on the ability to read and make metalinguistic decisions about words, since as Figure 1 shows it is displayed in essentially identical form by adults, by 12-year-olds and by 6-year-olds (McQueen et al., 2012; indeed, similar accent adaptation is shown even by 2-year-olds: White & Aslin, 2011). It is also not dependent on highly accurate acoustic processing since it is still in place and apparently unaltered in older listeners despite hearing decline (Adank & Janse, 2010; Scharenborg, Janse, & Weber, 2012).

Most importantly, the learning improves subsequent perception, in that it generalises from the exposure instances to other possible contexts (McQueen et al., 2006; Sjerps & McQueen, 2010). Hearing someone saying *giraffe* with a weird /f/ helps us understand the same talker conversing about *fat*, *traffic* and *life* in

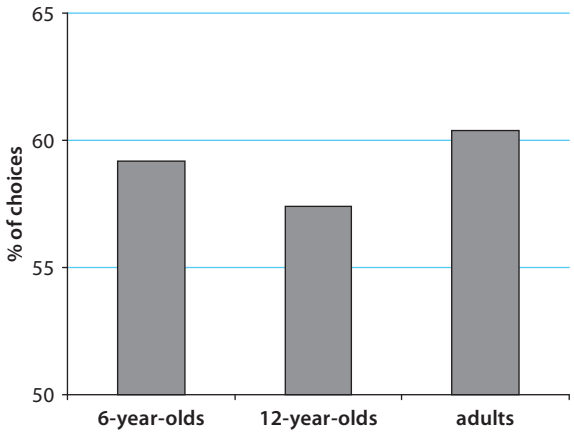

Figure 1. Perceptual learning consistency across the lifespan (McQueen et al., 2012): six-year-old participants (before reading acquisition), 12-year-olds, and adults tend to choose a character's name (*Fimpy or Simpy*?) in accord with the exposure they received in a picture-verification task; if an ambiguous fricative was heard in words like *giraffe*, they are more likely to identify the name as *Fimpy*, if it was heard in words like *horse*, they are more likely to decide for *Simpy*. Chance performance would be 50%, and these values are averaged over a seven-step continuum, thus including relatively clear /f/ and /s/ cases. The shift towards the trained category is equivalently significant at all ages.

general. Minimal-pair words such as *knife* and *nice* show this most easily; *ni-* followed by an ambiguous (f/s) is interpreted as *knife* by listeners who heard that sound in *giraffe* contexts, but as *nice* by those who heard the sound in /s/-words like *horse*. Note that the learning must therefore involve abstract phoneme representations. In the minimal pairs, the ambiguous phoneme occurred in phonetic environments unlike any presented in the exposure phase, and the learning also transfers across position in the word (Jesse & McQueen, 2011), thus across different positional allophones.

Finally, generalisation shows that the information supporting learning did not have its effect by top-down feedback from the vocabulary controlling phonetic processing during recognition. Such feedback would remove perceptual ambiguity on a case-by-case basis, but this would have no implications for later processing and would not affect the category boundary as a whole, nor would it resolve ambiguity in minimal pairs where both phoneme alternatives yield a viable word. The resolution of such ambiguity is exactly the reason why perceptual learning is so important a contribution to listening flexibility.

3. Flexibility at the lexical level

In all languages, the vocabulary contains many tens or even hundreds of thousands of words, all of them constructed from a mere handful of phonemes (around 31 on average, with the most common total being even less, at 25 (Maddieson, 1984). An inevitable consequence of this is that words closely resemble one another and longer words contain shorter words embedded within them. Moreover, as speech is continuous and word boundaries are often unclear, juxtaposition of words can create sequences corresponding to yet more words. Thus *worst* contains *were* and *worse*, and *acre* has *ache*; putting them together as *worst acre* then adds *steak*, *take* and *taker* to the mix. As a result, speech may contain many spurious words that are just as well supported by the acoustic signal as the words that speakers actually choose and utter.

Listeners deal with this by entertaining multiple possible hypotheses about what the speech input is. All words contained in such a signal can be, albeit temporarily, activated in the listener's mind. This even includes accidentally present words such as *steak* in *worst acre*. Recognition is not a matter of dealing sequentially with the input starting with the first available word, and, at its end, assuming a new word's beginning, because of the embedding problem (so *were* and *ache* are the first but not the right candidates in *worst* and *acre*, respectively). Instead, candidate words supported by the input compete with one another until winners emerge. The winners are words that are not mismatched and that form a sequence

accounting for the input with no leftover residue. Mismatching input is used immediately it arrives (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001) to remove word candidates that are no longer viable; *worst* might initially activate *word* and *worth*, but these would disappear from the competition as soon as the [s] arrived. In this example, neither *worst acre* nor *worst ache* is mismatched, but the former triumphs by virtue of accounting for the whole sequence, whereas the latter would leave the final syllable as a leftover residue.

Evidence of multiple concurrent activation and competition is plentiful. Words that happen to begin in the same way are momentarily activated together (Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Zwitserlood, 1989). Embedded words are harder to spot in contexts supporting another word; e.g., *mess* is harder to find in *domess* than in *vomess* because of competition from *domestic* in the former case (McQueen, Norris, & Cutler, 1994). The more competing words are activated, the harder recognition is (Norris, McQueen, & Cutler, 1995).

Recent research (Brouwer, Mitterer, & Huettig, 2012; McQueen & Huettig, 2012) has revealed this level of listening to be surprisingly flexible. The parameters of the activation and competition process may be adjusted to make processing more efficient, especially in difficult listening conditions. Mismatch is normally drastically effective, but this is not so if other evidence suggests that acoustic-phonetic identification errors are likely. In eye-tracking studies (that measure what

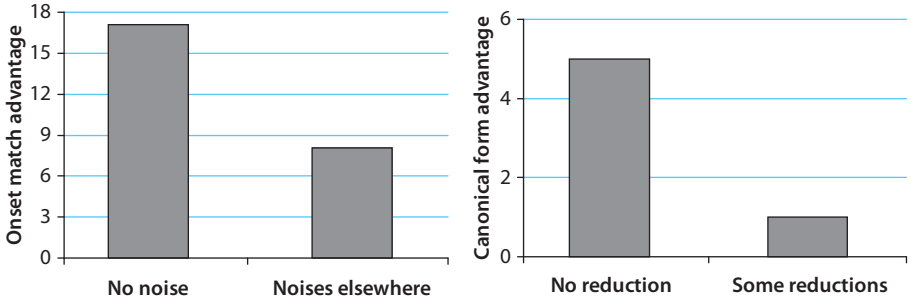

Figure 2. Modulation of competition dynamics in eye-tracking experiments by McQueen and Huettig (2012; left) and Brouwer et al. (2012; right). Without added noise, listeners tend to look more often at competitor words that begin in the same way as a target word they are hearing than at competitors that end in the same way, but this asymmetry is reduced when there is some noise (elsewhere) in the experiment. Also competitor words that sound like a canonical pronunciation of the target receive more looks than words that sound like a reduced pronunciation; this tendency is likewise reduced when there is reduced speech (elsewhere) in the experiment. Uncertainty about signal reliability leads listeners to modulate the degree to which they will consider alternative interpretations of incoming speech.

people look at as they hear speech), competitors for word onsets typically attract more looks than competitors for offsets — e.g., hearing *candle* will at first induce looks to a picture of a candy, and can also induce looks to a handle, but the former tendency is stronger (Allopenna et al., 1998; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; McQueen & Viebahn, 2007). This asymmetry becomes, however, less pronounced if there is some noise around, as shown in the left panel of Figure 2. Even words that are themselves clearly pronounced and unaffected by noise are more likely to induce looks to offset competitors and less likely to induce looks to onset competitors if occasional background crackle (as with an imperfectly tuned AM radio) interrupts other utterances (McQueen & Huettig, 2012).

The same result is observed when some other utterances are spoken casually, with consequent speech reductions such as *ornry* instead of *ordinary*; even for target words that are themselves realised clearly, acoustic mismatch with competing words is less penalised and there are more looks to non-target words (Brouwer et al., 2012). This is depicted in the right panel of Figure 2. So exactly the same speech input is processed more leniently if the listening conditions suggest that the acoustic signal may not be reliable, versus more strictly if there is no reason to doubt such reliability.

4. The fruits of flexibility

The adjustment of listening tolerances at the lexical level contributes to the adaptability of listening in adverse conditions, and the adjustment of phonetic boundaries, where necessary at the speaker-specific level, makes us able to adapt to new talkers. The pay-off of native listening flexibility can be seen in aspects of listening where we well know that L2 users tend to be at a disadvantage. Three situations will be highlighted in this section: Listening in noise, adjustment to dialects, and identification of individual talkers.

4.1 Listening in noise

There is no need to tell anyone who has ever functioned well in a second language that noisy conditions can pull the rug out from under an apparently secure mastery of listening. The phenomenon is not disputed (see the review of 50 years of research on this topic by Garcia Lecumberri, Cooke, & Cutler, 2010); its explanation is another matter. There are basically two classes of explanation. On the one hand, speech processing in L2 can be assumed to be fallible at the level of phoneme identification and discrimination, such that L2 listeners perform best with particularly clear acoustic information concerning phonetic segments. If the available

information is masked, and the phonemic level of processing thereby fails to deliver input on which higher levels of processing can reach decisions about words and sentence structure, then essentially the system grinds to a halt. On this phonetic hypothesis, the difficulty of L2 listening in noise is basically a problem with the identification of phonetic segments.

An alternative type of explanation locates the difficulty more at higher levels of processing. This type of explanation is preferred by researchers who work with spontaneous speech and view acoustic-phonetic information, in L1 or in L2, as always likely to be impoverished with respect to an ideal citation pronunciation. On this account, L1 and L2 listening are equally fallible at the level of phoneme processing, but this fallibility is compensated for by a wide range of resources on which the listener can call to recover from the inadequacy of the input and to reconstruct what the input must have been. These resources are much more extensive, and are exploited with much greater efficiency, in the native language. This type of account thus locates the difficulty of L2 listening in noise at the higher levels of processing on which listeners rely for recovery from the problems that speakers and listening conditions always present them with.

A way to test these classes of explanation against one another is to restrict the effect of noise in listening to the phoneme level alone, for instance by requiring listeners to identify the phonetic structure of meaningless input. Under these conditions, no recourse to knowledge of words, of higher-level structure or contextual plausibility can help the listeners out of the uncertainty that the noise-masking has brought about. Attempts at this kind of restriction were undertaken independently by Cutler, Weber, Smits and Cooper (2004) and by Garcia Lecumberri and Cooke (2006).

In the first of these studies, American English CV or VC syllables were embedded centrally in babble noise (several talkers speaking at once, typical of the background that makes listening in a crowded bar hard). Although structure was predictable (separate sets of CVs or VCs), the central embedding made the exact moment of onset unpredictable (because the syllables varied in length), and the phonetic context was also unpredictable (because the materials used all vowels and consonants of the language bar the reduced vowel schwa). Listeners had either American English or Dutch as native language.

In the second study, American English consonants were presented in several kinds of noise, including babble noise, embedded in a constant [a_a] context and always preceded by the same amount of noise (so both the phonetic context and the exact moment of onset were predictable). Listeners had either British English or Spanish as native language. The prediction from the phonetic hypothesis is that exactly the result that is so often observed (greater effect of noise on L2 than on L1) will be observed in these studies too, because this pure phonetic processing

situation represents the type of processing that is at the root of the whole phenomenon. The prediction from the higher-level account, however, is that the phenomenon is rooted in the levels of processing that have been removed from these experimental situations, so that a quite different result may be observed.

Although the two studies were in many respects very similar, the results in fact differed. Cutler et al. (2004) found that both L1 and L2 listeners were affected by noise, but not asymmetrically; the Dutch listeners' identification scores were about 80% of the Americans' scores, both in quiet and in noise. Moreover, the average scores per phoneme correlated very highly indeed (r = .91) across the two listener groups — the sounds that were hard for the native group were also hard for the L2 group. Cutler et al. concluded against the phonetic hypothesis. Garcia Lecumberri and Cooke (2006), in contrast, found that their Spanish listeners were not much worse than the L1 listeners in quiet, but were significantly worse in noise, and concluded in favour of the phonetic hypothesis.

However, consider the small differences in predictability allowed by the experimental design of the Spanish study. If even a constant phonetic context and a predictable moment of onset count as resources for L1 recovery, then that design might have offered the L1 listeners an advantage. The obvious way to test this is to present the materials from the Spanish study to the listener population from the Dutch study. If the results of the Spanish study were due to the experimental design offering native listeners an advantage, and Dutch listeners are as unable

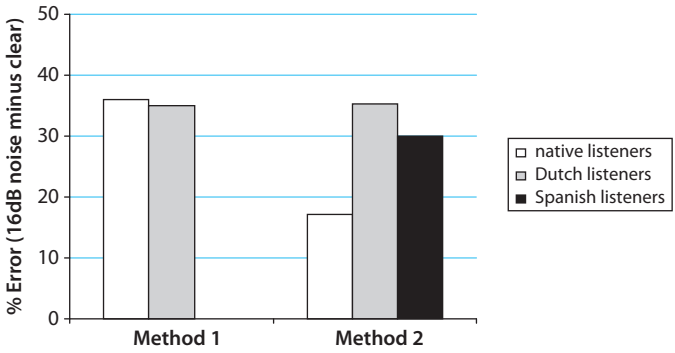

Figure 3. Impact of 16dB signal-to-noise-ratio on English-speaking and non-native listeners' identification of 15 American English consonants in the absence of lexical or sentential information, tested with two methods. In Method 1, consonant-vowel or vowel-consonant syllables were centrally embedded in noise. Compared with a clear presentation, the effect of this noise-masking was equivalent for native and non-native listeners. In Method 2, consonants were identified in tokens such as *aba*, *ana*, with noise onset a constant interval before token onset. Native listeners profited from even the minimal predictability offered by this method, such that the impact of noise was significantly less for native than for two groups of non-native listeners.

as any other L2 listener to use such a native advantage, then the Dutch listeners should perform like the Spanish listeners in this case. As Figure 3 shows for the phonemes contained in both materials sets, this is indeed what happened (Cutler, Garcia Lecumberri, & Cooke, 2008). Overall, therefore, the higher-level account is supported; native listeners recover better from the impact of noise on speech signals, by using whatever predictability, at whatever level, the input affords.

4.2 Adjustment to dialectal varieties

Something else that native listeners do well is identify the variety being spoken by another user of their language (Clopper & Pisoni, 2004), and indeed recognise when a different variety is being presented; non-native listeners are very bad at this (Clopper & Bradlow, 2009). The perceptual learning described in Section 2 has as its goal adjustment across speakers, and native listeners accomplish such adjustment rapidly, both at the individual talker level and at the level of dialectal variety (Dahan, Drucker, & Scarborough, 2008; Trude & Brown-Schmidt, 2012); indeed, it is likely that these learning processes underlie phonetic change and thus the differentiation that gives dialectal varieties their uniqueness (Cutler et al., 2010). Our ability to recognise varieties, to discriminate between them, and to adjust to their characteristics, is another terrain on which native listening displays its flexibility.

In a series of investigations of listener adjustment to casual speech processes across languages and dialects, Tuinman (2011) compared how the British English process of /r/-insertion was dealt with by native listeners versus listeners with another dialect (American English) or another language (Dutch). This process causes the appearance of /r/ between a word-final vowel and a following word-initial vowel (e.g., after the first word in *saw a film*, or *Canada aided Lesotho*). The process can lead to speech signals that are in principle compatible with unintended words (in the second example, *raided*). British English listeners proved to be highly sensitive to the relevant acoustic cue (inserted /r/ is significantly shorter than intended /r/), not to pay attention to any other factors in deciding whether they are hearing, for example, *saw ice* or *saw rice*, and not to show significant activation of the meaning of the unintended words that an /r/ would support (Tuinman, Mitterer, & Cutler, 2011a; 2012; see Figure 4).

Dutch listeners in the same studies were much less sensitive to the acoustic evidence, drew on irrelevant orthographic and semantic features to make their decisions, and showed significant activation of, for instance, *raided* when they heard *Canada aided*. American listeners, whose varieties are in general /r/-pronouncing postvocalically, like Dutch, and do not show the insertion process, are also not particularly good at making decisions based on the acoustics alone and tend to allow orthography to play a role, but crucially, they do not show significant

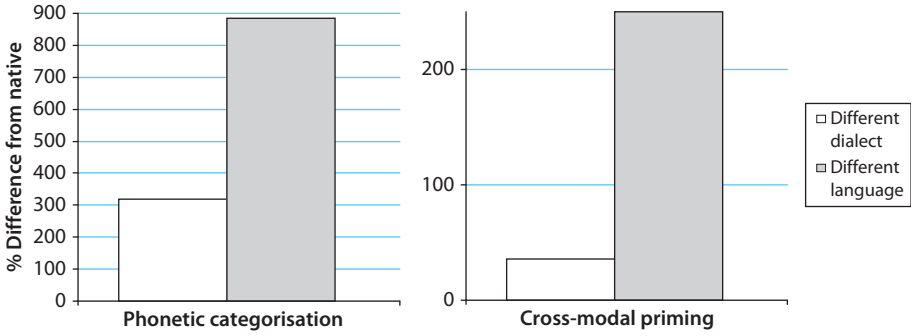

Figure 4. Difference from native performance on a phonetic judgement versus a word recognition task involving an unfamiliar casual speech process (British English /r/-insertion). In the phonetic task, the figure shows the degree to which irrelevant (non-durational) factors affected listeners' judgements, as a percentage of the effects of such factors for the native British listeners. Listeners with a different dialect (U.S. English) used such information more than three times as much as native listeners, and listeners with a different language (Dutch) nearly nine times as much. In word recognition (with cross-modal priming), however, the difference from the native results in amount of irrelevant priming is greatly attenuated (and insignificant) for the cross-dialect listeners, while remaining high (and significant) for the cross-language listeners.

activation of unintended words (Tuinman et al., 2011a; 2011b; again see Figure 4). This indicates that they are capable of recovering rapidly from the unusual pronunciation, as expected of native listening flexibility. Indeed, eye-tracking studies show that even the native British listeners show momentary consideration of /r/-initial words such as *raided* when they hear an insertion process as in *Canada aided*, though they reject such interpretations very rapidly (Tuinman et al., 2012). In other words, the native listeners' accurate identification performance rests on a recovery process, just as is the case with native comprehension in noisy conditions.

Casual speech processes of American English can also create difficulties for British English listeners. In words like *writer*, the medial /t/ can become a flap in many American varieties, and this can sound like a voiced stop, as in *rider*. The flapping process is sensitive to syntactic structure, in that it tends not to occur across a phrase boundary. Thus in the utterance *If you want to eat early, lunch will be served*, the final sound of *eat* is likely to be a flap, but this is much less likely in *If you want to eat, early lunch will be served*. British listeners are poor at using the presence of a flap to extract cues to phrase structure (Scott & Cutler, 1984). But interestingly, the same study showed that British speakers resident in the US had learned to do this, and their performance was uncorrelated with their length of residence. The latter lack of relationship suggests that immersion may be a 'tipping-point' experience, different in kind from simple accrual of sufficient

exposure; if more or less every talker one hears is using particular sounds or processes, adaptation is highly encouraged (at least in a variety of the native tongue).

Perceptual learning for talker adaptation can indeed occur in L2 listening (Reinisch, Weber & Mitterer, 2012), at least with highly proficient L2 listeners in an immersion environment (German students at Radboud University Nijmegen). Perceptual learning for dialect adaptation has been tested in an experiment in which the exposure condition used movie subtitles as the source of disambiguating information. Dutch listeners watched extracts from a movie spoken in heavily accented English, either Scottish or Australian. In the test phase they were asked to repeat back new utterances from the same two dialects. Of course they were better in repeating utterances from the variety that they had just heard, even if they had been in the baseline condition with no subtitles. They were even better if they had watched the movie with English-language subtitles to tell them exactly what words were spoken. But they were actually worse than the baseline if they had seen the movie with Dutch subtitles, which had presumably been quite efficient at letting them follow the action, but had given them no clue as to the mapping of pronunciation to underlying word form (Mitterer & McQueen, 2009). This result offers significant hope of generalisation of perceptual learning to the L2 situation.

4.3 Talker identification

Because of the talker-specific speech settings that each vocal tract produces, we are able to learn the characteristics of voices and thus recognise talkers whom we have heard before. Recent studies of how voice recognition is achieved in the brain (Andics et al., 2010; Perrachione, Pierrehumbert, & Wong, 2009) show that identification of voices involves left hemisphere neural circuitry used for language processing, and thus is separate from simple acoustic processing. Consistent with this, voice identification appears to be easier in a known language. Thus Thompson (1987) found that American English listeners more accurately identified talkers producing English than talkers producing Spanish. This native language advantage was observed even when the talkers were the same ones, i.e., German-English bilinguals (Goggin, Thompson, Strube, & Simental, 1991); in this study listeners heard a talker reading a text, then six further talkers, and had to identify the first talker among the latter six. Native English speakers performed better with English texts, while native German speakers performed better with German texts. Since the set of talkers was exactly the same, accuracy in talker identification was clearly dependent on listeners' knowledge of the language. Training effects do not remove the known language advantage (Perrachione & Wong, 2007; Winters, Levi, & Pisoni, 2008).

The known language need not be necessarily the L1, as Schiller and Köster (1996; Köster & Schiller, 1997) showed by testing listeners with German as L1,

with German as L2, or with no German, on recognition accuracy for voices speaking German. Listeners who knew no German performed worse than all those who knew German, while in the latter group, Spanish and Chinese L2 listeners performed less well than the native Germans and the L2 English. English is phonologically closer to German than either Spanish or Chinese is, so the known language advantage may be phonologically based. Rhythm (stress-based in both English and German) is not enough, because the advantage disappeared in reiterant speech in which rhythm was preserved but segmental structure lost (Schiller, Köster, & Duckworth, 1997); note that individual segments differ in how well they support voice discrimination (Andics, McQueen, & Van Turennout, 2007).

The known language also need not be understood for the advantage to hold: even seven-month-olds can better distinguish speakers in the language they are acquiring than in an unfamiliar language (Johnson, Westrek, Nazzi, & Cutler, 2011). In this study, Dutch infants listened to several voices speaking either in Dutch or in a foreign language. Once they showed signs of being bored, the input changed to a different voice speaking either the same language as before, or a different language. The infants always noticed a change of language, but a change of talker was only noticed when the talker was speaking Dutch. Talker switches in Italian or in Japanese were overlooked. Though seven-month-olds do not understand sentences, they presumably are sufficiently familiar with the phonological structure of the input for the known language advantage to hold.

Adults in a control experiment in the same study provided further support for the phonologically based account. In order to rule out intrinsic discriminability of the Dutch voices as an explanation of the Dutch babies' better performance with Dutch talkers, the materials were also presented to adult listeners who knew no Dutch (or Italian or Japanese), using the procedure of Goggin et al. (1991) described above. These adult listeners were speakers of Canadian English, students at the University of Toronto, and they also received Canadian English input, which they of course discriminated most accurately, while they had great difficulty discriminating Japanese talkers, and greatest difficulty of all with the Dutch talkers. Their performance with the Italian talkers, though, was less poor than might be predicted from the fact that none of them could actually use Italian at all; as Figure 5 shows, they hardly differed on Italian from their performance on English. This is interesting because Italian is the second most widely-spoken mother tongue in the Greater Toronto area (after English, and somewhat counter-intuitively well before French). Thus all the Toronto students would have had experience of hearing Italian spoken, even if they understood none of it. Apparently even such regular exposure gives access to the phonological infrastructure that supports voice discrimination.

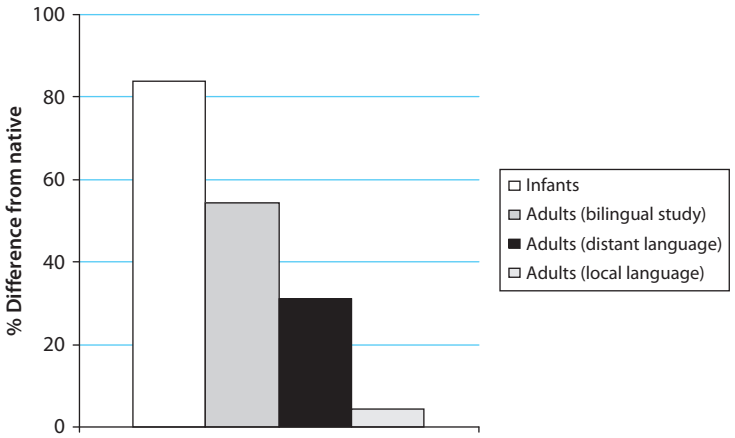

Figure 5. Native advantage in talker identification: percent performance increment for native language over (a) either Japanese or Italian, for Dutch infants; (b) the unknown language, for either English or German listeners hearing German-English bilinguals; (c) Japanese or Dutch, for Canadian English listeners; (d) Italian, an unknown but local language for Canadian English listeners (a, c, d: Johnson et al., 2011; b: Goggin et al., 1991).

5. Conclusion

The argument presented here may be viewed as a research program. The dimensions of flexibility that have been described in this paper are inherently gradable, and offer new ways of comparing native and non-native listening. Only a handful of studies directly addressing such comparisons of relative flexibility as yet exist. But the new techniques by which we have created windows on the adaptation of phoneme category boundaries for adjusting to talkers, and the modulation of lexical competition dynamics for adjusting to noisy listening conditions, present us with tools that could, eventually, enable direct assessment of an individual listener's degree of flexibility. If flexibility is an important dimension on which L1 and L2 listening differ, these methods may make it possible to quantify such flexibility and in so doing, establish when listening is indeed native.

Acknowledgments

A version of this paper was delivered as an invited address at the seventh Anéla Conference, Lunteren, May 2012, under the title "Native listening: can it be diagnosed?". Much of the reported research was supported by the NWO-SPINOZA project "Native and nonnative listening".

References

Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*, 736–740.

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

Andics, A., McQueen, J.M., Petersson, K.M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage*, *52*, 1528–1540.

Andics, A., McQueen, J.M., & Van Turennout, M. (2007). Phonetic context influences voice discriminability. In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences, saarbrücken, Germany* (pp. 1829–1832). Dudweiler: Pirrot.

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O.-S. Bohn (Eds.), *Second language speech learning* (pp. 13–34). Amsterdam: John Benjamins.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of spoken word recognition. *Language and Cognitive Processes*, *27*, 539–571.

Clopper, C.G., & Bradlow, A.R. (2009). Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics*, *37*, 436–451.

Clopper, C.G., & Pisoni, D.B. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, *32*, 111–140.

Costa, A., Cutler, A., & Sebastián-Gallés, N. (1998). Effects of phoneme repertoire on phoneme decision. *Perception & Psychophysics*, *60*, 1022–1031.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–142.

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 10* (pp. 91–111). Berlin: De Gruyter.

Cutler, A., Garcia Lecumberri, M.L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *Journal of the Acoustical Society of America*, *124*, 1264–1268.

Cutler, A., McQueen, J.M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In J. Fletcher, D. Loakes, R. Goecke, D. Burnham, & M. Wagner (Eds.), *Proceedings of Interspeech 2008*, *Brisbane* (p. 2056). Adelaide, Australia: Causal Productions.

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,* 385–400.

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

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

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, *116*, 3668–3678.

Dahan, D., Drucker, S.J., & Scarborough, R.A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*, 710–718.

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

Eisner, F., & McQueen, J.M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America*, *119*, 1950–1953.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*, 345–360.

Garcia Lecumberri, M.L., & Cooke, M. (2006). Effect of masker type on native and nonnative consonant perception in noise. *Journal of the Acoustical Society of America*, *119*, 2445–2454.

Garcia Lecumberri, M.L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*, 864–886.

Goggin, J.P., Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*, 448–458.

Jesse, A., & McQueen, J.M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*, 943–950.

Johnson, E.K., Westrek, W., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002–1011.

Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech*, *51*, 342–358.

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

Köster, O., & Schiller, N.O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*, 18–28.

Kraljic, T., Brennan, S.E., & Samuel, A.G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 51–81.

Kraljic, T., & Samuel, A.G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178.

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

Kraljic, T., Samuel, A.G., & Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*, 332–338.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N., & Dahan, D. (2003). The time course of spoken word recognition and learning: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202–227.

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.

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

McQueen, J.M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America*, *131*, 509–517.

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*, 621–638.

McQueen, J.M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech*, *49*, 101–112.

McQueen, J.M., Tyler, M., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*, 317–339.

McQueen, J.M. & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*, 661–671.

Mitterer, H., Chen, Y., & Zhou, X.L. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, *35*, 184–197.

Mitterer, H., & McQueen, J.M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE, 4*, e7785.

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

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

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

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 358–378.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, *32*, 373–389.

Perrachione, T.K., Pierrehumbert, J.B., & Wong, P.C.M (2009). Differential neural contributions to native-and foreign-language talker identification. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1950–1960.

Perrachione, T.K., & Wong, P.C.M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*, 1899–1910.

Reinisch, E., Weber, A., & Mitterer, H. (2012). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *38*. doi:10.1037/a0027979.

Scharenborg, O., Janse, E., Weber, A. (2012). Perceptual learning of /f/-/s/ by older listeners. *Proceedings of Interspeech 2012,* Portland, OR.

Scharenborg, O., Mitterer, H., & McQueen, J.M. (2011). Perceptual learning of liquids. *Proceedings of Interspeech 2011*, Florence, Italy (pp. 149–152).

Schiller, N.O., & Köster, O. (1996). Evaluation of a foreign speaker in forensic phonetics: A report. *Forensic Linguistics*, *3*, 176–185.

Schiller, N.O., Köster, O., & Duckworth, M. (1997). The effect of removing linguistic information upon identifying speakers of a foreign language. *Forensic Linguistics*, *4*, 1–17.

Scott, D.R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450–466.

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

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Tanenhaus, M.K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J.C. (1995). Integration of visual and linguistic information is spoken-language comprehension. *Science*, *268*, 1632–1634.

Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*, 121–131.

Trude, A.M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*, 979–1001.

Tuinman, A. (2011). *Processing casual speech in native and non-native language*. Doctoral dissertation, MPI Series in Psycholinguistics 60, Radboud University Nijmegen.

Tuinman, A., Mitterer, H., & Cutler, A. (2011a). Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *Journal of the Acoustical Society of America*, *130*, 1643–1652.

Tuinman, A., Mitterer, H., & Cutler, A. (2011b). The efficiency of cross-dialectal word recognition. In *Proceedings of the 12th annual conference of the international speech communication association (Interspeech 2011), Florence, Italy*. Adelaide, Australia: Causal Productions [CD-ROM].

Tuinman, A., Mitterer, H., & Cutler, A. (2012). Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language*, *66*, 530–544.

Van der Hulst, H.G. (1999). Word accent. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 3–116). Berlin: Mouton de Gruyter.

Van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1483–1494.

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

Wells, J.C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

White, K.S., & Aslin, R.N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*, 372–384.

Winters, S.J., Levi, S.V., & Pisoni, D.B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America*, *123*, 4524–4538.

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