Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 864–886
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Non-native speech perception in adverse conditions: A review

Maria Luisa Garcia Lecumberri [a,*], Martin Cooke [a,b], Anne Cutler [c,d,e]

[a] Language and Speech Laboratory, Facultad de Letras, Universidad del País Vasco, Paseo de la Universidad 5, 01006 Vitoria, Spain
[b] IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain
[c] Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
[d] MARCS Auditory Laboratories, University of Western Sydney, NSW 1797, Australia
[e] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

Abstract

If listening in adverse conditions is hard, then listening in a foreign language is doubly so: non-native listeners have to cope with both imperfect signals and imperfect knowledge. Comparison of native and non-native listener performance in speech-in-noise tasks helps to clarify the role of prior linguistic experience in speech perception, and, more directly, contributes to an understanding of the problems faced by language learners in everyday listening situations. This article reviews experimental studies on non-native listening in adverse conditions, organised around three principal contributory factors: the task facing listeners, the effect of adverse conditions on speech, and the differences among listener populations. Based on a comprehensive tabulation of key studies, we identify robust findings, research trends and gaps in current knowledge.

Keywords: Non-native; Speech perception; Noise; Review

Contents

* Corresponding author.
E-mail address: garcia.lecumberri@ehu.es (M.L. Garcia Lecumberri).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.08.014

1. Introduction

For many of us, the first non-native (NN) language experience outside the classroom is a shock. Not only are the answers to the carefully-practised stock phrases not those that appeared in the textbook, but the raw acoustic material reaching our ears lacks the clarity of the speakers in the quiet language laboratory. Thus unprepared, we enter the real world of the non-native listener, characterised by the dual challenges of *imperfect signal* and *imperfect knowledge*. And the problem persists even as we gain experience, exposure and confidence in the non-native language. Use of the telephone seems harder than it should be. Conversations in restaurants and bars are difficult to follow and join. The television never seems quite loud enough. We continue to prefer hearing non-native speakers of English at international conferences rather than highly-fluent natives. We finally take solace in the fact that even "true" bilingual listeners never quite reach the ability of monolinguals in the presence of noise (Mayo et al., 1997; Rogers et al., 2006).

Knowing about the extent of problems faced by non-native listeners in adverse conditions is important in developing theories of general speech perception. Comparing adult, normal-hearing populations who differ only in their native-language experience has the potential to provide insights into the role of linguistic factors in speech decoding. Since all listeners routinely handle acoustically-complex scenarios containing competing sound sources, reverberant energy and other forms of distortion, the use of native and non-native populations allows us to explore the extent to which linguistic knowledge is used in tasks such as sound source separation and identification. To give an example, consider the processes used by listeners to handle variability in formant frequencies due to factors such as differences in vocal tract sizes. Can vocal tract length normalisation be performed purely on the basis of the speech signal itself, or does it depend on the identification of units such as vowels in the speech stream, thereby engaging higher level representations which differentiate between native and non-native listeners? More generally, studies which compare native and non-native listeners help determine what processes in speech understanding are universal and which are language-specific.

The study of non-native speech perception in imperfect conditions has clear practical applications in a world where increased mobility means that large minorities work or study in a second or third language setting. Knowing the extent to which non-native listeners suffer in spaces with moderate-to-strong reverberation should be taken into account when designing classrooms and workspaces (Picard and Bradley, 2001; Nelson et al., 2005). Understanding non-native deficits in noisy settings such as factories ought to lead to revised standards for acceptable noise output levels. Differential effects of transmission channels on native and non-native listeners for speech broadcast over public address systems need to be understood to optimise information transfer to all listener groups, particularly in safety-critical situations involving emergency procedures. Similarly, studying the effect on non-native learners of differences in speaking style, from casual to careful speech, will inform the design of materials, training programmes and guidelines for educators.

This review focuses on studies involving non-native listener groups performing speech perception tasks in simulated adverse conditions. Work in this area is dominated by additive noise, with only a very limited number of reports employing reverberation or channel distortion. Added noise has been motivated both by a desire to reduce scores from ceiling level when comparing native and non-native listeners, and more directly to study the effect of simulating everyday conditions in which speech perception takes place. We consider both types of study here. The review is limited to acoustic stimuli, and excludes the extra dimension brought in by audio-visual studies (e.g., Hardison, 1996; Wang et al., 2008; Hazan et al., 2010). Further due to the paucity of non-native studies, we do not review the effects of noise on speech production. Another area outside the scope of the review is the perception of foreign accent in noise (e.g., Munro, 1998; Burki-Cohen et al., 2001; Rogers et al., 2004; Adank et al., 2009; Volin and Skarnitzl, 2010).

The review is organised as follows. Three background sections describe the parameters of the task faced by listeners in processing speech (Section 2), what constitutes non-native listening (Section 3), and how adverse listening conditions (noise, reverberation) affect the perception of speech (Section 4). In Section 5, we then provide a compre-

hensive overview of studies of non-native perception in adverse conditions, and in Section 6 we compare across the studies, in particular in the light of the features of non-native speech perception which emerged from the earlier sections. This section highlights limitations of work to date, and concludes with suggestions of some areas for further studies.

2. The listener's task in recognising speech

Whatever the language of input or the background of the listener, the goal of speech recognition is to extract meaning. Certain aspects of the speech recognition task effectively impose severe constraints on how listeners go about it. The first constraining factor is that all languages have a phonemic inventory that is trivially small in comparison with the size of the vocabulary it supports; the average number of phonemes across a representative sample of the world's languages is 31, with the most common inventory size being 25 (Maddieson, 1984). Vocabulary size runs into the tens if not hundreds of thousands, however, in all languages.

A simple calculation suffices to reveal the inevitable implication of this size relationship: the members of the vocabulary cannot exhibit a high degree of dissimilarity. Words resemble one another, and shorter words occur embedded fortuitously[1] in longer words. This means that whenever a listener hears spoken words, the input temporarily supports a range of alternative possibilities. *Star* could become *stark* or *starling* or *start* or *startle* or *starch*, or maybe *star* + *ch* was part of *star chart* after all. The task of spoken-word recognition is one of sorting out what is intended to be there from a large set of alternatives that are only partially or accidentally there.

The second constraining factor is the nature of speech production. The articulatory gestures by which sounds are produced flow smoothly in sequence just like any other human movements do; thus the movement producing a given sound in one phonetic context may differ from the movement producing the same sound in another context. Such coarticulation of sounds adds variability to speech, and the smooth continuity of articulation also makes the constituent units of speech – phonemes, but even words – difficult to discern separately. While *star* in *starch* and *star chart* may be easy to differentiate in isolated pronunciations, the disambiguating differences are not always guaranteed to be present in running speech.

[1] For example, consider that *fortuitous* which means "by chance" has nothing to do with *fortunate* meaning "lucky". People who mix these words up have fallen prey to the accidental similarities in the vocabulary described here. Latin, which of course also had a large vocabulary constructed from a small phoneme repertoire, passed the fortuitous similarity between its words *fortuitus* and *fortuna* straight on to English, where the present authors are fortunate enough to be able to make a point out of it.

The third constraining factor is that speech is spoken by people, and can happen more or less anywhere. People differ in the size and shape of their vocal tract, and factors such as the distance between speaker and listener, or where the conversation takes place, introduce further variability even without the listening conditions being in any way adverse in the sense of this review.

Thus the nature of the speech communication task presents the listener with signals that are variable, fast, continuous and non-unique. Yet listeners usually manage the task of speech recognition successfully, and apparently without great effort. Over the past few decades research has revealed how this happens; see the review articles by Frauenfelder and Floccia (1998) and McQueen (2007) for more details of the experimental evidence. The next two sections provide a brief summary.

2.1. The process of spoken-word recognition

The stored representations that listeners hold in memory, and against which incoming speech needs to be matched, are in most cases representations of words; some recurring phrases and useful productive affixes apart, words form the core of the recognition process. Across languages, what counts as a word can differ widely, and the number of processes affecting a word's actual form in speech can vary too; but in all languages the words in speech are variable in their form, are continuously appended to other words, and overlap in form with other possible words in the vocabulary.

Listeners do not wait for the resulting ambiguity to be resolved. Potential interpretations of the incoming input are considered simultaneously, and additional input immediately evaluated with respect to its import for the candidate set. This makes word recognition effectively a process in which rival lexical alternatives compete, in that evidence in favour of any one of them simultaneously counts against all rival candidates. If listeners hear a nonword which might have continued to become a word, they take longer to reject it than if it could not have become a word (e.g., *driz*, which might have become *drizzle*, is harder than *drim*; Taft, 1986). This is even true if the cue to a word is in the way sounds are articulated; so *troot* made up of *troo-* from *troop* plus /t/ is harder to reject than *troot* made up of *troo-* from *trook* plus /t/ (Marslen-Wilson and Warren, 1994; McQueen et al., 1999). The real words *drizzle* and *troop* were competing and so causing the longer response times; in the latter case, even tiny cues in the vowel of *troo-*, suggesting an upcoming bilabial sound, were apparently enough to favour an available real-word candidate.

When only a fragment has been heard, alternative continuations of it are equally available (Zwitserlood, 1989), but as soon as one of the alternatives is mismatched by the input, its probability of acceptance is reduced (Soto-Faraco et al., 2001). Many experiments examining this effect, including these two, have used a task in which listen-

ers hear speech but make a response about a visually presented word. This task ("cross-modal priming") exploits the fact that responses to a word are always faster on a second presentation. This includes responses to a visually presented word being faster after auditory presentation of the same word, or part of it, compared with after presentation of some unrelated word or fragment. Thus if visually presented CAPTAIN and CAPTIVE are both recognised faster after *capt-* has been heard, we assume that both were made available by the spoken fragment. And if the fragment *sardi-* makes responses to SARDINA 'sardine' similarly faster, while responses to SARDANA 'a dance' after *sardi-* are actually slower than after the unrelated control fragment, we assume that the second vowel in the fragment has added matching evidence to the probability of SARDINA, but has significantly reduced the probability of SARDANA by mismatching it. Soto-Faraco et al. found that any such mismatching evidence was immediately exploited by listeners, and it did not matter whether the mismatch came from a vowel or a consonant, or from stress pattern, or from a single phonetic feature (e.g., the place of articulation difference between /m/ and /n/) or many features (e.g., the difference between /p/ and /s/); anything which distinguished between rival word candidates was used right away to favour the winner and disfavour the alternatives.

This result makes very clear what the role of phonemes is in spoken-word recognition; although the distinction between any word and its nearest rival – *troop* rather than *troot*, or *sardi*(*na*) rather than *sarda*(*na*) – is by definition a phonemic one, the importance for listeners lies not in deciding what phonemes they have heard, but in deciding between words. The minimal elements that can be matched to memory representations are those that are stored in the lexicon; incoming speech information is processed with respect to the information it gives about these representations.

2.2. Levels of processing

Although the incoming speech information is processed in a continuous manner, so that initial contact to stored lexical forms involves no conversion into any intermediate representation, listeners do draw on abstract representations of the phonemes of their language. They can, for instance, learn that a particular talker has a deviant pronunciation of a certain phoneme, and they can adjust their phonemic categorisations for that talker's speech very rapidly and on the basis of very little experience (Norris et al., 2003; Eisner and McQueen, 2005). The adjusted categorisations then apply to tokens of the phoneme in any word, not just to the words already heard from that talker; so if an ambiguous sound between /f/ and /s/ has been heard in words like *gira*[f/s], it will be identified as that talker's version of /f/, and if it is then presented in the context [naif/s], the word will be understood immediately to be *knife* and not *nice* (McQueen et al., 2006). The phonemic difference for this talker between the adjusted /f/ and a real /s/ is exactly as effective in distinguishing between words in a cross-modal priming test as another talker's normal /f/ and /s/ (Sjerps and McQueen, 2010). Again, the point is that listeners are using phoneme information in the way that best enables them to recognise speech rapidly (in this case, learning from their experience of a talker's pronunciation so far, and using what has been learned to facilitate future processing of that person's speech).

The forms that participate in the competition process, the earliest stage of word recognition, are phonological only; they are separate from the conceptual representations associated with the same word (Norris et al., 2006). This too supports efficiency; the speech information may activate lots of alternative words, but this does not lead in turn to consideration in parallel of lots of alternative semantic interpretations of the message. Instead, the most well-supported alternatives are passed on to be conceptually evaluated in the context so far. Sentence context does not rule out alternatives in advance, but as soon as one word is better supported by the input, its mismatched rivals can be ruled out by contextual evidence too. Thus in *The crew mourned their capt-*, the continuation *captain* is far more likely than *captive*, but nevertheless both remain available until just a bit more vowel support for the contextually appropriate alternative arrives and rules out the alternative (Zwitserlood, 1989).

The picture of native spoken-word recognition that arises from these decades of research is thus one of maximal efficiency. No decisions are made too early but the decision-making is cascaded so that the information from speech processing flows on from stage to stage as rapidly as possible. There are many models of spoken-word recognition, and they differ in particular in whether they allow higher-level processing to exercise control over processing at lower levels (e.g., McClelland and Elman, 1986), or whether, as the evidence summarised here suggests, higher-level information does not rule out anything in advance but rather allows rapid selection between the best-supported alternatives (Norris, 1994; Gaskell and Marslen-Wilson, 1997; Luce and Pisoni, 1998). The most recent model (Norris and McQueen, 2008) allows various sources of information to be evaluated probabilistically, and this is also the first model which has fully incorporated frequency information, known to play a significant role in speech processing. All models agree on the necessary central features of the spoken-word recognition process, namely the simultaneous availability of, and competition between, multiple lexical alternatives.

3. The non-native listener

The listener's task in recognising speech is in principle the same in any language; the same constraints of vocabulary makeup, the nature of speech production, and inter-speaker variability apply whether the language is long-known, relatively new, or unknown. Nonetheless, there

are many ways in which the broad classes of listener compared in the literature we review – native versus non-native – can vary. Comparing native and non-native performance in speech perception can be rendered difficult by the fact that the latter broad grouping may be very heterogeneous; learners in the classroom are not like strict bilinguals, etc. In this section, we examine the principal relevant dimensions of variability that can affect speech perception – listener type, first language (L1) influence, and input differences. More general reviews of issues in second language (L2) perception can be found in (Flege, 1988; Strange, 1995; Markham, 1997; Bohn, 2000; Bohn and Munro, 2007).

3.1. Types of listener

In most research related to non-monolingual speakers there is considerable uncertainty over the terms which define the linguistic profile of such populations. A frequent division in the literature is that between native versus non-native speakers/listeners. In the case of speakers who are monolingual until adulthood or at least until adolescence, there is little doubt as to which of their languages is their native or first one. In these cases, languages learned after adolescence can be classified as second languages (L2s) or foreign languages (FLs). These terms are frequently used interchangeably but, at least for perceptual studies, it is necessary to make a distinction. In broad terms, whether a language is considered to be 'second' or 'foreign' can be defined as a function of geographical setting and amount of presence in the community. The term L2 applies to languages learned after the L1 is fully established, and which are in widespread use in the community where the speaker is located at the time of acquisition, as is the case for many immigrants. In contrast, a FL is not widely present in the speaker's environment, even if contact with it through the media or other sources is frequent. A FL is typically learned through instruction and lacks the massive, natural and native input which characterises L2 acquisition. This difference in input is what makes the distinction particularly relevant for phonetic studies. When dealing with L2 speakers we can assume that their behaviour is not provoked by defective input, whereas in the case of FL speakers, not only the quantity but even more so the quality of input maybe suspect and a dominant factor in speech perception.

It is usually assumed that speakers possess native competence in their L1. However, this is not always so. There are cases in which, through attrition (e.g., immigrants who lose contact with their L1) or interrupted exposure before full acquisition (e.g., children who stop receiving L1 input before the language has been completely established), individuals may not have so-called 'native' competence in their L1 or native language.

Another issue which is relevant to phonetic and perceptual studies is homogeneity of competence in the native population. In any one group of native listeners there may be relevant differences with respect to their hearing acuity, to their familiarity (perceptual competence and experience) with different varieties of the language, including non-native varieties, and with respect to their phonetic awareness, all of which can contribute variability in test results. The question of variable competence reaches an extreme in bilingualism. There are frequent cases in the literature in which L2 and even FL speakers are denoted 'bilingual'. While correct in the literal sense of possessing two languages, a close definition is required to properly assess the competence of listener populations used in perception studies. A strict definition of a bilingual supposes equal and native competence in the two languages (Bloomfield, 1933) which, it may be argued, is rather a rare situation. The term 'bilingual' is sometimes used for those situations in which the two languages are acquired at the same time, though a more precise denomination for this situation is 'simultaneous bilingualism'. In practice, most bilinguals have a dominant language for particular domains of life, which may vary at different stages of their life. This is different from the most active language at different points of their daily life depending on the topic, context and interlocutors. In fact it has been suggested that bilinguals move along a continuum of language activation, at one end of which only one of their languages is active and at the other end their activation is totally bilingual because they are operating simultaneously in the two languages (Grosjean, 2001). Just as language activation maybe seen as a continuum, bilingualism itself may be a scale ranging from true monolinguals to strict bilinguals (Macnamara, 1969; Garland, 2007). In between lie speakers with different degrees of competence in each of the two (or more) languages (however, see Grosjean, 2010).

Despite the complexity of the issue, when comparing across experimental studies it is useful to distinguish L2 speakers, FL speakers and those who have native-like or near-native competence in two or more languages. However, as we shall see, these distinctions are not observed frequently in the literature and the true nature of listener populations can be difficult to gauge from the written record. Likewise, without strict criteria of bilingualism, division points amongst speaker groups according to age of arrival (AOA) are somewhat arbitrary and variable. Again this makes comparisons amongst differing groups problematic, and means that careful comparisons between reported studies are needed to ascertain the linguistic profile of the analysed groups. Differences in results should not be attributable to methodological differences in subject selection, after the fact, as is often the case (see Grosjean, 1998). Given the many different criteria used in NN studies to select populations and the inter- and intra-study heterogeneity of listener populations, it is difficult to draw sound generalizations and comparisons about non-monolingual listener perception. In what follows we will use the term "bilingual" to refer to speakers with near-native competence in two languages, typically from childhood. For the later acquisition of another language we will use L2/FL jointly or separately as the need arises.

3.2. L1 influence

It is a long-held view that we hear foreign sounds in terms of our native language ones (Polivanov, 1932; Lado, 1957; Stockwell and Bowen, 1965): the L1 sound system acts as a 'sieve' for the perception of new sounds (Trubetzkoy, 1939). Later research has shown that this view is too sweeping and that there are other mechanisms present such as developmental and universal processes (Eckman, 1977; Wode, 1980; Major, 1998; Major and Kim, 1999). Nevertheless, the influence of the L1 on L2/FL sound perception has been found to be stronger than in other linguistic areas (Ioup, 1984; Leather and James, 1991; Ellis, 1994) and it informs most models of L2 speech perception. In this section we will analyse some of these models and examine some factors which may affect the weight of L1 influences.

The best-known models of L2/FL acquisition place L1 influences at their core. In brief, they propose that, at least initially (Major, 2001), target language (TL) sounds are interpreted in terms of the L1 sound system and the degree of difficulty in acquiring L2 sounds may be predicted or explained according to their similarity to or distance from learners' L1 sounds. Kuhl proposed the Native Language Magnet theory (Kuhl, 1993a) in which L1 sounds act as prototypes (magnets) in L2/FL acquisition. The prototype attracts perceptually sounds within its sphere so that differences are less noticeable the closer a sound is to the prototype. Thus, it is hard to separate perceptually L2 sounds from L1 ones if they are within the scope of the native prototype. If two TL sounds fall within the area of a single prototype, it will also be difficult to distinguish them because "exposure to a primary language distorts the underlying perceptual space by reducing sensitivity near phonetic prototypes" (Iverson and Kuhl, 1995, p. 561).

The perceptual assimilation model (PAM) (Best, 1995; Best et al., 2001) accounts for the (lack of) discrimination of TL sounds on the basis of L2 sounds' similarity (exemplar rating) to L1 sounds, using the framework of articulatory phonology. Learners compare the gestures and timings of their L1 sounds to those of the TL and this comparison results in TL sounds being interpreted as either (i) exemplars of a L1 category varying in goodness fit, (ii) as sounds which may not be categorized in terms of the L1 system because they are sufficiently different, or (iii) as sounds which do not fall within the learner's experience of any speech sounds (e.g. clicks, which speakers of Indo-European languages recognise as extra-linguistic). PAM predicts degrees of difficulty for the acquisition of NN sounds in pairs, depending on the similarity of each member of the pair to L1 sounds. The most difficult discrimination would correspond to 'single category assimilation' in which two NN sounds are perceived as being equally good or deviant exemplars of one TL category.

Flege's speech learning model (Flege, 1995) claims that the acquisition of L2/FL sounds follows three possible courses depending on whether a particular sound is perceived by the learner to be totally different (new), identical or similar to one of the L1's sounds. Identical sounds may be transferred correctly from the L1 to the TL. New sounds are sufficiently unlike any TL one so that the learner will be aware of the differences and establish a new category for them. The greatest degree of difficulty will be faced with similar sounds, since they are erroneously assimilated to an L1 category from which it will be hard to separate them.

The ontogeny and phylogeny model (Major, 2001) proposes that, during the course of L2/FL speech learning, both L1 influences and latent universal developmental phenomena become apparent together with the influence of the TL sound system and, crucially, that L1 influences decrease as acquisition proceeds. Depending on the stage of acquisition, these factors have different roles and relative importance. When learning begins, the influence of the L1 system is the dominant force. Gradually, universal and TL phenomena manifest themselves. At the end of the learning process, the TL system is dominant while L1 influences and universal phenomena are no longer in evidence.

3.3. Input differences

A key difference between L1 and L2/FL acquisition is the starting point of language acquisition. The critical period hypothesis (Lenneberg, 1967) maintains that after a certain point in a person's maturation process, the ability to learn languages to a native-like standard is lost. The current consensus view is that, although it is not impossible to achieve native-like performance after a particular age, it is the case that for most individuals, native-like competence, and in particular native-like phonological competence is not possible if languages are learned after childhood. Puberty has traditionally been considered to be the cut-off point (Scovel, 1988), although it is currently believed that different linguistic abilities may have different sensitive periods, and that speech is the earliest (Seliger, 1978; Walsh and Diller, 1981). By 4–6 months of age, infants show preferences for the sounds of their native language (Kuhl, 1993b) or for languages rhythmically similar to their L1 (Bosch and Sebastián, 1997).

The reasons given for a sensitive period for languages have ranged from loss of brain plasticity or neurological specialisation (Lenneberg, 1967; Long, 1990; Penfield and Roberts, 1959; Singleton, 1989; Walsh and Diller, 1981) to deep-seated L1 habits (Flege, 1987, 1999; Bohn, 2000). Kuhl (2000, 2004) offers a comprehensive account in which both neurological and input factors account for the age factor. In her view, exposure to the L1 very early on creates neural patterns corresponding to what has been learned and these patterns then act as a filter which interferes in the processing of later stimuli which differ from the established patterns.

Another important difference between L1 and L2/FL acquisition which bears on the configuration of the TL sound system is the amount of exposure (Flege et al., 1999; Flege and Liu, 2001). The amount of native speech input obtained varies with the age of the learners and their

occupation and social contacts. In general, late learners do not receive as much native speech input as natives and early bilinguals, who interact abundantly through school and other activities from the beginning whereas adult immigrants are more likely to maintain contact with other L1 speakers. Quantity can be viewed as a scale ranging from minimal aural exposure (e.g. old-fashioned FL teaching methods based solely on grammar and translation) to total immersion in the TL natural context, with 100% of the learners' speech interaction being carried out in the TL. An L2 acquisition context is a necessary but not sufficient condition for the latter.

Another factor which differentiates L1 from later language acquisition is the quality and diversity of the input, ranging from single-source, non-native, heavily L1-accented pronunciations of the TL found in some FL learning situations to diverse, native and variable speech characteristic of natural contexts. Training studies have shown that for the formation of robust categories, diverse native input is necessary, which is the basis of the high-variability training methods introduced by Pisoni and his colleagues (Logan et al., 1991; Lively et al., 1993, 1994). It is notable that there have been few studies of the effect of simulated adverse conditions on training.

3.4. Native versus non-native spoken-word recognition

Given the parallelism of the listener's task in recognising speech in a second or a first language, certain skills already in place for listening to the native language will be available when second-language speech is heard. The architecture of the spoken-word recognition system is in principle not language-dependent; multiple activation of word candidates and competition between them will work whatever the vocabulary. Nonetheless, this ready availability does not necessarily work in the non-native listener's favour.

First, consider the crucial role of phonemic distinctions in modulating the availability of candidate words, and especially in getting rid of potential candidates as soon as a mismatch between input and stored representation is detected. This function depends upon accurate phonemic perception. Yet, as described in Section 3.2 above, wherever the phoneme repertoires of L1 and L2 differ, phonemic perception is notoriously inaccurate.

Second, consider the inter-word competition that is the basis of all word recognition. The vocabulary of an L2 user may be very impoverished compared with that of an L1 listener, so that the correct candidate may not even be available, or the competition may be skewed. But far worse, the L2 listener's competitor set may contain candidates that would simply not bother a native listener. This may on the one hand be words that are not even in the input language – namely, words of the L2 listener's L1. It has repeatedly been demonstrated in non-native listening experiments with advanced learners of an L2 that L2 input can call up words from the L1 vocabulary in parallel with words from the L2 vocabulary. For instance, in experiments of the kind described in Section 2.1, in which listeners' looks to various displays are recorded, Russian listeners to English have been found to look at a stamp (Russian: *marka*) when they hear the English word *marker*, and Dutch listeners to English have been found to look at a lid (Dutch: *deksel*) as the English word *desk* is spoken (Spivey and Marian, 1999; Weber and Cutler, 2004). On the other hand, candidates in the L2 may be unnecessarily activated if the listener's phonemic perception fails to rule them out. If a listener to English cannot tell the difference between /r/ and /l/, for example, then the input *regis-* may activate candidate words beginning with /l/ as well as words beginning with /r/ – both *register* and *legislate*, for example, may be activated, and competing with one another. The listener will not be able to resolve the competition until the sixth phoneme (e.g., *regist-*) whereas the native listener would have resolved that difference on the first phoneme. This also has been demonstrated in the laboratory (e.g., Cutler et al., 2006).

Each of these factors induces far more competition for the L2 listener than for the L1 listener. In cross-modal priming experiments of the type described in Section 2.1, this added competition has been shown to be not only present in L2 listening, but also especially hard to get rid of (Broersma and Cutler, 2008, 2010; Cooper et al., 2002).

Third, to complicate the issue still further, consider that, as described in Section 2.2, the information processed in listening flows in a cascade through the speech recognition system, so that higher-level processes can resolve lower-level alternatives with optimal efficiency. If the L2 listener's phonemic processing is inaccurate, more alternatives are passed on to the word recognition stage, and if the word recognition stage involves more competition than it should, sentence-level context has more work to do to sort it out. In other words, at all levels there will be more uncertainty to be resolved for the L2 than for the L1 listener. Yet the higher-level processes that are here called upon to do more work than is in principle necessary are also themselves less efficient in the case of the L2 listener. L2 listeners do not have the extensive experience with lexical and syntactic transitional probabilities that an L1 listener can call upon, they do not have as fine a sensitivity to the subtle syntactic and pragmatic differences that choice of words can convey, and so on. Again, there is abundant empirical evidence for the poorer syntactic processing skills of the L2 listener (Clahsen and Felser, 2006), and the L2 disadvantage here can even be observed when L2 proficiency is high (Sorace, 1993). A similar L2 disadvantage can be seen with the processing of idioms (Vanlancker-Sidtis, 2003; Cieslicka, 2006), and with prosodic processing, at the word level (Tremblay, 2008), at the phrasal level (Harley et al., 1995), and at the semantic level (Akker and Cutler, 2003). Thus while L2 listeners are forced back on higher-level context to resolve persistent uncertainty resulting from the disruptive effects of noise, their higher-level resources also let them down. We return to this issue in Section 6.

4. The effect of adverse conditions on speech recognition

Adverse conditions experienced in everyday speech understanding result in added energy from other source sources, reverberant energy from reflecting surfaces, and channel distortions. In many situations these elements act in concert (e.g. listening to an announcement via a public address system or using a mobile telephone in a social space). We do not review here every type of adverse condition, but concentrate on the conditions that have overwhelmingly been used in studies of non-native speech perception, namely additive noise sources, plus (to a lesser extent) the effects of reverberation. We first consider the listener's task in recognising speech in the presence of added noise (particularly the differential impact of speech and non-speech maskers), and in the presence of reverberation.

4.1. Additive noise: energetic masking

Most environments contain energy intruding from other sound sources at a range of signal-to-noise ratios (SNRs) and varying in attentional capture capacity from sources with a slowly-varying spectrum to competing speech. Sound source distance is often a critical factor too: even relatively quiet noise sources close to the listener can make a more distant target source more difficult to comprehend. In the following we adopt a wide definition of noise to mean any signal other than the target.

Fig. 1 illustrates predicted masking effects of the nonsense sentence *His travels show in our fear* from the Picheny corpus (Picheny, 1981) by speech-shaped noise, 6-talker babble and competing speech. In each case, the noise signal was added to the target speech to produce a global SNR of

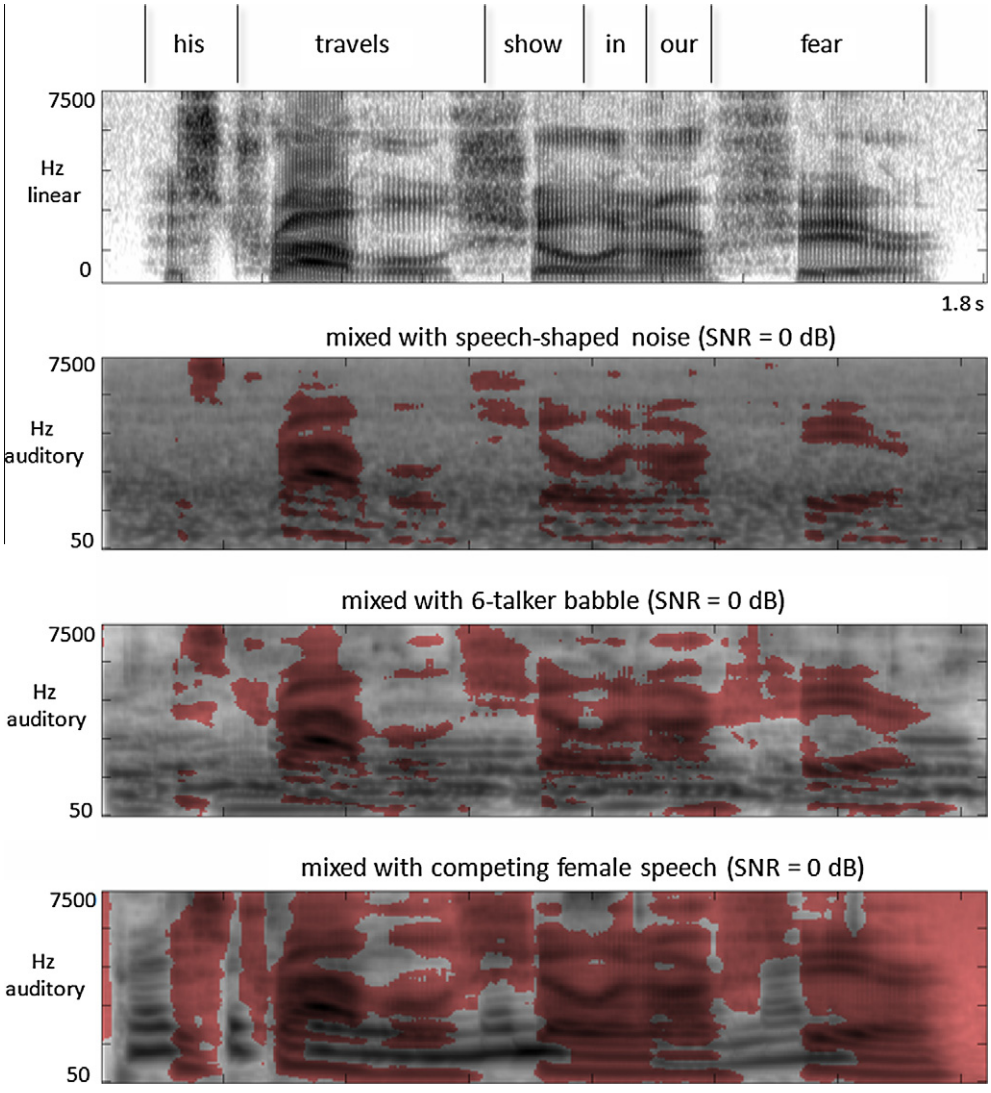

Fig. 1. *Upper panel*: Conventional spectrogram of the utterance "his travels show in our fear" spoken by a male American English speaker. The remaining panels depict auditory spectrograms of this utterance mixed with speech-shaped noise, 6-talker babble and competing female speech, in each case at an SNR of 0 dB. The red patches highlight those time–frequency regions where the male speech contains more energy than the masker. The frequency range for the auditory spectrograms is linear in ERB-rate: the middle of the range corresponds to about 1000 Hz. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0 dB, which is typically near the middle of the range of SNRs employed in non-native studies. The regions picked out in red indicate those portions of the time–frequency plane where the target speech has more energy than the masker. These regions are derived from a "glimpsing" account of speech perception in noise based on the idea that the auditory system makes use of only those time–frequency regions where the target source is locally dominant (Cooke, 2006). Since two sound sources are unlikely to possess similar amounts of energy at any given time–frequency location, the speech signal is either clearly 'visible' or masked in nearly all spectro-temporal locations: the proportion of ambiguous locations is very small. Regions which survive masking are shown superimposed upon *auditory* spectrograms since this is the domain in which the masking model operates. For reference, a conventional spectrogram of the clean utterance is shown in the top panel of Fig. 1.

Fig. 1 depicts *energetic masking (EM)*, a form of masking caused by the interaction of speech and masker at the level of the auditory periphery. Energetic masking affects speech perception by rendering unavailable potential cues to the identity of segments and their boundaries as well as interfering with access to prosodic cues (although complete energetic masking of prosody is uncommon due to the longer-term nature of suprasegmental information). Sounds such as the dental fricatives which are inherently low in energy are most susceptible to EM. In addition, the spectral profile of the masker also dictates the likelihood of cues surviving in different frequency regions. White, pink and speech-shaped noise have a different masking effects which influence segmental confusions in speech perception.

Fig. 1 illustrates several features of masking by additive noise. The overall quantity of information available about the target speech signal is not determined solely by SNR but depends on the properties of the masker. Speech-shaped noise obscures significantly more information than a competing talker, with babble noise producing an intermediate degree of masking. While long runs of all three maskers would have the same long-term spectrum equal to the average spectrum of speech, they differ both in degree of temporal modulation and preservation of spectral detail. The spectro-temporal variation in masker energy is greatest for the competing speech source and least for speech-shaped noise. In essence, for stationary maskers like speech-shaped noise, glimpses are determined by the modulations of the target speech, while for nonstationary maskers it is the interaction of target and masker modulations which define those parts of the target which lies above the noise. Thus, for speech-shaped noise, only those parts of the speech signal which are intrinsically energetic (typically resolved harmonics, formants and strong frication) are likely to survive at SNRs below 0 dB, while for a modulated masker, weaker target signal components may well be audible if they coincide with a spectro-temporal gap in the masker. In the example utterance, cues such as the /t/ burst in *travels* and the weak fricative in *fear* are virtually obliterated by the stationary masker yet retain some presence for the other maskers.

Temporal modulations in the competing speech masker produce epochs when information from across much of the spectrum is visible. Indeed, at 0 dB SNR, most words from each of a pair of talkers are intelligible, demonstrating that less than half of the time–frequency plane is perfectly adequate for speech perception for listeners with normal hearing. Masker modulation depth declines as increasing numbers of talkers are added to create babble until, in the limit, a stationary signal is reached. For six talkers, significant modulations remain, giving rise to a greater number of glimpses than in the stationary speech-shaped noise case. Studies of the effect of masker modulation on speech perception (Miller, 1947; Festen and Plomp, 1990; Simpson and Cooke, 2005) confirm that, when the global SNR is held constant, highly-modulated maskers such as speech or speech-modulated noise are significantly less effective than maskers with smaller modulation depth such as multi-talker babble or speech-shaped noise: typically, 6–8 dB of additional modulated noise is required to reduce sentence intelligibility to the level seen in the stationary condition. Spectral variations are also larger in the competing speech case than for speech-shaped noise since the latter represents an average spectrum while, for most speech sounds, at any instant a speech masker is likely to contain energy concentrated in certain spectral regions joined by spectral valleys of lower energy. These valleys provide an opportunity for glimpses of the target signal.

4.2. Additive noise: informational masking

Energetic masking is not the only effect of signal degradation; a second important aspect is the role played by any additional energy contributed by the masker. While the focus is usually directed towards the effect of masking on a target signal such as speech, masking is a two-way process: the noise is itself masked by the target, and audible masker fragments can interfere with the perception of the target. The regions not depicted in red in Fig. 1 are dominated by audible contributions from the masker. This information has the capacity to interfere with decisions at higher levels of processing, resulting in what has been called *informational masking (IM)*. IM was first studied for speech signals by Carhart et al. (1969) and more recently has been investigated by Brungart et al. (2001) and Freyman et al. (2004).

Unlike EM, informational masking covers more than one process whose relative contributions to a reduction in intelligibility remain poorly-understood. IM can arise due to supposedly low-level processes such as the grouping of time–frequency regions into larger units. Consider the spectral "holes" at the start of the word *fear* caused by the competing speech masker in Fig. 1. These contain audible speech components from the masker that need to be excluded from the target speech. The success or failure of grouping will depend on factors such as the similarity of

competing sources in F0 and vocal tract length (Darwin et al., 2003) as well as how much the patches of energy make sense in the ongoing context formed during interpretation of the utterance. Although competing speech or multitalker babble for smaller numbers of talkers are usually regarded as the most effective form of informational masker, it is worth noting that all maskers have some occasional IM potential: audible time–frequency regions belonging to stationary maskers could, in principle, incorrectly cohere with the target signal in a form of informational masking based on target-masker misallocation.

4.3. Reverberation

Sound reaching the listener from indirect paths following reflections contributes reverberant energy to the direct signal received at the ear. This additional energy can lead to masking of speech components. However, unlike additive noise, reverberant energy is correlated with the sound source which produced it, leading to different masking patterns. Fig. 2 depicts the effect of moderate and high levels of reverberation on the utterance used in the previous section. These examples were produced by recording the clean signal reproduced over loudspeakers in a games room and a bathroom.

Reverberation smears energy in time, enhancing "horizontal" structures such as static formants and blurring "vertical" structure such as bursts and abrupt transitions. The effects of reverberation are felt as both within- and across-segment distortions. Information conveyed by time-synchronous activity is particularly affected: offsets are less precise and bursts are smoothed. Vertical structure cueing the /t/ burst in *travels* is absent even in moderate reverberation. Onsets are affected to a lesser extent, particularly when the dominant frequency components of the preceding sound occupy a different part of the spectrum. For example, formant onsets in *show* are well-preserved in moderate reverberation, although in the high reverberation condition energy arriving from the previous word blurs the inter-word boundary.

Low-energy regions such as those resulting from plosive closures are susceptible to energy invading from preceding segments (see, for example, the initial weak fricative in *fear*). Within segments, formant transitions are broadened and, in the limit, transitional information is lost altogether. The second formant of the diphthong in *show* provides evidence of these kinds of distortion. By contrast, static vowel formants are enhanced and vowel identification typically suffers little in reverberation (Nabelek, 1988).

Reverberation affects prosodic as well as segmental information. Temporal smearing reduces the availability of cues to duration and rhythm. Information about F0 is distorted in both the low frequency region, where the motion of resolved harmonics is affected in the same way as formant transitions, and also at higher frequencies, where the modulation depth of amplitude changes at the period of the fundamental is reduced. The disruption of harmonic relationships is particularly evident in the more severe reverberation condition shown in Fig. 2. The effects of reverberation are particularly detri-

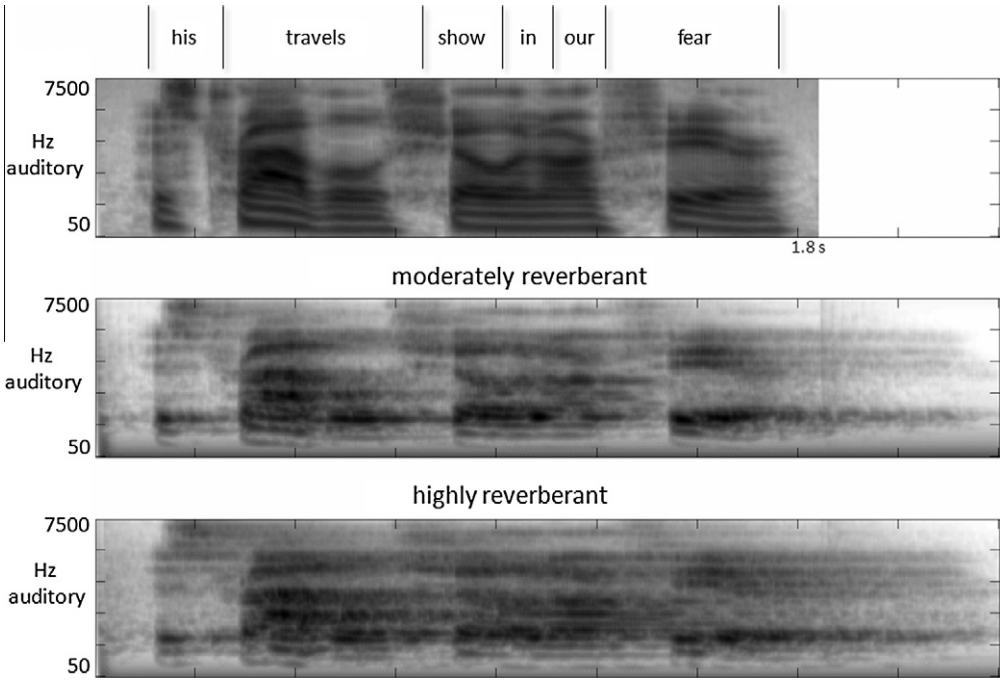

Fig. 2. Auditory spectrograms of the utterance "his travels show in our fear" recorded in near-anechoic conditions (top), then reproduced via loudspeakers and recorded in a 5 × 9 m games room (middle) and a 7 × 3 m bathroom (lower).

mental to speech perception in children (Neuman and Hochberg, 1983), listeners with hearing impairment (Nabelek and Pickett, 1974) and older listeners (Nabelek and Robinson, 1982).

4.4. The effects of adverse conditions on spoken-word recognition

The processes described in Section 2 are affected by the presence of noise; evidence on this issue is available from the many studies of human speech recognition by native listeners that have used noise as a diagnostic tool, to examine the distribution of information in the speech signal, or to test between alternative models of speech processing.

The neighborhood activation model of spoken-word recognition proposed by Luce and Pisoni (1998), for instance, was based *inter alia* on data showing that the relative accuracy and the range of responses for noise-masked spoken words reflected the density of their lexical neighborhood, and the relative frequency of the word itself and its neighbours. It has long been known that noise masking can produce a greater advantage for frequent over uncommon words (Howes, 1957; Savin, 1963), but this effect disappears if listeners are choosing from a known set of words rather than an open set (Pollack et al., 1959). Thus there are strong effects of guessing under uncertainty. Similarly, voice familiarity affects recognition under noise; speech by a single speaker is better recognised than speech by multiple speakers (Mullennix et al., 1989), speech in previously learned voices better than speech in new voices (Nygaard et al., 1994), and speech in common dialects better than speech in less familiar dialects (Clopper and Bradlow, 2006).

Speech is a temporal signal, so that the beginnings of words are heard before the ends; but this does not mean that stored representations of words in memory can only be successfully contacted if speech information arrives in this canonical order. There is substantial evidence that listeners are able to exploit whatever information is on offer – for instance, that they can extract useful information from the ends of words if they did not hear the beginnings because of noise interference. Van der Vlugt and Nooteboom (1986) found that beginnings and ends of synthesised words were equally effective prompts for correct recognition responses when the rest of the word was masked with noise. Slowiaczek et al. (1987) measured word recognition in noise as a function of prior primes. They found similar effects of primes which overlapped with either the beginning or the end of the target, whether by a single phoneme (e.g. *flock* preceded by *fret* or *steak*), two phonemes (*flap, stock*) or three (*flop, block*). The usefulness of all relevant information fits well with our current understanding of spoken-word recognition as described in the preceding section. Effects of informational masking have been argued to depend in part on competition from lexical items activated by the masking speech (Hoen et al., 2007).

Native listener adjustment to noise is in fact very sensitive, and also accords with our understanding of the spoken-word recognition process. Consider the Lombard effect, whereby people automatically start to talk louder when there is masking noise (see Lane and Tranel, 1971). They can also adjust their speech to maintain intelligibility against differing kinds of noise, e.g., by adapting the formant frequencies of their vowels (Van Summers et al., 1988). Listeners can likewise modulate the probability of words in the competition process as a function of masking noise. If AM radio crackle is presented as a masker, so that it occasionally masks a phoneme here or there in the input, listeners reduce their certainty about the words they hear, including the words actually unaffected; in an experiment in which looks to a display of objects or printed words are recorded, the word onset usually determines where listeners' looks go, but with the radio crackle there were more looks than otherwise to words with a different onset (e.g., to a *tent* when the input was *cent*; McQueen and Huettig, submitted for publication).

Many studies, finally, have used temporary or complete noise-masking to address the question of whether spoken-word recognition models should allow higher-level influence on lower-level processing. The first outcome of this type of work was the "phoneme restoration" phenomenon discovered by Warren (1970). When part of an utterance is excised and replaced with a noise that sounds like a cough, listeners report hearing a cough simultaneously with an intact speech signal – something which, after all, they have probably often experienced hearing. That is, listeners "restore" the sound, which was actually completely excised, to its rightful place in the utterance. The listeners are also not very good at judging exactly where in the utterance the cough occurred.

An extension of the technique was introduced by Samuel (1981), who compared listeners' judgements of stimuli in which the noise actually replaced a phoneme (as it did in Warren's stimuli), versus stimuli in which the noise was overlaid on an intact speech signal (as Warren's listeners reported hearing). Samuel argued that the phoneme restoration effect might arise due to direct influence from the lexical level on perceptual judgements about phonemes; in this case, the replaced and overlaid versions of a word should sound the same. He therefore examined the discriminability of the two types of stimuli, using signal detection techniques. The discriminability between replaced versus overlaid phonemes turned out to be significantly worse when the phonemes were in real words rather than in nonwords. Samuel argued that these results indeed indicated lexical influence; nonwords show no reduction of discriminability because nonwords have no lexical representations which can exercise influence.

However, note that discriminability here concerns an individual masked phoneme, not the word as a whole. Repp and Frost (1988) compared the detectability of whole real words and nonsense words under complete noise masking. Although words are more likely to be correctly

identified in noise than nonwords are, this could be because in uncertainty listeners tend to guess real words, as pointed out above. Repp and Frost argued, like Samuel, that direct lexical influence on the earliest stages of speech perception should make real words more detectable than nonwords when the listeners' task is simply to decide as to the presence of some spoken signal underneath the noise. However, they found no difference at all: nonwords and words were equally detectable. Thus the discriminability advantage is limited to phonemes within words. Warren's original explanation of phoneme restoration was that judgements of noise location are based on the signal itself, not on a derived representation of it; after a word has been recognised, the acoustic trace of it in the speech signal is no longer needed and is then no longer available for the performance of precise location determination. By studying the characteristics of restoration effects when rate of speech was varied, Bashford and Warren (1987) and Bashford et al. (1988) showed that the limits for illusory continuity corresponded to the average word duration in any given utterance. Once a lexical representation has emerged as the winner of the competition process, it is thenceforth the only representation available to the processor. For native listeners, in other words, noisy or other adverse conditions do exert a detrimental effect on the processes of speech perception, but recovery is efficient and relatively rapid.

5. Non-native speech perception in adverse conditions

As we began our review: the effect of noise on speech perception is much greater for non-native than for native listeners, and this has been familiar as long as there have been listeners to a second language. The question was, however, first addressed in the laboratory by Black and Hast (1962), now nearly half a century ago. Their study compared word reception scores in quiet and white noise at SNRs of +4, 0 and −4 dB. A modest disadvantage for non-native relative to native listeners in quiet of about 11 percentage points grew to 16, 25 and 29 percentage points with increasing noise.

Since then, the topic has been explored systematically, especially in the last decade. As already noted, non-native studies on this topic have focused on additive noise sources, with only a few considering reverberation, and only one investigating the combined effect of noise and reverberation (Rogers et al., 2006). Channel distortion has received very little attention apart from an early study by Singh (1966), while Bergman (1980) employed interrupted speech, finding no significant non-native disadvantage. Two further aspects of real speech usage are similarly absent from most studies, namely the effect of communicative context and noise on speech production (see, e.g., Garnier, 2007; Lu, 2010 for recent summaries). The effects of noise, reverberation and channel distortion on speech perception and production by native listeners are examined in detail in Assmann and Summerfield (2004), while Darwin (2008) provides a comprehensive review of listening to speech in the presence of other sounds, including the special case of interfering speech maskers.

Table 1 lists many of the published articles (but excluding those in the current Special Issue) addressing non-native speech perception in adverse conditions. Some commonalities and trends are evident in this table. English is nearly always used as the target language, perhaps raising a question concerning applicability of findings to other languages, particularly to those with tone-based distinctions. The type of noise employed has changed over the years, with predominantly white or pink noise giving way to speech-shaped noise and more informative maskers such as low-order babble and competing talkers. Tasks have varied from low-level consonant determination to the identification of keywords in low/high predictability sentences.

Several types of factor are at play in determining non-native listener performance in the studies listed in Table 1: (i) *listener* factors such as exposure, L2 competence and age of acquisition; (ii) *noise* factors such as type and level of noise or reverberation; and (iii) *task* factors which define the amount and kind of speech processing knowledge required to identify consonants, words or sentences. We return to the explanatory weight of these factors in Section 6. In the present section we describe the empirical studies, grouped principally by type of noise, but we begin with a summary of the principal findings.

- Non-native listeners suffer more from increasing noise than natives when the task involves words (Black and Hast, 1962; Gat and Keith, 1978) or sentences (Bergman, 1980; Meador et al., 2000) but for tasks which minimise or eliminate the possibility of using high-level linguistic information, noise has an *equivalent* overall effect on native and non-native listeners (Flege and Liu, 2001; MacKay et al., 2001a,b; Cutler et al., 2004; Rogers et al., 2006).
- There are indications that reverberation affects non-native listeners more than natives for low-level tasks involving consonant identification (Nabelek and Donahue, 1984; Takata and Nabelek, 1990).
- Intelligible maskers in an L2 appear to reduce L1 interference and hence aid non-native listeners (Rhebergen et al., 2005; García Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007).
- However, competing speech in the background increases overall informational masking for non-natives more than for natives (Cooke et al., 2008, 2010).
- In noise, early bilinguals are better than later L2 learners in the perception of sentences (Mayo et al., 1997; Weiss and Dempsey, 2008), words (Meador et al., 2000) and segments (Mackay et al., 2001a,b), though not for all positions (Mackay et al., 2001a).
- Even bilinguals from infancy manifest differences in speech perception in noise as compared to monolinguals, although these differences are not apparent in quiet nor in speech production (Mayo et al., 1997; Rogers

Table 1
Studies of non-native speech perception in adverse conditions.

Study	Target lang.	Listener L1s	Noise type (s)	SNRs (dB)	Task
Golestani et al. (2009)	English, French	French	SSN	−4, −5, −6, −7	Word recognition in sentences with semantically related/unrelated primes
Weiss and Dempsey (2008)	English, Spanish	Spanish–English bilinguals	SSN	Variable (SRT)	HINT
Cooke et al. (2008)	English	English, Spanish	(i) SSN and (ii) sentences	(i) 6, 0, −6 and (ii) 6, 3, 0, −3, −6, −9	GRID (keywords in short sentences)
Cutler et al. (2008)	English	Dutch, English, Spanish	8-Babble	0	16 consonants in VCV position
Jones et al. (2007)	English	English plus non-native group	Multitalker babble	6	Synthetic speech comprehension for short descriptions (mean 47 words)
Bradlow and Alexander (2007)	English	English, various NNLs	SSN	2 (NNs), −2 (Ns)	High- and low-predictability sentences
Cutler et al. (2007)	English	English, Dutch, Spanish	(i) SSN and (ii) CS	0 dB	23 consonants in VCV position
Van Engen and Bradlow (2007)	English	English	(i) Mandarin 6-babble and (ii) English 6-babble	5, 0, −5	Revised BKB
Rogers et al. (2006)	English	English, Spanish–English biling.	(i) SSN and (ii) SSN + reverb	(i) 0, −2, −6 and (ii) 4, 2, 0	Words
Garcia Lecumberri and Cooke (2006)	English	English, Spanish	(i) SSN, (ii) 8-babble, and (iii) CS	0 for all	16 consonants in VCV position
Rhebergen et al. (2005)	Dutch	Dutch	(i) Swedish CS and (ii) Dutch CS	Variable (SRT)	Short sentences
Cutler et al. (2005)	English	AmEng, AusEng, Dutch	6-Babble	16, 8, 0	Vowel in VC and CV
Imai et al. (2005)	English: native and Spanish-accented	English, Spanish	12-Babble	14	1-Syllable words
von Hapsburg et al. (2004)	English	English, Spanish–English biling.	SSN	Variable (SRT)	HINT (sentences)
Cutler et al. (2004)	English	English, Dutch	6-Babble	16, 8, 0	Consonants and vowels in CV and VC syllables
Bent and Bradlow (2003)	English (spoken by N and NN)	English, Chinese, Korean	White	5	Revised BKB
Bradlow and Bent (2002)	English	English, various NNLs	White	−4, −8	Revised BKB (simple sentences)
Shimizu et al. (2002)	English	Japanese	(i) White, (ii) pink, and (iii) aircraft	6, 1, −4	CID (words)
Wijngaarden et al. (2002)	Dutch, German, English	Dutch	SSN	Variable (SRT)	
Mackay et al. (2001a)	English	English, Italian	Pink	12, 6, 0, −6	18 consonants in initial and final positions of non-words
Mackay et al. (2001b)	English	English, Italian	Pink	16, 10, 4	Six plosives in bisyllabic nonwords in initial, medial and final position
Flege and Liu (2001)	English	English, Chinese	Pink	16, 10	Six plosives in word final bisyllabic non-words
Meador et al. (2000)	English	English, Italian	Pink	6, 0, −6	Semantically-unpredictable sentences
Hazan and Simpson (2000)	English	English, Spanish, Japanese	SSN	0	VCVs with enhancement
Mayo et al. (1997)	English	English, Spanish–English bilinguals	Babble	Variable (SRT) listener-dependent (range −14 to +8)	SPIN
Takata and Nabelek (1990)	English	English, Japanese	(i) Babble and (ii) reverb	(i) −3 and (ii) $T60 = 1.2$ s	MRT
Buus et al. (1986)	English	English, French	White	Variable (SRT)	Simple sentences
Florentine (1985)	English	English, various	Babble	Variable, subject specific	SPIN (final noun in sentence)
Florentine (1984)	English	English, French	Not stated	Variable (SRT)	Simple sentences
Nabelek and Donahue (1984)	English	English, various NN	Reverberation	T60 = 0.4, 0.8, 1.2	MRT (consonants from words in carrier sentences)
Bergman (1980)	Hebrew	Hebrew native born versus non-native born	(i) 15-Babble and (ii) temporal interruptions	(i) 3 dB and (ii) 10 Hz	Sentences
Gat and Keith (1978)	English	English, various NN	White	12, 6, 0 dB	Words (CID test W-22) with fixed precursor phrase
Spolsky et al. (1968)	English	English, various NN	White	1–10 dB	Sentences
Singh (1966)	English, Hindi	English, Hindi	(i) Temporal gating and (ii) bandpass filtering		CV for six plosives
Black and Hast (1962)	English	English, various NN	White	4, 0, −4 dB	Words

et al., 2006; von Hapsburg et al., 2004). The bilingual–monolingual difference is not observed if the bilinguals use the other language less frequently (Mackay et al., 2001a).

- Some of the errors found in speech in noise tasks are due to the influence of the L1 sound system (Mackay et al., 2001b; García Lecumberri and Cooke, 2006), and the relative degree of activation of the L1 correlates with the strength of its influence (Meador et al., 2000; Mackay et al., 2001a,b).
- Performance in speech perception tasks in adverse conditions correlates with the quality and quantity of input (Mayo et al., 1997; Rogers et al., 2006; Mackay et al., 2001a,b; Meador et al., 2000; Bradlow and Bent, 2002; Quene and van Delft, 2010).

5.1. Non-native speech perception under energetic masking

Until recently, non-native studies employed largely uninformative noise types such as stationary noise (white, pink or speech-shaped) or babble/cafeteria noise composed of six or more talkers. Lately, maskers with audible speech components, such as competing speech or a mixture of small number of talkers, have been used. The stationary/modulated and intelligible/unintelligible distinctions lead to different masking effects, as do the types of noise. For instance, Shimizu et al. (2002) found significantly lower scores for English word identification in white noise than in pink or aircraft noise when matched for overall level. Broersma and Scharenborg (2010) also highlight the highly-variable effect of different maskers which vary both in spectral and temporal patterning on sound perception in an L2.

Those studies using more than one noise type have demonstrated an equivalent NL versus NNL ranking of masking effectiveness of modulated and unmodulated noises. In a consonant identification task, both NLs and NNLs in (Garcia Lecumberri and Cooke, 2006) produced highest scores for a competing speech masker, lowest for 8-talker babble and intermediate results for speech-shaped noise, with all maskers presented at 0 dB SNR. This finding was extended to three listener groups in (Cutler et al., 2007) and listener groups from eight languages in (Cooke et al., 2010).

Within a single type of background, signal-to-noise ratio impacts directly on the availability of low-level information for speech judgements, although redundancy of cues allows native listeners to tolerate quite significant amounts of noise. Increasing noise levels might be expected to have proportionally more impact on non-native listeners. Several studies investigated the effect of SNR differences in native versus non-native listener speech perception. Two very clear conclusions emerge from these studies. Non-native listeners do indeed suffer more from increasing noise than natives when the task involves word or sentence processing in fixed or variable noise levels (Black and Hast, 1962; Gat and Keith, 1978; Florentine et al., 1984; Florentine, 1985; Buus et al., 1986; Mayo et al., 1997; Meador et al., 2000; van Wijngaarden et al., 2002; von Hapsburg et al., 2004; Cooke et al., 2008). However, for tasks which minimise or eliminate the possibility of using high-level linguistic information, noise has an *equivalent* overall effect on native and non-native listeners (Flege and Liu, 2001; Mackay et al., 2001a,b; Cutler et al., 2004; Rogers et al., 2006); even low-level predictability is better exploited by native listeners (Cutler et al., 2008). The finding that word and sentence processing is more adversely affected for non-native or bilingual listeners suggests that less effective use of phonotactic and contextual knowledge (particularly at the semantic level), is responsible for the non-native disadvantage. Recently, Golestani et al. (2009) isolated the contribution of semantics by measuring the effect of semantically-related or –unrelated items on previously-presented degraded words. Native language items facilitated the identification while non-native items produced a reduction in identification of the degraded words.

5.2. Non-native speech perception under informational masking

There are good reasons to expect informational masking to have a different impact on native versus non-native listeners. On the one hand, dealing with more than one intelligible speech source requires additional cognitive loading and the ability to correctly piece together fragments from each source to make a coherent interpretation, both of which might affect listeners more adversely when processing a second language. However, listeners might suffer more from competing attention if the masker is in a language with which they are familiar, while unfamiliar language interferers may be easier to block out. In other words: the first effect could cause more difficulty for non-native listeners, while the second effect could cause more difficulty for native listeners. Several recent studies have started to measure the effect of intelligible maskers on native and non-native listeners. One paradigm has the same listener group identifying items in their native language in the presence of native and non-native language maskers (e.g. Rhebergen et al., 2005; Van Engen and Bradlow, 2007) while another compares the effect of a single masking language and varies the L1 of the listener group (Cooke et al., 2008, 2010), or crosses masker language and listener group L1 (García Lecumberri and Cooke, 2006).

The first study to show an informational masking effect due to L1 interference was Rhebergen et al. (2005), whose Dutch listeners gained a benefit of 4.3 dB from time-reversal of Dutch speech but only 2.3 dB for Swedish material. Van Engen and Bradlow (2007) demonstrated that native English listeners are more adversely affected by English babble than Mandarin babble for two-talker babble but not for 6-talker babble, suggesting that audible words in babble have a greater impact if they are in the listeners' L1. In García Lecumberri and Cooke (2006), monolingual

native English listeners scored better on a consonant recognition task in the presence of a competing Spanish speech masker than for an English masker, while a non-native Spanish group of university-level English learners suffered equally from the two maskers. These studies all point to increased interference when the language of the masker consists of known, recognisable words, whereas unknown and unrecognisable speech is easier to block out. As Hoen et al. (2007) showed with native listeners, the more words potentially activated by the competing speech, the greater the interference. This effect thus causes more problems for native listeners.

The other possibility raised above, that *non-native* listeners might suffer more in the presence of competing speech, is also borne out by two recent studies in which native and non-native listeners identified speech in the presence of a speech masker. Cooke et al. (2008) used a classic informational masking paradigm (Brungart et al., 2001) where listeners heard pairs of simple syntactically-equivalent and semantically-neutral sentences such as "place red at B4 now" spoken by talkers with the same or different gender, or by the same talker, and presented at a range of SNRs. This task forces listeners to separate and track sentences based on low-level cues such as F0 and gender-related differences in formant spacing. In all speaker-pairing conditions (same/different gender, same talker) non-native listeners suffered significantly more than natives, with a disadvantage that increased markedly at higher masker levels. In (Cooke et al., 2010), eight listener groups differing in L1 identified consonants in a competing speech masker and in a speech-modulated noise maskers designed to produce the same amount of energetic masking. While scores for the native group were equivalent for the two maskers, most of the non-native listener groups showed a greater masking effect for the competing speaker. Moreover, the additional impact of competing speech showed a negative correlation with individual listener scores in noise, suggesting that competence in the target language also helps when dealing with the effects of a competing speaker.

The influence of target and background language similarity on speech perception in competing talker conditions was tackled by Van Engen (2010), who compared native English and native Mandarin speakers in identifying English sentences in the presence of either English or Mandarin 2-talker babble. Both groups suffered more in the English babble condition, suggesting that a common foreground–background language is detrimental regardless of whether it is in a listener's L1 or L2.

Recently, other forms of informational masking have been investigated. Mattys et al. (2010) tested the reliance of native and non-native participants on acoustic or lexical cues to word segmentation under conditions of cognitive load, provided by a simultaneous visual search task. Their results suggest that relative to native listeners, non-natives attend more to acoustic detail when a cognitive load is present, highlighting a reduced ability to make use of lexical information under noisy conditions, a finding compatible with the outcome of studies reviewed in the previous section.

In summary, the few studies of informational masking on non-native speech perception to date thus support the idea of both a facilitating effect for non-natives due to release from L1 interference, and a degrading impact due to other factors such as increased cognitive load when processing pairs of L2 sentences or their lesser experience with the masker language in speech separation. The relative contribution of these factors has yet to be clarified, and is likely to vary as a function of both situation-dependent factors (e.g., SNR) and listener-dependent factors (e.g., proficiency).

5.3. Non-native speech perception under reverberation

Compared to additive noise, relatively few studies have considered the effect of reverberation on non-native listeners (Nabelek and Donahue, 1984; Takata and Nabelek, 1990; Rogers et al., 2006). Of these, only the latter examined the most common listening situation in which the effects of noise and reverberation are combined. Using a consonant identification task, Nabelek and Donahue (1984) and Takata and Nabelek (1990) demonstrated that non-native listeners who performed at comparable levels to natives in anechoic conditions suffer substantially more in the presence of mild, moderate and severe reverberation, with the native advantage reaching a peak for the moderate level of reverberation. This finding contrasts with the additive noise case for consonants, described above, where no interaction of nativeness and noise level is typically observed. Indeed, Rogers et al. (2006) found no significant additional effect of reverberation on monosyllabic word recognition in speech-shaped noise for bilingual listeners who had learned English prior to the age of six.

It is not clear why reverberation might have a disproportionate impact on non-native listeners for tasks which focus on acoustic information at the segmental level while additive noise does not. In fact, since reverberation is a universally-experienced feature which listeners appear to compensate for, at least in part (Watkins, 2005), it might be expected to have similar effects on all listener groups. A comparison can be made with the use of low-level cues such as spatial separation (Von Hapsburg et al., 2004; Ezzatian et al., 2010) or F0 (Cooke et al., 2008) by non-native listeners, where no native advantage has been found. The suggestion that higher-level linguistic knowledge is required to mitigate the impact of reverberation would be an interesting finding. Given the level of reverberation in typical instructional settings (see Section 1), this is an area where further studies are needed with a wider variety of listener groups and tasks.

5.4. Noise and non-native phonetic cue perception

For tasks which involve making segmental distinctions (e.g., from nonsense or rhyming words, alone or in sen-

tence context), phonetic feature analysis allows a more detailed comparison of native- and non-native listener performance in noise. Feature analysis operates directly for tasks such as the diagnostic rhyme test (Fairbanks, 1958), where word pairs differ in a single feature such as voicing or place, or indirectly from consonant or vowel discrimination tasks where the proportion of transmitted information (TI) for each feature can be derived from the confusion matrix (Miller and Nicely, 1955; Wang and Bilger, 1973). Miller and Nicely (1955) reported that TI values for voicing and nasality survived noise better than affrication and durational features, with place-based confusions being most adversely-affected. However, a recent repetition by Lovitt and Allen (2006) found a substantially different feature ranking, with voicing worse than place in noise, a finding echoed by Van Dommelen and Hazan (2010) and Cooke et al. (2010). Many studies agree that sibilance is particularly robust in noise (Miller and Nicely, 1955; Wang and Bilger, 1973; Wright, 2004; Cooke et al., 2010). Wang and Bilger (1973) highlighted the importance of position in the syllable for consonant perception in noise. Hazan and Simpson (1998) found that voice was by far the best transmitted feature in noise followed by manner and lastly by place. They suggest that manner confusions between plosives and fricatives are particularly frequent in noise and that the best cues for place are bursts for plosives, the frequency of fricative noise for fricatives and in general F2 and F3 transitions to following vowels. Jiang et al. (2006) looked at factors and cues for plosive voicing distinction in noise. Whilst VOT duration was the best voicing cue in quiet conditions, the context /Ca/ promoted greatest intelligibility in noise, perhaps related to the fact that F1 onset is a good cue for voicing contrasts in noise. Additionally they found that plosive burst and aspiration are easily masked by noise, that F0 is not an important cue in noise and that the contrast is least robust in the case of bilabial plosives.

Comparisons of native and non-native sound perception demonstrate that native listeners use different cues and cue weightings and that non-natives' cue use comes to resemble the native pattern as competence in the L2 advances (Bohn and Flege, 1990; Fox et al., 1995; Bohn, 1995; Flege et al., 1997; Cebrian, 2006). Given that native listeners use different cues and cue rankings in noise compared to quiet, and that non-native listeners differ from natives also in the relative use of cues, it is reasonable to assume that NN will use different cues to natives in noise. However, it could be that noise induces the use of universal robust cues, leading to greater similarity in cue use in noise compared to quiet. Indeed, several recent studies in noise (e.g. Cooke et al., 2010; Van Dommelen and Hazan, 2010) have suggested the presence of universal acoustic confusions, which may be less apparent in quiet due to ceiling effects. In practice, a mixture of L1-influences and language-independent effects are evident. Spanish listeners suffered from poor voice reception in Hazan and Simpson (2000) compared to Japanese and English listeners, presumably due to the absence of voiced fricative phonemes in Spanish, but place was the least well transmitted feature for all three L1 groups, pointing to a basis in acoustic masking. Cutler et al. (2007) suggest that the effect of noise on native versus non-native cue use may differ depending on the relevance of the cue in the L1. In particular, they showed that the differential reliance on transitional cues, which are more fragile in noise, could explain the considerable drop in fricative performance for both native (English) and non-native (Spanish) listeners, who use these cues in their L1s for fricative distinctions. The other non-native group tested (Dutch), whose fricative inventory does not require these cues, was relatively less affected by noise. Heinrich and Hawkins (2010) also show differential use of cues in noise, in their case short and long-term *r*-resonances. While native English listeners experienced a strong facilitating effect of *r*-resonances in English sentences, only those German listeners with relatively little experience in English were able to exploit the long-term cues, while Germans with more experience in English ignored them.

6. Cross-study comparisons

The background review in the earlier sections of this article highlighted a number of ways in which research on native listening, or research on listening in the laboratory or in otherwise non-adverse conditions, has isolated variables which affect listener performance. A striking outcome of our review of the studies of non-native listening in adverse conditions has been that these variables are not always controlled as strictly as one would wish for comparability, both across non-native studies in noise and between these and other studies. We highlight some of these issues in the following sections.

6.1. Listener type

For the listener populations in the studies listed in Table 1, the most frequently reported data are age of arrival (AOA) and length of residence (LOR) in the L2 country. However, there is enough variation across studies that it is difficult to determine whether the target language (TL) of the listeners in a given study should be regarded as a FL or L2.

Mackay et al. (2001a) establish a cut-off point around puberty to distinguish between early and late bilinguals: early bilinguals had AOAs up to 13 and late bilinguals between 14 and 26. In this case a more strict use of terms would define the groups as bilingual versus L2 learners, respectively, although even some of the learners in the 'early' group might be considered L2 speakers rather than bilingual. Then again, in (Mackay et al., 2001b) and in (Meador et al., 2000) the (presumably same) sample of subjects were divided into three age of learning groups with the cut-off points at ages 7, 14 and 19, in which case the two older groups are best considered L2 speakers. Jones et al. (2007) also use an age around puberty (AOA = 15) as a

defining criterion for their late-learners. Flege and Liu (2001) refer to L2 speakers for groups of Chinese speakers of English arriving in America after the age of 21. Since in the previous studies, late 'bilingual' groups contained subjects with AOA up to 20 years, it may be the case that for these authors this is the cut-off point between bilinguals and L2 speakers. In general, Flege and his collaborators apply the term bilingual in a literal sense to refer to speakers of L2s, irrespective of their AOA.

This literal sense is also adopted by van Wijngaarden et al. (2002), who speak of trilinguals despite the fact that the second and third languages were learned in a formal setting in the L1 country from secondary school. Mayo et al. (1997) provide a clearer separation of listener groups and, at least in 2 out of 3 groups, the term bilingual seems appropriate (bilinguals from infancy and from toddlerhood). However, the group of post-puberty learners, who started learning their L2 ranging from age 14 and with a reported use of the language varying between 3 and 26 year is probably better defined as an L2 group, rather than bilingual. Rogers et al. (2006) follow this approach when selecting their bilingual group in enlisting only speakers of English and Spanish who were exposed to the two languages before the age of six.

Thus as noted earlier, comparisons across the groups tested in different studies of non-native speech perception in noise are problematic. Indeed, several authors (von Hapsburg and Peña, 2002; von Hapsburg et al., 2004; Weiss and Dempsey, 2008) advocate a thorough description and control of several speaker variables that introduce heterogeneity in subject sample populations and have been shown to influence linguistic performance. However, in their own work, language proficiency is measured by self-assessment, which is an intrinsically subjective measure and may vary culturally (Hazan and Simpson, 2000). Von Hapsburg et al. (2004) set out to obtain a homogenous bilingual sample by selecting immigrants to the USA arriving after age 10, a cut off point based on neurological maturation. However, their upper limit (age 20) may introduce heterogeneity since for some authors (e.g. Mackay et al., 2001a) the cut off between early and late 'bilinguals' is around puberty. Later teenager immigrants typically do not receive as much school immersion and formal education in the L2, which has been found to be a relevant variable (Flege et al., 1999; Flege and Liu, 2001).

In early studies particularly, NN listeners are described very succinctly. For example, Black and Hast (1962) provide no details concerning the L1, TL competence level or AOA of their 32 NN listeners. The only information provided is that they were 'foreign-born', which suggests their TL was a FL. Similarly, Spolsky et al. (1968) only indicate that the NN speakers were graduate students in the USA. The listeners in (Nabelek and Donahue, 1984) had differing L1s, learned English as teenagers and spoke it with a foreign accent. There, a minimum level of performance on the modified rhyme test (Kreul et al., 1968) was used a criterion for inclusion. Takata and Nabelek (1990) pool together listeners who could be considered to be FL speakers with others who could be L2 speakers. Their LOR in the TL country ranged from 1 to 13 years. The earliest AOA was 19, and all of them had studied the TL (English) in their L1 country (Japan) prior to arriving to the TL country. These details suggest that for some but perhaps not all of the 10 listeners, English could be considered a FL, but it is difficult to be certain. However, even in current studies there can be ambiguity. For example, participants in (Ezzatian et al., 2010) were university students distributed in four groups according to AOA and previous experience with the TL. However, since LOR is not mentioned, it is unclear whether for some of them the TL was still a FL. Mattys et al. (2010) studied listeners who had learned the TL as a FL in their native country but had since migrated to the TL, where they had been living between 3 and 10 years, which suggests their FL had become a L2.

Several studies tested listeners who learned the target language (TL) as an FL in their country of origin, and that was also where testing took place (Shimizu et al., 2002; van Wijngaarden et al., 2002; Cutler et al., 2004; Golestani et al., 2009). These are similar to the populations analysed by Hazan and Simpson (2000) and Bradlow and her collaborators (Bradlow and Bent, 2002; Van Engen, 2010), being mainly visiting students with a very short LOR, making results more comparable. The main difference is that the latter authors carried out testing in the TL country, so their listeners may have had more native input than the other studies since they were living there, albeit temporarily. In a later study (Bradlow et al., 2010) the authors have two listener populations tested in the aforementioned conditions (in the TL country where they have resided for under 1 month) and two other listener groups tested in their own NL country where they used the FL on a regular basis. In terms of TL sound identification, as we shall see below, the amount of native input is an important variable. Indeed, Bradlow and Bent (2002) consider that the initially surprising lack of correlation between 'classic' variables such as AOA or LOR and speech perception in noise is due to the fact that they are dealing with FL listeners with limited experience with the TL.

Imai et al. (2005) use a foreign accent rating test to divide the listeners into high and low proficiency L2 participants. Even more directly, some studies employ performance in quiet conditions as a classification measure for perceptual proficiency (Nabelek and Donahue, 1984; Takata and Nabelek, 1990; Buus et al., 1986; García Lecumberri and Cooke, 2006). In general, NN population proficiency classifications for perceptual research tend to be carried out less directly. NN competence is often estimated by self-assessment (Hazan and Simpson, 2000; van Wijngaarden et al., 2002; von Hapsburg et al., 2004; Weiss and Dempsey, 2008; Broersma and Scharenborg, 2010; Mattys et al., 2010), which has proved to be unreliable given its intrinsic subjectivity (Cooke et al., 2010) and the potential impact of cultural differences (Hazan and Simpson, 2000). Still, in other studies, competence is roughly

estimated (Mayo et al., 1997; Bradlow and Bent, 2002; Cutler et al., 2004; Golestani et al., 2009; Bradlow et al., 2010). Other authors make use of years of instruction (Florentine et al., 1984; Cutler et al., 2004; Lee et al., 2010; Mattys et al., 2010), LOR (Gat and Keith, 1978; Meador et al., 2000; Flege and Liu, 2001; Mackay et al., 2001a,b; Heinrich and Hawkins, 2010; Mattys et al., 2010), frequency of use or contact with the TL (van Wijngaarden et al., 2002; von Hapsburg et al., 2004; Rogers et al., 2006; Gooskens et al., 2010) or general proficiency tests (Bradlow and Bent, 2002; Bradlow and Alexander, 2007; Cooke et al., 2008; Bradlow et al., 2010; Van Engen, 2010; Volin and Skarnitzl, 2010). It is known that general competence is often unrelated to phonological competence (Scovel, 1969, 1988), and in particular precise timing control is a separate skill that non-native speakers often fail to master (Quene and van Delft, 2010); nonetheless some studies (Ezzatian et al., 2010) have found correlations between perception in adverse conditions and non-perceptual measures of competence (such as vocabulary).

6.2. Influence of L1 and AOA

Some studies which fall under the criteria of the present review had the goal of investigating the predictions made by the models of non-native speech perception described in Section 3.2. In those studies, noise is used to elicit differences in perceptual competence between populations which may not surface in quiet conditions. Flege and his collaborators tested the SLM using pink noise to avoid ceiling performance effects and explore in more detail subtle differences between monolingual and bilingual listeners (Mackay et al., 2001a,b; Meador et al., 2000; Flege and Liu, 2001). Hazan and Simpson (2000) looked at the effect of systemic differences between the L1 and the FL (finding, against their predictions, poor perceptual results also for sounds which are shared by the two languages). Van Dommelen and Hazan (2010) analysed consonant perception to observe the effect of adverse conditions on novel sounds versus sounds which exist in both language inventories, applying a category goodness fit criterion as suggested in PAM. Again, contrary to expectations but agreeing with Hazan and Simpson (2000), novel sounds present better discrimination in noise than shared categories, which the authors suggest could be due to differences in phonetic detail between the apparently identical categories. It seems that current L2 perception models do not well predict the results of speech perception in noise.

It has been suggested that the weight of the L1 differs according to the stages of acquisition. Major's (2001) model explicitly states that the influence of the L1 is greatest at the initial stages of acquisition and that it decreases as acquisition of the TL sound system progresses. If we see the advance of acquisition as a gain in linguistic competence, we could say that the weight of the L1 sound system is inversely proportional to the level of phonological competence in the TL. Hazan and Simpson (2000) found L1 interference to be stronger in adverse conditions than in quiet. However, the study by van Dommelen and Hazan (2010) found no increase in L1-based confusions in noise. The authors suggest that differences in the level of FL competence might explain the conflicting findings: listeners in the later study had a higher FL competence which led to less L1 interference.

Grosjean (2001) proposed a model in which the level of language activation or bilingual 'language mode' affects the quality of speakers' performance. Several studies have investigated the amount of L1 activation and its effect on L2 speech production and perception, based on the idea that a more active L1 system (through more exposure to, and use of, the L1) will have more influence on the L2/FL. There is evidence indicative of this being the case for L2/FL degree of foreign accent (Flege et al., 1995, 1997; Guion et al., 2000; Piske and Mackay, 1999). On the other hand, Flege et al. (1999) found no detrimental effect of L1 use on vowel discrimination for two groups of early L2 learners who differed in amount of L1 activation. However, these authors found the same subject groups to differ in consonant identification and word recognition when embedded in pink noise, with an advantage for the group who used the L1 less frequently (Mackay et al., 2001a,b; Meador et al., 2000). This is an interesting indication of how simulated adverse conditions might provide the fine-grained tool necessary to uncover subtle perceptual phenomena which are not apparent in quiet conditions.

Adverse listening conditions may themselves favour a greater L1 influence, since in challenging circumstances listeners might be expected to fall back on familiar categories and cues. Similarly, when insufficient information is available, listener cue use may be appropriate for their L1 categories but not necessarily for the TL (Heinrich and Hawkins, 2010). Energetic masking (see Section 4.1) has been shown to increase the relative weight of acoustic detail at the expense of lexical-semantic knowledge (Mattys et al., 2009) which in turn may reduce L2 activation. Takata and Nabelek (1990) consider that L1-induced confusions rise in degraded conditions. On the other hand, Cutler et al. (2004) and van Dommelen and Hazan (2010) find that the influence of the native system does not increase in noise relative to less adverse or quiet conditions. Noise seems to induce a good deal of similarity in responses from different listener populations (Cooke et al., 2010).

Amount of input has been varied in a number of the studies in Table 1. Mackay et al. (2001a) found differences in stop perception in quiet and noise between bilinguals and L2 learners. Mackay et al. (2001b) analysed English consonant perception in noise for three groups of L2 speakers who differed in their learning starting age, finding that early learners showed an advantage for onset consonant identification but not for coda consonants. More research is needed to elucidate the possible effect of L2 learning age on perception in adverse conditions. Other studies (von Hapsburg et al., 2004; Weiss and Dempsey, 2008) found that the age at which the L2 exposure started

is inversely correlated with performance in a L1 in noise perception test; that is, earlier bilinguals performed worse than late bilinguals in their L1, which is a sign of loss of competence in the L1 due to L2 interference. Finally, Flege and Liu (2001) presented final plosives for discrimination embedded in different levels of noise to Chinese listeners differing in length of residence in the TL country and in their occupation during their stay (students versus workers). Length of residence produced significant perceptual differences in two student groups but not in two worker groups, a lack of effect ascribed to the smaller amount of TL input for the latter groups. Mackay et al. (2001a) find differences in short-lag voiced plosive perception between native listeners and (early) bilinguals versus late bilinguals who have a more active L1 but not for late bilinguals with a less frequent use of the L1. Accordingly, these differences are put down not to age of learning itself but to the differences in quality and quantity of the input in favour of the early learners and less frequent L1 users.

6.3. The next steps

This review examined the impact of listener, task and noise-related factors on speech perception by non-native listeners in adverse conditions. We conclude by noting some of the limitations of existing studies and suggest research themes which may become important in the near future.

One key issue is realism. As noted at the outset, much of our experience as non-native communicators comes from everyday situations rather than formal teaching settings, yet most of what we know about the effect of adverse conditions has been gleaned from controlled laboratory conditions and narrow tasks. For instance, only one study to date has combined additive noise and reverberation even though that is by far the most common adverse condition. Noise also affects speech production (Lombard, 1911), so it will be important to explore how non-native listeners cope with speech modifications induced by noise. Little work has been carried out using conversational tasks, so we do not know how noise affects the ability of non-native listeners to process and generate high-level interactional cues (e.g. Schegloff, 2000) and the extent to which non-native listeners are capable of aligning in conversations (e.g. Pickering and Garrod, 2004).

One step towards more realistic tasks has been taken by Lee et al. (2010), who consider the case of processing multiple talkers rather than a fixed talker, demonstrating that changes of talker do affect native and non-native listeners equally. Another factor from which non-native listeners appear to derive equivalent benefit is the release from masking due to spatial separation of target and masker, irrespective of whether the masker was energetic or informational in nature (Ezzatian et al., 2010). The potential differential use of multimodal information is a further real-world aspect which is being tackled in non-native speech perception studies (e.g. Hazan et al., 2010).

The lack of unified criteria when selecting and describing NN populations makes some of the results less robust or reliable than they appear to be, as pointed out by van Wijngaarden et al., 2002), and renders comparisons between studies tentative (see Grosjean, 1998; von Hapsburg and Peña, 2002; von Hapsburg et al., 2004 and Weiss and Dempsey, 2008). Some of the variables and results that apply to L2 populations cannot be extrapolated to FL listeners because of their smaller degree of contact with the TL (Bradlow and Bent, 2002), which makes it crucial to distinguish the two types of population. A related issue concerns the distance between languages, which is predicted to affect how different NN populations perform in speech tasks. Bradlow et al. (2010) make a start at deriving a phonetic similarity space for languages.

While models of non-native sound perception exist, few cater for adverse conditions. An exception is that of van Wijngaarden et al., (2004), which adapts the speech transmission index (Steeneken and Houtgast, 1980) to non-native listeners in additive noise, reverberation and band-limiting using a single parameter modification to the L1 psychometric function. While the non-native index provides a global indication of intelligibility, we currently lack explanatory models which produce more detailed predictions of, for example, specific sound confusions suffered by non-native listeners in noise.

A feature of all studies involving non-native listeners is the large degree of between-listener variation in identification scores, with the most competent listeners often achieving native levels of performance. Further work is needed to study the relationship between individual performance in L1 and L2 tasks. A related issue concerns non-native listener baselines. Few studies have measured the performance of *monolingual* non-native listeners in low-level tasks involving L2 sound perception in noise. Also, the role of formal training in noise is currently unknown, but it is certainly worth asking whether exposure in adverse conditions can be beneficial, (for example, by highlighting robust cues to L2 distinctions).

Finally, although research focussing on the recognition of spoken words has addressed non-native listening in ever growing detail in recent years (see Section 3.4), and has also used noise as an important tool in critically distinguishing between theories (Section 4.4), the intersection of these topics (that is to say, lexical processing in adverse conditions by non-native listeners) is as yet untrodden research terrain.

Acknowledgments

M.L. Garcia Lecumberri and M. Cooke were supported by the EU Marie Curie RTN "Sound to Sense", Basque Government Grant IT311-10 and Spanish Government grant FFI2009-10264. We thank Madhu Shashanka for recording the reverberant sound example.

References

Adank, P., Evans, B.G., Stuart-Smith, J., Scott, S.K., 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. J. Exp. Psychol. Human Percep. Perform. 35, 520–529.

Akker, E., Cutler, A., 2003. Prosodic cues to semantic structure in native and nonnative listening. Bilingualism: Lang. Cognit. 6, 81–96.

Assmann, P.F., Summerfield, Q., 2004. The perception of speech under adverse acoustic conditions. In: Greenberg, S., Ainsworth, W.A., Popper, A.N., Fay, R.R. (Eds.), Speech Processing in the Auditory System, . In: Springer Handbook of Auditory Research, Vol. 18. Springer, Berlin.

Bashford, J.A., Meyers, M.D., Brubaker, B.S., Warren, R.M., 1988. Illusory continuity of interrupted speech: speech rate determines durational limits. J. Acoust. Soc. Amer. 84, 1635–1638.

Bashford, J.A., Warren, R.M., 1987. Multiple phonemic restorations follow the rules for auditory induction. Percept. Psychophys. 42, 114–121.

Best, C.T., 1995. A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. Issues in Cross-Language Research. York Press, Timonium, MD, pp. 171–204.

Best, C.T., McRoberts, G.W., Goodwell, E., 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. J. Acoust. Soc. Amer. 109 (2), 775–794.

Bent, T., Bradlow, A.R., 2003. The interlanguage speech intelligibility benefit. J. Acoust. Soc. Amer. 114, 1600–1610.

Bergman, 1980. Aging and the Perception of Speech. University Park Press, Baltimore, pp. 123–133.

Black, J.W., Hast, M.H., 1962. Speech reception with altering signal. J. Speech Hearing Res. 5, 70–75.

Bloomfield, L., 1933. Language. Holt, New York.

Bohn, O.S., Munro, M.J., 2007. Language Experience in Second Language Speech Learning: In Honor of James Emil Flege. John Benjamins, Amsterdam/Philadelphia.

Bohn, O.S., 2000. Linguistic relativity in speech perception: an overview of the influence of language experience on the perception of speech sounds from infancy to adulthood. In: Niemeier, S., Dirven, R. (Eds.), Evidence for Linguistic Relativity. John Benjamins, Amsterdam and Philadelphia, pp. 1–28.

Bohn, O.S., 1995. Cross-language speech perception in adults: first language transfer doesn't tell it all. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. Issues in Cross-Language Research. York Press, Timonium, MD, pp. 279–304.

Bohn, O.S., Flege, J.E., 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. Appl. Psycholinguist. 11, 303–328.

Bosch, L., Sebastián, N., 1997. The role of prosody in infants NL discrimination abilities. In: Eurospeech '97: Proceedings of the 5th European Conference on Speech Communication and Technology, vol. 1, pp. 231–234.

Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Amer. 112, 272–284.

Bradlow, A.R., Alexander, J.A., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Amer. 121, 2339–2349.

Bradlow, A.R., Clopper, C.G., Smiljanic, R., Walter, M.A., 2010. A perceptual phonetic similarity space for languages: evidence from five native language listener groups. Speech Communication 52, 930–942.

Broersma, M., Scharenborg, O., 2010. Native and non-native listeners' perception of English consonants in different types of noise. Speech Communication 52, 980–995.

Broersma, M., Cutler, A., 2010. Competition dynamics of second-language listening. Quart. J. Exp. Psychol.

Broersma, M., Cutler, A., 2008. Phantom word recognition in L2. System: Internat. J. Educat. Technol. Appl. Linguist. 36, 22–34.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Amer. 100, 2527–2538.

Burki-Cohen, J., Miller, J.L., Eimas, P.D., 2001. Perceiving non-native speech. Lang. Speech 44, 149–169.

Buus, S., Florentine, M., Scharf, B., Canevet, G., 1986. Native French listeners' perception of American-English in noise. In: Proceedings of InterNoise 86, Cambridge, USA, pp. 895–898.

Carhart, R., Tillman, T.W., Greetis, E.S., 1969. Perceptual masking in multiple sound backgrounds. J. Acoust. Soc. Amer. 45, 694–703.

Cebrian, J., 2006. Experience and the use of duration in the categorization of L2 vowels. J. Phonetics 34, 372–387.

Cieslicka, A., 2006. Literal salience in on-line processing of idiomatic expressions by second language learners. Second Lang. Res. 22, 115–144.

Clahsen, H., Felser, C., 2006. Grammatical processing in language learners. Appl. Psycholinguist. 27, 3–42.

Clopper, C.G., Bradlow, A.R., 2006. Effects of dialect variation on speech intelligibility in noise. J. Acoust. Soc. Amer. 119, 3424.

Cooke, M., Garcia Lecumberri, M.L., Barker, J.P., 2008. The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. J. Acoust. Soc. Amer. 123, 414–427.

Cooke, M., Garcia Lecumberri, M.L., Scharenborg, O., van Dommelen, W.A., 2010. Language-independent processing in speech perception: identification of English intervocalic consonants by speakers of eight European languages. Speech Communication 52, 954–967.

Cooke, M., 2006. A glimpsing model of speech perception in noise. J. Acoust. Soc. Amer. 119, 1562–1573.

Cooper, N., Cutler, A., Wales, R., 2002. Constraints of lexical stress on lexical access in English: evidence from native and nonnative listeners. Lang. Speech 45, 207–228.

Cutler, A., Weber, A., Smits, R., Cooper, N., 2004. Patterns of English phoneme confusions by native and non-native listeners. J. Acoust. Soc. Amer. 116, 3668–3678.

Cutler, A., Smits, R., Cooper, N., 2005. Vowel perception: effects of non-native language vs. non-native dialect. Speech Commun. 47, 32–42.

Cutler, A., Cooke, M., Garcia Lecumberri, M.L., Pasveer, D., 2007. L2 consonant identification in noise: cross-language comparisons. In: Proceedings of Interspeech 2007, pp. 1585–1588.

Cutler, A., Garcia Lecumberri, M.L., Cooke, M., 2008. Consonant identification in noise by native and non-native listeners: effects of local context. J. Acoust. Soc. Amer. 124, 1264–1268.

Cutler, A., Weber, A., Otake, T., 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. J. Phonetics 34, 269–284.

Darwin, C.J., 2008. Listening to speech in the presence of other sounds. Philos. Trans. Roy. Soc. B 363, 1011–1021.

Darwin, C.J., Brungart, D.S., Simpson, B.D., 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. J. Acoust. Soc. Amer. 114, 2913–2922.

Eckman, F.R., 1977. Markedness and the contractive analysis hypothesis. Lang. Learn. 27, 315–330.

Eisner, F., McQueen, J.M., 2005. The specificity of perceptual learning in speech processing. Percept. Psychophys. 67, 224–238.

Ellis, R., 1994. The Study of Second Language Acquisition. Oxford University Press, Oxford.

Ezzatian, P., Avivi, M., Schneider, B., 2010. Do non-native listeners benefit as much as native listeners from spatial cues that release speech from masking? Speech Communication 52, 919–929.

Fairbanks, G., 1958. Test of phonemic differentiation: the rhyme test. J. Acoust. Soc. Amer. 30, 596–600.

Festen, J.M., Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. J. Acoust. Soc. Amer. 88, 1725–1736.

Flege, J.E., Liu, S., 2001. The effect of experience on adults' acquisition of a second language. Stud. Second Lang. Acquisit. 23, 527–552.

Flege, J.E., Yeni-Komshian, G., Liu, S., 1999. Age constraints on second language learning. J. Memory Lang. 41, 78–104.
Flege, J.E., Bohn, O.S., Jang, S., 1997a. The effect of experience on nonnative subjects' production and perception of English vowels. J. Phonetics 25, 437–470.
Flege, J.E., Frieda, A., Nozawa, T., 1997b. Amount of native-language (L1) use affects the pronunciation of an L2. J. Phonetics 25, 169–186.
Flege, J.E., 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), Speech Perception and Language Experience. Issues in Cross-language Research. York, Baltimore, MD, pp. 233–277.
Flege, J.E., Munro, M.J., Mackay, I.R.A., 1995. Factors affecting strength of perceived foreign accent in a second language. J. Acoust. Soc. Amer. 97, 3125–3134.
Flege, J.E., 1987. A critical period for learning to pronounce foreign languages? Appl. Linguist. 8, 162–177.
Flege, J.E., 1988. The production and perception of foreign language speech sounds. In: Winitz, H. (Ed.), . In: Human Communication and Its Disorders. A Review, Vol. 1. Ablex Publishers, Norwood, NJ, pp. 224–401.
Flege, J.E., 1999. Age of learning and second-language speech. In: Birdsong, D. (Ed.), Second Language Acquisition and the Critical Period Hypothesis. Lawrence Erlbaum, Hillsdale, NJ, pp. 101–132.
Florentine, M., Buus, S., Scharf, B., Canevet, G., 1984. Speech reception thresholds in noise for native and non-native listeners. J. Acoust. Soc. Amer. 75, S84.
Florentine, M., 1985. Non-native listeners' perception of American-English in noise. In: Proceedings of InterNoise, Munich, pp. 1021–1024.
Fox, R.A., Flege, J.E., Munro, M.J., 1995. The perception of English and Spanish vowels by native English and Spanish listeners: a multidimensional scaling analysis. J. Acoust. Soc. Amer. 97, 2540–2552.
Frauenfelder, U.H., Floccia, C., 1998. The recognition of spoken words. In: Friederici, A. (Ed.), Language Comprehension: A Biological Perspective. Springer, Berlin, pp. 1–40.
Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Amer. 115, 2246–2256.
Gat, I.B., Keith, R.W., 1978. An effect of linguistic experience: auditory word discrimination by native and non-native speakers of English. Audiology 17, 339–345.
Garcia Lecumberri, M.L., Cooke, M., 2006. Effect of masker type on native and non-native consonant perception in noise. J. Acoust. Soc. Amer. 119, 2445–2454.
Garland, S., 2007. The Bilingual Spectrum. Guirnalda Publishing, Orlando, FL.
Garnier, M., 2007. Communiquer en environement bruyant: de l'adaptation jusqu'au forcage vocal. These de Doctorat de l'Universite Paris 6.
Gaskell, M., Marslen-Wilson, W.D., 1997. Integrating form and meaning: a distributed model of speech perception. Lang. Cognit. Process. 12, 613–656.
Golestani, N., Rosen, S., Scott, S.K., 2009. Native-language benefit for understanding speech-in-noise: the contribution of semantics. Bilingualism: Lang. Cognit. 12, 385–392.
Gooskens, C., van Heuven, V.J., van Bezooijen, R., Pacilly, J.J., 2010. Is spoken Danish less intelligible than Swedish? Speech Communication 52, 1022–1037.
Grosjean, F., 1998. Studying bilinguals: methodological and conceptual issues. Lang. Cognit. 1, 131–149.
Grosjean, F., 2001. The bilingual's language modes. In: Nicol, J. (Ed.), One Mind, Two Languages: Bilingual Language Processing. Blackwell, Oxford, pp. 1–22.
Grosjean, F., 2010. Bilingual: Life and Reality. Harvard University Press, Cambridge, Massachusetts.
Guion, S., Flege, J.E., Loftin, J., 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. J. Phonetics 28, 27–42.
Hardison, D.M., 1996. Bimodal speech perception by native and non-native speakers of English: factors influencing the McGurk effect. Lang. Learn. 46, 3–73.
Harley, B., Howard, J., Hart, D., 1995. Second language processing at different ages: do younger learners pay more attention to prosodic cues to sentence structure. Lang. Learn. 45, 43–71.
Hazan, V., Kim, J., Chen, Y., 2010. Audiovisual perception in adverse conditions: language, speaker and listener effects. Speech Communication 52, 996–1009.
Hazan, V., Simpson, A., 2000. The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. Lang. Speech 43, 273–294.
Hazan, V., Simpson, A., 1998. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. Speech Commun. 24, 211–226.
Heinrich, A., Hawkins, S., 2010. Influence of English *r*-resonances on intelligibility of speech in noise for native English and German listeners. Speech Communication 52, 1038–1055.
Hoen, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., 2007. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun. 12, 905–916.
Howes, D., 1957. On the relationship between intelligibility and frequency of occurrence of English words. J. Acoust. Soc. Amer. 29, 296–305.
Imai, S., Walley, A., Flege, J.E., 2005. Lexical frequency and neighborhood density effects on the recognition of native and Spanish accented words by native English and Spanish listeners. J. Acoust. Soc. Amer. 117, 896–907.
Ioup, G., 1984. Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. Lang. Learn. 34, 1–17.
Iverson, P., Kuhl, P.K., 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. J. Acoust. Soc. Amer. 99, 553–562.
Jiang, J., Chen, M., Alwan, A., 2006. On the perception of voicing in syllable initial plosives in noise. J. Acoust. Soc. Amer. 119, 1092–1105.
Jones, C., Berry, L., Stevens, C., 2007. Synthesized speech intelligibility and persuasion: speech rate and non-native listeners. Comput. Speech Lang. 21, 641–651.
Kreul, E.J., Nixon, J.C., Kryter, K.D., Bell, D.W., Lang, J.S., Schubert, E.D., 1968. A proposed clinical test of speech discrimination. J. Speech Hear. Res. 11, 536–552.
Kuhl, P.K., 1993a. An examination of the perceptual magnet effect. J. Acoust. Soc. Amer. 93, 2423, 2423.
Kuhl, P.K., 1993b. Early linguistic experience and phonetic perception: implications for theories of developmental speech production. J. Phonetics 21, 125–139.
Kuhl, P.K., 2000. A new view of language acquisition. Proc. Nat. Acad. Sci. USA 97, 11850–11857.
Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. Nat. Rev. Neurosci. 5, 831–843.
Lado, R., 1957. Linguistics across Cultures. University of Michigan Press, Ann Arbor.
Lane, H., Tranel, B., 1971. The Lombard sign and the role of hearing in speech. J. Speech Hear. Res. 14, 677–709.
Leather, J., James, A., 1991. The acquisition of second language speech. Stud. Second Lang. Acquisit. 13, 305–341.
Lee, C.-Y., Tao, L., Bond, Z.S., 2010. Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. Speech Communication 52, 900–910.
Lenneberg, E., 1967. Biological Foundations of Language. Wiley, New York.
Lively, S.E., Logan, J.S., Pisoni, D.B., 1993. Training Japanese listeners to identify English /r/ and /l/. The role of phonetic environment and talker variability in learning new perceptual categories. J. Acoust. Soc. Amer. 94, 1242–1255.
Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., Yamada, T., 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-

term retention of new phonetic categories. J. Acoust. Soc. Amer. 96, 2076–2087.
Logan, J.S., Lively, S.E., Pisoni, D.B., 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. J. Acoust. Soc. Amer. 89, 874–886.
Lombard, E., 1911. Le signe de l'elevation de la voix. Ann. Maladiers Oreille, Larynx, Nez, Pharynx 37, 101–119.
Long, M., 1990. Maturational constraints on language development. Stud. Second Lang. Acquisit. 12, 251–285.
Lovitt, A., Allen, J.B., 2006. 50 years late: repeating Miller-Nicely 1955. In: Proceedings of Interspeech 2006, Pittsburgh, pp. 2154–2157.
Lu, Y., 2010. Production and Perceptual Analysis of Speech Produced in Noise. Ph.D. Thesis, University of Sheffield.
Luce, P.A., Pisoni, D.B., 1998. Recognizing spoken words: the neighborhood activation model. Ear Hearing 19, 1–36.
Mackay, I.R.A., Meador, D., Flege, J.E., 2001a. The identification of English consonants by native speakers of Italian. Phonetica 58, 103–125.
Mackay, I.R.A., Flege, J.E., Piske, T., Schirru, C., 2001b. Category restructuring during second-language speech acquisition. J. Acoust. Soc. Amer. 110, 516–528.
Macnamara, J., 1969. How can one measure the extent of a person's bilingual proficiency? In: Kelly, L.G. (Ed.), Description and Measurement of Bilingualism. University of Toronto Press, Toronto, pp. 80–119.
Maddieson, I., 1984. Patterns of Sounds. Cambridge University Press, Cambridge.
Major, R.C., 2001. Foreign Accent: The Ontogeny and Philogeny of Second Language Phonology. Lawrence Erlbaum, New Jersey.
Major, R.C., 1998. Interlanguage phonetics and phonology: an introduction. Stud. Second Lang. Acquisit. 20, 131–137.
Major, R.C., Kim, E., 1999. The similarity differential rate hypothesis. In: Leather, J.H. (Ed.), Phonological Issues in Language Learning. Blackwell, Malden, MA, pp. 151–184.
Markham, D., 1997. Phonetic Imitation, Accent and the Learner. Lund University Press.
Marslen-Wilson, W., Warren, P., 1994. Levels of perceptual representation and process in lexical access: words, phonemes, and features. Psychol. Rev. 101, 653–675.
Mattys, S.L., Carroll, L.M., Li, C.K., Chan, S.L., 2010. Effects of energetic and informational masking on speech segmentation by native and non-native speakers. Speech Communication 52, 887–899.
Mattys, S.L., Brooks, J., Cooke, M., 2009. Recognising speech under a processing load: dissociating energetic from informational factors. Cognit. Psychol. 59, 203–243.
Mayo, L., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech Lang. Hearing Res. 40, 686–693.
McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cognit. Psychol. 18, 1–86.
McQueen, J.M., 2007. Eight questions about spoken-word recognition. In: Gaskell, M.G. (Ed.), The Oxford Handbook of Psycholinguistics. Oxford University Press, Oxford, pp. 37–53.
McQueen, J.M., Norris, D., Cutler, A., 1999. Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. J. Exp. Psychol.: Human Percept. Perform. 25, 1363–1389.
McQueen, J.M., Cutler, A., Norris, D., 2006. Phonological abstraction in the mental lexicon. Cognit. Sci. 30, 1113–1126.
McQueen, J.M., Huettig, F., submitted for publication. Changing the probability of radio interference changes how spoken words are recognized. Speech Communication.
Meador, D., Flege, J.E., Mackay, I.R.A., 2000. Factors affecting the recognition of words in second language. Bilingualism: Lang. Cognit. 3, 55–67.
Miller, G.A., 1947. The masking of speech. Psychol. Bull. 44, 105–129.
Miller, G.A., Nicely, P., 1955. An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer. 27, 338–352.
Mullennix, J.W., Pisoni, D.B., Martin, C.S., 1989. Some effects of talker variability on spoken word recognition. J. Acoust. Soc. Amer. 85, 365–378.
Munro, M.J., 1998. The effects of noise on the intelligibility of foreign-accented speech. Stud. Second Lang. Acquisit. 20, 139–154.
Nabelek, A.K., 1988. Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. J. Acoust. Soc. Amer. 84, 476–484.
Nabelek, A.K., Donahue, A.M., 1984. Perception of consonants in reverberation by native and non-native listeners. J. Acoust. Soc. Amer. 75, 632–634.
Nabelek, A.K., Robinson, P.K., 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. J. Acoust. Soc. Amer. 71, 1242–1248.
Nabelek, A.K., Pickett, J.M., 1974. Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. J. Speech Hearing Res. 17, 724–739.
Nelson, P., Kohnert, K., Sabur, S., Shaw, D., 2005. Classroom noise and children learning through a second language: double jeopardy? Language, Speech Hearing Services Schools 36, 219–229.
Neuman, A.C., Hochberg, I., 1983. Children's perception of speech in reverberation. J. Acoust. Soc. Amer. 73, 2145–2149.
Norris, D., Cutler, A., McQueen, J.M., Butterfield, S., 2006. Phonological and conceptual activation in speech comprehension. Cognit. Psychol. 53, 146–193.
Norris, D., McQueen, J.M., 2008. Shortlist B: a Bayesian model of continuous speech recognition. Psychol. Rev. 115, 357–395.
Norris, D., McQueen, J.M., Cutler, A., 2003. Perceptual learning in speech. Cognit. Psychol. 47, 204–238.
Norris, D., 1994. Shortlist: a connectionist model of continuous speech recognition. Cognition 52, 189–234.
Nygaard, L.C., Sommers, M.S., Pisoni, D.B., 1994. Speech perception as a talker-contingent process. Psychol. Sci. 5, 42–46.
Penfield, W., Roberts, L., 1959. Speech Brain Mechanisms. Princeton University Press, Princeton, NJ.
Picard, M., Bradley, J.S., 2001. Revisiting speech interference in classrooms. Audiology 40, 221–244.
Picheny, M.A., 1981. Speaking Clearly for the Hard of Hearing. Ph.D. Thesis, MIT.
Pickering, M.J., Garrod, S., 2004. Towards a mechanistic psychology of dialogue. Behav. Brain Sci. 27, 169–226.
Piske, T., Mackay, I.R.A., 1999. Age and L1 use effects on degree of foreign accent in English. In: Ohala, J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A. (Eds.), Proceedings of the Fourteenth International Congress of Phonetic Sciences. Department of Linguistics, San Francisco/Berkeley, CA, pp. 1433–1436.
Polivanov, E., 1932. La perception des sons d'une langue étrangère. Travaux du Cercle Linguistique de Prague 4, 79–96.
Pollack, I., Rubenstein, H., Decker, L., 1959. Intelligibility of known and unknown message sets. J. Acoust. Soc. Amer. 31, 273–279.
Quene, H., van Delft, L.E., 2010. Non-native durational patterns decrease speech intelligibility. Speech Communication 52, 911–918.
Repp, B.H., Frost, R., 1988. Detectability of words and nonwords in two kinds of noise. J. Acoust. Soc. Amer. 84, 1929–1932.
Rhebergen, K.S., Versfeld, N.J., Dreschler, W.A., 2005. Release from informational masking by time reversal of native and non-native interfering speech. J. Acoust. Soc. Amer. 118, 1274–1277.
Rogers, C., Lister, J., Febo, D., Besing, J., Abrams, H., 2006. Effects of bilingualism, noise and reverberation on speech perception by listeners with normal hearing. Appl. Psycholinguist. 27, 465–485.
Rogers, C.L., Dalby, J., Nishi, K., 2004. Effects of noise and proficiency on intelligibility of Chinese-accented English. Lang. Speech 47, 139–154.
Samuel, A.G., 1981. Phonemic restoration: insights from a new methodology. J. Exp. Psychol. Gen. 110, 474–494.
Savin, H., 1963. Word frequency effect and errors in the perception of speech. J. Acoust. Soc. Amer. 35, 200–206.

Schegloff, E.A., 2000. Overlapping talk and the organization of turn-taking for conversation. Lang. Soc. 29, 1–63.
Scovel, T., 1969. Foreign accents, language acquisition and cerebral dominance. Lang. Learn. 19, 245–253.
Scovel, T., 1988. A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech. Newbury House, New York.
Seliger, H., 1978. Implications of a multiple critical periods hypothesis for second language learning. In: Ritchie, W. (Ed.), Second Language Acquisition Research. Academic Press, New York, pp. 11–20.
Shimizu, T., Makishima, K., Yoshida, M., Yamagishi, H., 2002. Effect of background noise on perception of English speech for Japanese listeners. Auris Nasus Larynx 29, 121–125.
Simpson, S., Cooke, M., 2005. Consonant identification in *N*-talker babble is a nonmonotonic function of *N*. J. Acoust. Soc. Amer. 118, 2775–2778.
Singh, S., 1966. Cross-language study of perceptual confusion of plosive phonemes in two conditions of distortion. J. Acoust. Soc. Amer. 40, 635–656.
Singleton, D., 1989. Language Acquisition: The Age Factor. Multilingual Matters Ltd., Clevedon.
Sjerps, M.J., McQueen, J.M., 2010. The bounds on flexibility in speech perception. J. Exp. Psychol.: Human Percept. Perform. 36, 195–211.
Slowiaczek, L.M., Nusbaum, H.C., Pisoni, D.B., 1987. Phonological priming in auditory word recognition. J. Exp. Psychol.: Learn. Memory Cognit. 13, 64–75.
Sorace, A., 1993. Incomplete vs. divergent representations of unaccusativity in near-native grammars of Italian. Second Lang. Res. 9, 22–48.
Soto-Faraco, S., Sebastián-Gallés, N., Cutler, A., 2001. Segmental and suprasegmental mismatch in lexical access. J. Memory Lang. 45, 412–432.
Spivey, M., Marian, V., 1999. Cross talk between native and second languages: partial activation of an irrelevant lexicon. Psychol. Sci. 10, 281–284.
Spolsky, B., Sigurd, B., Sato, M., Walker, E., Arterburn, C., 1968. Preliminary studies in the development of testing techniques for testing overall second language proficiency. Lang. Learn. 18, 79–101.
Steeneken, H.J.M., Houtgast, T., 1980. A physical method for measuring speech-transmission quality. J. Acoust. Soc. Amer. 67, 318–326.
Stockwell, R.P., Bowen, J.D., 1965. The Sounds of English and Spanish. University of Chicago Press, Chicago.
Strange, W. (Ed.), 1995. Speech Perception and Linguistic Experience: Issues in Cross-Language Research. York Press, Timonium, MD, pp. 3–45.
Taft, M., 1986. Lexical access codes in visual and auditory word recognition. Lang. Cognit. Process. 1, 297–308.
Takata, Y., Nabelek, A., 1990. English consonant recognition in noise and in reverberation by Japanese and American listeners. J. Acoust. Soc. Amer. 88, 663–666.
Tremblay, A., 2008. Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. Appl. Psycholinguist. 29, 553–584.
Trubetzkoy, N., 1939. Principles of Phonology (Grundzüge der Phonologie). University of California Press (Translated by C.A.M. Baltaxe, Berkeley, 1949.
van der Vlugt, M., Nooteboom, S., 1986. Auditory Word Recognition is Not More Sensitive to Word-initial than to Word-final Stimulus Information. IPO Annual Progress Report 21, pp. 41–49.
van Wijngaarden, S.J., Steeneken, H., Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native listeners. J. Acoust. Soc. Amer. 111, 1906–1916.
van Wijngaarden, S.J., Bronkhorst, A.W., Houtgast, T., Steeneken, H.J.M., 2004. Using the speech transmission index for predicting non-native speech intelligibility. J. Acoust. Soc. Amer. 115, 1281–1291.
Vanlancker-Sidtis, D., 2003. Auditory recognition of idioms by first and second speakers of English. Appl. Psycholinguist. 24, 45–57.
Van Dommelen, W.A., Hazan, V., 2010. Perception of English consonants in noise by native and Norwegian listeners. Speech Commun. 52, 968–979.
Van Engen, K.J., 2010. Similarity and familiarity: second language sentence recognition in first- and second-language multi-talker babble. Speech Commun. 52, 943–953.
Van Engen, K.J., Bradlow, A.R., 2007. Sentence recognition in native- and foreign-language multi-talker background noise. J. Acoust. Soc. Amer. 121, 519–526.
Van Summers, W., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., Stokes, M.A., 1988. Effects of noise on speech production: acoustic and perceptual analyses. J. Acoust. Soc. Amer. 84, 917–928.
Volin, J., Skarnitzl, R., 2010. The strength of foreign accent in Czech English under adverse listening conditions. Speech Commun. 52, 1010–1021.
von Hapsburg, D., Champlin, C.A., Shetty, S.R., 2004. Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. J. Amer. Acad. Audiol. 15, 88–98.
von Hapsburg, D., Peña, E.D., 2002. Understanding bilingualism and its impact on speech audiometry. J. Speech Lang. Hearing Res. 45, 202–213.
Walsh, T., Diller, K., 1981. Neurolinguistic considerations on the optimum age for second language learning. In: Diller, K. (Ed.), Individual Diff. Universals Lang. Learn. Aptitude. Newbury House, Rowley, MA, pp. 3–21.
Wang, Y., Behne, D.M., Jiang, H., 2008. Influence of native language phonetic system on audio-visual speech perception. J. Phonetics 37, 344–356.
Wang, M.D., Bilger, R.C., 1973. Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Amer. 54, 1248–1266.
Watkins, A.J., 2005. Perceptual compensation for effects of echo and of reverberation on speech identification. Acta Acustica United with Acustica 91, 892–901.
Warren, R.M., 1970. Perceptual restoration of missing speech sounds. Science 167, 392–393.
Weber, A., Cutler, A., 2004. Lexical competition in non-native spoken-word recognition. J. Memory Lang. 50, 1–25.
Weiss, W., Dempsey, J.J., 2008. Performance of bilingual speakers on the English and Spanish versions of the hearing in noise test (HINT). J. Amer. Acad. Audiol. 19, 5–17.
Wode, H., 1980. Phonology in L2 acquisition. In: Felix, S. (Ed.), Second Language Development: Trends and Issues. Gunter Narr, Tübingen.
Wright, R., 2004. A review of perceptual cues and cue robustness. In: Hayes, B., Kirchner, R., Steriade, D. (Eds.), Phonetically-based Phonology. Cambridge University Press.
Zwitserlood, P., 1989. The locus of the effects of sentential-semantic context in spoken-word processing. Cognition 32, 25–64.