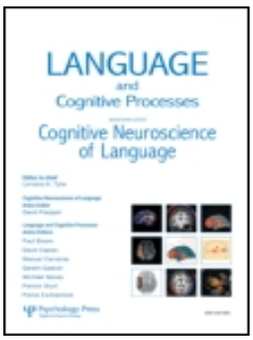

Routledge
Taylor & Francis Group

Language and Cognitive Processes

ISSN: 0169-0965 (Print) 1464-0732 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp20

Speech recognition in adverse conditions: A review

Sven L. Mattys , Matthew H. Davis , Ann R. Bradlow & Sophie K. Scott

To cite this article: Sven L. Mattys , Matthew H. Davis , Ann R. Bradlow & Sophie K. Scott (2012) Speech recognition in adverse conditions: A review, Language and Cognitive Processes, 27:7-8, 953-978, DOI: 10.1080/01690965.2012.705006

To link to this article: https://doi.org/10.1080/01690965.2012.705006

Published online: 12 Jul 2012.

LANGUAGE AND COGNITIVE PROCESSES
2012, 27 (7/8), 953–978

Psychology Press
Taylor & Francis Group

Speech recognition in adverse conditions: A review

Sven L. Mattys[1], Matthew H. Davis[2], Ann R. Bradlow[3], and Sophie K. Scott[4]

[1]Department of Psychology, University of York, York, UK
[2]Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, UK
[3]Department of Linguistics, Northwestern University, Evanston, IL, USA
[4]Institute of Cognitive Neuroscience, University College London, London, UK

This article presents a review of the effects of adverse conditions (ACs) on the perceptual, linguistic, cognitive, and neurophysiological mechanisms underlying speech recognition. The review starts with a classification of ACs based on their origin: Degradation at the source (production of a noncanonical signal), degradation during signal transmission (interfering signal or medium-induced impoverishment of the target signal), receiver limitations (peripheral, linguistic, cognitive). This is followed by a parallel, yet orthogonal classification of ACs based on the locus of their effect: Perceptual processes, mental representations, attention, and memory functions. We then review the added value that ACs provide for theories of speech recognition, with a focus on fundamental themes in psycholinguistics: Content and format of lexical representations, time-course of lexical access, word segmentation, feed-back in speech perception and recognition, lexical-semantic integration, interface between the speech system and general cognition, neuroanatomical organisation of speech processing. We conclude by advocating an approach to speech recognition that includes rather than neutralises complex listening environments and individual differences.

Keywords: Speech recognition; Lexical access; Adverse conditions; Signal degradation; Cognitive load; Masking.

In everyday life, listeners recognise speech under a wide range of suboptimal or adverse conditions. Here, we define an adverse condition (AC) as any factor leading to a decrease in speech intelligibility on a given task relative to the level of intelligibility when the same task is performed in optimal listening situations, i.e., healthy native listeners hearing carefully recorded speech in a quiet environment and under focused attention. Speech recognition in ACs has been a familiar area of

Correspondence should be addressed to Sven L. Mattys, Department of Psychology, University of York, Heslington, York YO10 5DD, UK. E-mail: sven.mattys@bris.ac.uk

This article and the organisation of this special issue were made possible thanks to support from: The Leverhulme Trust (F/00 182/BG), the ESRC (RES-062-23-2746), the Marie Curie Foundation (MRTN-CT-2006-035561), the Experimental Psychology Society (EPS), the Wellcome Trust (WT090961AIA), and the MRC (MHD: MC_US_A060_0038).

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2012.705006

research in computer science, engineering, and hearing sciences for several decades (e.g., Juang, 1991; Junqua & Haton, 1995). In contrast, most psycholinguistic theories of speech recognition are built on evidence gathered in optimal situations that can only be created in laboratory-controlled conditions. A problem with these conditions of "artificial normality" (Mattys & Liss, 2008) is that they often fail to capture the range of processes involved in everyday speech recognition, miscalculating the extent of compensatory mechanisms, overlooking the contribution of short-term memory and attention and, more generally, underestimating the flexibility of the speech recogniser and the degree to which it interfaces with higher-level cognition. The goal of this review is to provide a formal classification of ACs relative to the optimal benchmark described above and, in doing so, highlight the research opportunities, present and future, that these ACs offer for theories of speech recognition.

CLASSIFICATION OF ADVERSE CONDITIONS

We propose a classification of ACs based on their origin and, independently, on their effect (see Assmann & Summerfield, 2004, for a complementary classification system).[1] Origin refers to the locus or cause of the disruption, whether it is external to the listener (e.g., a speaker's atypical pronunciation, background noise) or internal (e.g., non-native linguistic knowledge, cochlear implant, multi-tasking). Effects refer to the types of perceptual processes, mental representations, linguistic functions, and cognitive mechanisms affected by the disruption, and the compensatory behaviour that this disruption might elicit. A summary of this classification system is displayed in Table 1.

CLASSIFICATION OF ADVERSE CONDITIONS BASED ON THEIR ORIGIN

Source degradation

This category includes any intrinsic variation of the speech signal leading to reduced intelligibility compared to speech carefully produced by healthy native speakers. By "intrinsic", we mean aspects of the speech as produced, rather than degradation due to limitations of the communication channel (e.g., noise, competing talker). The latter kind of degradation is reviewed in the Environmental/transmission degradation section.

Conversational speech

Although the features of conversational speech in relation to hyper-articulated (clear) speech are beyond the scope of this special issue (see Smiljanic & Bradlow, 2009 and Uchanski, 2005, for reviews of clear and conversational speech), spontaneously produced, conversational speech can be treated as an AC insofar as it reduces intelligibility relative to citation form, read speech. Features likely to lead to reduced intelligibility are syllable deletion, segment elision, and segment reduction (e.g.,

[1] Our review of ACs in speech recognition does not directly address short-term cues to segment-level features. For some data and discussion of cue re-weighting for segment identification in noise, see, e.g., Parikh and Loizou (2005), Jiang, Chen, and Alwan (2006); see also Assmann and Summerfield (2004) for a review).

TABLE 1

Summary of the origins and effects of ACs. The shade of each cell indicates our estimation of the approximate frequency of occurrence or importance of each origin-effect combination (light grey: rare/mild; dark grey: common/moderate; black: frequent/severe). The ordering of rows and columns is based on their summed ''severity'' (we arbitrarily used the following quantification: light grey = 1; dark grey = 2; black = 3). Thus, the most severe ACs are clustered in the top left portion of the table. The intention of this table, and this review in general, is to spark new research on ACs with this framework in mind. Specifically, our hope is that this framework will help guide future research towards asking and answering questions such as: Are there any ACs that have disparate sources yet similar effects and are, therefore, both candidates for similar amelioration strategies?

	Adverse condition effect				
Adverse condition origin	*Failure of recognition*	*Reduced attentional capacity*	*Reduced memory capacity*	*Perceptual learning*	*Perceptual interference*
Environment/transmission degradation with EM					
Receiver limitation impaired language model					
Source degradation speech disorders					
Source degradation accented speech					
Environment/transmission degradation without EM					
Receiver limitation cognitive load					
Receiver limitation peripheral deficiency					
Receiver limitation incomplete language model					
Source degradation conversational speech					
Source degradation disfluencies					

Ernestus, Baayen, & Schreuder, 2002; Mitterer, 2006; Picheny, Durlach, & Braida, 1985, 1986) and possibly faster speech rate (though this is a disputed fact, e.g., Bradlow, Torretta, & Pisoni, 1996; Kraus & Braida, 2002; Picheny, Durlach, & Braida, 1989).

Accented speech

Defined by Munro and Derwing (1995) as "non-pathological speech that differs in some noticeable respects from native speaker pronunciation norms" (p. 298), foreign-accented speech affects both segmental and suprasegmental aspects of the signal, and it can result in increased processing effort, segmental/lexical ambiguity, and mapping failure (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992). Unfamiliar native accents can present similar challenges (Adank, Evans, Stuart-Smith, & Scott, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006). An interesting feature of accented speech is the relative consistency of the speaker's productions compared to accidental mispronunciations. Such consistency allows listeners to recalibrate their phonemic and/or prosodic categories within the course of a conversation through perceptual

learning (e.g., Bradlow & Bent, 2008; Kraljic, Brennan, & Samuel, 2008; Maye, Aslin, & Tanenhaus, 2008; Sidaras, Alexander, & Nygaard, 2009).

Disfluencies

These include repairs, restarts, and fillers interrupting the flow of otherwise fluent speech. Disfluencies affect segment duration, intonation, voice quality, and coarticulation patterns (Shriberg, 1994). Although disfluencies can sometimes carry meaningful information (e.g., Brennan & Schober, 2001; Clark & Fox Tree, 2002), they tend to impair recognition performance (e.g., Levelt, 1989).

Speech disorders

Departures from typical speech vary widely across disorders. Neurogenic disorders, for example, dysarthria and apraxia of speech, result in a constellation of speech distortions, including problems with rhythm, rate intensity, voice quality, formant structure, coarticulation, etc. (Darley, Aronson, & Brown, 1969; Kent, Weismer, Kent, & Rosenbek, 1989). In contrast, structural disorders, which affect the anatomical structure of the vocal tract (e.g., cleft palate, bifid uvula), can have very highly feature-specific manifestations, such as increased nasal resonance, incorrect place of articulation, and hyper-glottalisation (e.g., Harding & Grunwell, 1996). Speech production disorders associated with hearing impairments, i.e., "deaf speech", often lead to more centralised, and hence, less discriminable vowels, as well as intonational irregularities and a wide range of misarticulations (Osberger & McGarr, 1982).

Environmental/transmission degradation

This category is conceptually independent of the previous one in that, here, the ACs originate in imperfections in the communication channel between the speaker and the listener. The degradation can be due to factors as obvious as the distance between the speaker and the listener, which will broadly result in better transmission of high- than low-intensity speech components. However, it will more commonly result from competing signals in the environment (e.g., noise, background babble) or from acoustic distortions caused by the physical environment (reverberation) or the channel (e.g., filtering of speech on a telephone). In the competing-signal category, an important distinction must be made between competing signals leading to energetic masking and competing signals causing distortion without energetic masking.

Degradation with energetic masking

Energetic masking occurs when the intelligibility of a target is reduced by a distractor due to a physical overlap, or super-imposition between the target signal and a nontarget signal, such as noise or background talkers (see a review in Brungart, 2001). Under energetic-masking conditions, signal separation (or "stream segregation", Bregman, 1990) and selective attention become central to the recognition process (Darwin, 2008). If a masking signal has a fluctuating amplitude envelope, "glimpses" of the target through lower-intensity parts of the distractor can aid recognition (Cooke, 2006; Festen & Plomp, 1990). With energetically constant distractors (e.g., multi-talker babble, steady-state noise), however, temporal glimpses are rarer but signal separation based on spectral contrast, common onset, or harmonicity cues may be possible. Note that a nontarget signal with intelligible and meaningful content (e.g., a competing talker in a native rather than non-native language) can additionally result in so-called informational

masking. Informational masking can be described as the consequence of the nontarget signal once its energetic effect has been accounted for (Cooke, Garcia Lecumberri, & Barker, 2008). Thus, it typically refers to the higher-level, postperiphery consequence of masking, e.g., attentional capture by the masker, semantic interference, and associated cognitive load (Cooke et al., 2008; Mattys, Brooks, & Cooke, 2009; for a review, see Kidd, Mason, Richards, Gallun, & Durlach, 2007). We return to the consequences of informational masking in a later section.

Degradation without energetic masking

Signal degradation can also occur in the absence of a separate distractor. For instance, telephone transmission typically filters out frequencies below 400 Hz and above 3,400 Hz (Nilsson & Kleijn, 2001), while the bulk of the information-carrying frequency range for human speech is between 100 and 5000 Hz (Borden, Harris, & Raphael, 2003). Reverberation also degrades the target signal without creating a extraneous signal in need of separation—at least at short reverberation delays—and has the effect of reducing intelligibility of late segments more than early segments, and of preserving vowel identity better than consonant identity (Helfer, 1994; Nábelek, 1988). Unlike degradation with energetic masking, degradation without energetic masking does not require sound separation or selective attention. Rather, its manifestations and consequences are often similar to ACs created by source degradation.

An important factor to consider when studying the extent of environmental/transmission degradation is whether conversational partners are aware of the degradation and, if so, whether they attempt to compensate for it when speaking. For instance, interlocutors over the phone rarely attempt to compensate for the impoverished signal, whereas speakers in noisy environments often try to adjust their articulation, adopting what is generally called a Lombard-speech style (Lombard, 1911; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988), which is distinct from the shouted speech style typically adopted by interlocutors separated by distance (Rostolland, 1982, 1985). Similarly, when addressing a listener with a hearing impairment or a non-native listener, talkers commonly adopt a "clear" speaking style that is intended to enhance intelligibility (e.g., Ferguson & Kewley-Port, 2002; see also Uchanski, 2005, and Smiljanic & Bradlow, 2009, for reviews). These adjustments involve multiple acoustic-phonetic dimensions and the overall benefit for intelligibility is quite robust across talkers and listeners from various populations (e.g., children, young and older adults, native and non-native speakers). Yet, the extent to which each individual acoustic-phonetic modification is responsible for overall enhanced intelligibility and whether talkers can modulate their clear speech strategies depending on the particular environmental/transmission degradation remain to be determined. However, data by Hazan and Baker (2011) suggest that talkers can tailor their clear-speech production strategies to the particular AC by, for example, making greater changes in F0 and mean energy in a noisy listening condition (background of 8-talker babble) than in a simulated cochlear-implant listening condition, where changes in speaking rate and vowel space are more prominent.

Receiver limitations

Adverse conditions arising from limitations in the perceptual or cognitive abilities of the listener fall into four categories.

Peripheral deficiency

This category primarily includes sensorineural hearing impairments. The magnitude and nature of the intelligibility decrement varies widely as a function of the type and severity of the impairment, the type of remediation in place (none, hearing aids, cochlear implants), and the presence of other ACs such as background noise or reverberation. However, these peripheral deficiencies are distinct from source and environmental degradations in that the degradation experienced by hearing-impaired individuals is relatively constant and applies to all auditory signals. The highly redundant nature of the speech signal can, therefore, be exploited to its fullest extent to compensate for sensory impoverishment; it is this redundancy that permits surprisingly good comprehension of speech even with only a limited amount of spectral detail (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Wilson et al., 1991).

Incomplete language model

Independent of peripheral limitations, ACs can arise from an incomplete knowledge of the language—phonological, lexical, morpho-syntactic, grammatical, idiomatic, etc. Leaving aside the case of developing children, this category is best represented by non-native listeners (see Garcia Lecumberri, Cooke, & Cutler, 2010, for a review). The detrimental effect of an incomplete language model is exacerbated when combined with source or environmental degradation. For example, non-native listeners are more affected by a noisy environment than are native listeners (e.g., Mayo, Florentine, & Buss, 1997; Nábelek & Donahue, 1984; Rogers, Lister, Febor, Besing, & Abrams, 2006), with the difference attributed to the accumulation of limitations in non-native representations across segmental, lexical, syntactic, and higher levels of processing (Bradlow & Alexander, 2007; Cutler, Garcia Lecumberri, & Cooke, 2008; Cutler, Webber, Smits, & Cooper, 2004).

Impaired access or use of the language model

This category includes mainly acquired neurological deficits of language functions (e.g., auditory agnosia, cortical deafness, pure word deafness, Wernicke's aphasia). Speech recognition deficits subsequent to brain injury vary widely depending on the site and severity of the injury. Acoustic-phonetic discrimination, phonemic categorisation, prosodic judgment, lexical activation, lexical inhibition, short-term memory, semantic retrieval, and syntactic parsing can be selectively affected or, alternatively (and more typically), lead to cascaded disruptions at multiple levels of the speech system (see Badecker, 2005, and Blumstein, 2007, for reviews). While a discussion of the impact of these neurological impairments on speech recognition is beyond the scope of this article, it is worth noting that the listener's *awareness* of this type of AC—and of its consequences—is generally far lower than that of the other types of ACs (e.g., Prigatano & Schacter, 1991). We also note with interest that perceptual degradation of speech has been proposed as a means of simulating these impairments in healthy listeners (Dick et al., 2001). This might suggest commonalities between behavioural profiles induced by source or environmental degradation and receiver limitations.

Cognitive load

We define cognitive load as any factor placing unusually high demands on central attentional or mnemonic capacities (Mattys & Wiget, 2011). Although cognitive load

often originates in the environment (e.g., visual distraction) or arises as a secondary consequence of another AC (e.g., signal degradation, accented speech, Rönnberg, Rudner, Lunner, & Zekveld, 2010), we classify cognitive load as a receiver limitation insofar as it involves mental activities that are unrelated to the speech stimuli or the listening task and insofar as the control of those mental activities is largely a function of the listener's individual attentional and memory resources. In the context of this review, we are particularly interested in cognitive load caused by an independent task competing with speech processing for limited processing resources (Kahneman, 1973), although we note that cognitive load can also arise from linguistic complexity per se, e.g., complex syntactic or discourse structure (e.g., Lewis, Vasishth, & Van Dyke, 2006).

CLASSIFICATION OF ADVERSE CONDITIONS BASED ON THEIR EFFECT

As the previous section suggests, ACs originating in different sources can impact on the same processes in speech reception. For instance, listening to an unfamiliar accent (a source degradation), listening to reverberant speech (an environmental degradation), or listening to speech through a cochlear implant (a receiver limitation) all lead to some degree of mismatch between the perceived segments and their canonical forms. The fact that very different ACs can have a common perceptual effect also suggests that similar compensatory mechanisms might be at work. We, therefore, can classify ACs in terms of the impact that they have on the comprehension system, and the cognitive (or neural) compensatory mechanisms that they elicit. Although this classification is by definition somewhat artificial, it maps onto traditionally distinct research areas and offers concrete ways of thinking about and experimenting with ACs.

Failure of recognition

Failure to map acoustic-phonetic features to segmental representations and, in turn, segmental to lexical representations is a common consequence of many ACs. It can arise from information loss in the time domain (e.g., intermittent transmission), the intensity domain (e.g., distance), and the spectral domain (e.g., telephone transmission, cochlear implants). It can also be due to acoustic-phonetic deviations from expectations, as with accented speech, disordered speech, mispronunciations, conversational speech, etc. The outcome in all these cases ranges from lexical uncertainty (often simulated as diffuse lexical activation or unresolved lexical competition), to erroneous lexical selection (recognising a word that was not presented), to no lexical selection at all.

How the speech system copes with loss of information depends on the nature and the quantity of the loss. Aside from catastrophic segmental/lexical mapping failure, listeners are usually good at bridging short gaps in speech thanks to temporal redundancy in the signal, e.g., coarticulation, and echoic memory (Huggins, 1975; Miller & Licklider, 1950). Information is also more easily recovered when it consists of permissible segment deletions in richly contextualised conditions, as commonly seen in conversational speech (e.g., Ernestus et al., 2002).

With respect to acoustic-phonetic deviations, diffuse lexical activation is likely to lead to broader competition and less robust lexical selection. The impact of this on comprehension will depend on lexical frequency and lexical confusability (cf. neighbourhood density, e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1999), as well

as the degree of contextual support for the target words (e.g., Kalikow, Stevens, & Elliott, 1977). All other things being equal, comprehension of more predictable, redundant, and unambiguous linguistic material will be less affected by signal degradation (Miller & Isard, 1963; Miller, Heise, & Lichten, 1951). Contextually aided recognition does not necessarily imply that perceptual mechanisms fundamentally change under ACs, but rather that the opportunity for contextual influences is greater when the signal is degraded. For example, in a Bayesian view of speech recognition (e.g., Norris & McQueen, 2008), the influence of prior probability, e.g., lexical-semantic knowledge, is greatest when there are multiple plausible interpretations of the speech signal—e.g., due to noise or ambiguity in the signal. Hence, the behavioural effect of sentence context on recognition will be largest when the signal is degraded but of intermediate intelligibility, a finding that is well captured by mathematical models of context effects (Boothroyd & Nittrouer, 1988).

Thus, a loss of clarity, degraded speech, or missing perceptual input can lead to a change in the apparent balance of signal-driven and knowledge-driven processes (shown in behavioural data), without any necessary change in the processing mechanisms recruited (see Davis, Ford, Kherif, & Johnsrude, 2011). However, predictable and consistent deviations (e.g., accented or disordered speech) can cause more fundamental recalibration over time, cf. perceptual learning, which will be reviewed in a later section.

Perceptual interference

Interference can occur when the speech signal is forced to compete with a nontarget signal. Interference can be low-level, for instance, when the competing signal simply masks parts of the target signal (cf. energetic masking, see earlier), but it can originate in upper levels as well, for example, when the content of the competing signal affects the interpretation of the target signal or when it draws the listener's attention away from it.

By and large, the speech system handles energetic masking similarly to a lack of information or acoustic-phonetic deviations (see *Failure of recognition*, above), except that the presence of a competing signal comes with the additional challenges of selective attention and signal/noise separation. Intelligibility in noise depends on the degree of spectral overlap between speech and noise (e.g., Brungart, Simpson, Darwin, Arbogast, & Kidd, 2005) and the duration and nature of the glimpses of speech (e.g., Cooke, 2006). While energetic masking generally leads to greater influence of higher-order knowledge (e.g., Mayo et al., 1997), Mattys et al. (2009) found that lexical access can in some cases be so compromised by sensory degradation that whatever salient acoustic cues can be glimpsed through the noise provide better heuristics than does lexical-semantic knowledge.

Like energetic masking, informational masking involves selective attention and is influenced by low-level grouping principles. Unlike energetic masking, however, informational masking depends on the segmental and lexical familiarity of the masker. Greater interference is observed with semantically noticeable distractors (cf. cocktail party effect, Cherry 1953), with babble noise made of intelligible talkers (e.g., Simpson & Cooke, 2005), and with native rather than non-native interfering speech (e.g., Van Engen & Bradlow, 2007; although see Mattys, Carroll, Li, & Chan, 2010, for a failure to find informational masking in a cross-linguistic segmentation task). Furthermore, and unlike energetic masking, informational masking is subject to listener control, at least to some degree. For instance, release from informational masking can be obtained through

training (Leek, 1987; Leek & Watson, 1984) or by factors that increase listeners' attention to the target (Freyman, Balakrishnan, & Helfer, 2004).

Reduced attentional capacity

Reduced attentional capacity often arises as a secondary consequence of other ACs. For instance, listeners' attention tends to be captured by the presence of a distractor (e.g., competing talker), especially when the distractor itself has semantic content—cf. informational masking (e.g., Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006). In this case, the attentional effort can be described as the cost of trying to ignore the distractor and selectively attending to the target. In contrast, ACs incurred under multi-tasking conditions (e.g., speech recognition while driving or monitoring dials) reduce attentional resources directly via divided attention. Under the assumption that attentional resources are in limited supply (Kahneman, 1973), any task executed concurrently with a speech task should interfere with the attentional demands of the latter.

A key question is whether decreased attention affects some components of the speech-recognition system more than others. Fernandes, Kolinsky, and Ventura (2010) found that the extraction of novel words from a continuous speech stream was more greatly affected by a concurrent visual task when the word boundaries in the stream were cued by statistical regularities than when they were cued by coarticulatory regularities. The authors concluded that divided attention was particularly detrimental to domain-general processes, e.g., statistical computation, and less detrimental to low-level acoustic processing. Mattys and Wiget (2011), in contrast, claim that higher-level manifestations of divided attention (e.g., greater reliance on lexical knowledge) are only the cascaded effect of a reduction in perceptual acuity rather than a change of lexical activation per se—see Alais, Morrone, and Burr, 2006, for psychophysical evidence of elevated auditory discrimination thresholds under within-modality dual-tasking conditions. Conversely, Mirman, McClelland, Holt, and Magnuson (2008) suggest that attentional modulation is best implemented within the lexicon, and hence, localises to domain-specific lexical computations. Thus, these results leave unclear whether effects of attention on speech perception are the result of domain-general or domain-specific processing deficits, and whether these effects arise at a perceptual or integrative stage.

It is important to note that the effect of reduced attention can be observed not only in the performance level on the speech task, but also in the degree of listening effort required in the speech task—which can be measured by assessing performance on another task (e.g., Gosselin & Gagné, 2010). Performance on a secondary task often declines when speech is more difficult to understand (e.g., Huckvale & Frasi, 2010, for speech in babble or car noise; Gaskell, Quinlan, Tamminen, & Cleland, 2008, for cross-spliced speech with mismatching phonetic cues; Rodd, Johnsrude, & Davis, 2010, for semantic ambiguity; Sarampalis, Kalluri, Edwards, & Hafter, 2009, for degraded speech). Thus, to the extent that challenges to speech perception and sentence comprehension impact performance of nonlinguistic secondary tasks, this suggests that domain-general processes contribute to effortful speech processing (see Rodd et al., 2010, for discussion).

Reduced memory capacity

Similar to attentional resources, memory demands can arise either indirectly, from other ACs, or directly, from a concurrent memory task. For instance, listening to

various talkers, even sequentially, is shown to engage more working memory resources than listening to a single one (Nusbaum & Morin, 1992), possibly because normalisation of linguistically relevant cues across talkers places demands on computations that would otherwise subserve working memory (Nusbaum & Magnuson, 1997). Recognition of degraded speech also places extra demands on working memory (e.g., Francis & Nusbaum, 2009), as does speech recognition in a background of noise or babble (e.g., Francis, 2010; Rabbitt, 1968), especially for older listeners (e.g., Pichora-Fuller, Schneider, & Daneman, 1995). Concurrent tasks explicitly tapping into working memory (e.g., word/digit list maintenance, detection of repeated items in rapid serial visual presentation) have a generally detrimental effect on speech perception (e.g., Francis, 2010) and word segmentation (e.g., Mattys et al., 2009; Toro, Sinnett, & Soto-Faraco, 2005).

Reduced memory capacity is particularly harmful to processes requiring computation over long stretches of speech, i.e., syntactic parsing and semantic integration (e.g., Caplan & Waters, 1999; Just & Carpenter, 1992). However, the fact that the representation of speech maintained in working memory is likely to be phonological (e.g., Baddeley, 1986) means that reduced memory capacity can conceivably affect sublexical processes as well. Mattys et al. (2009) indeed found that reliance on allophonic detail for word boundaries was attenuated in favour of lexical evidence when listeners performed a segmentation task while holding words or nonwords in short-term memory (see also Jacquemot, Dupoux, Decouche, & Bachoud-Lévi, 2006). Thus, reduced memory resources lead to a change in weights between sublexical and lexical processes.

What type of memory representations are most strongly affected by ACs? In research on short-term memory for spoken material, a distinction if often made between auditory echoic memory (Neisser, 1967), which is responsible for better recall of the last item in an auditory list (Crowder & Morton, 1969), and phonological or articulatory mechanisms that preserve information for a longer period through active rehearsal (e.g., Baddeley and Hitch's phonological loop, 1974). These two forms of auditory/verbal short-term memory have been proposed to be neuroanatomically distinct (dorsal vs. ventral pathways in the lateral temporal lobe, e.g., Buchsbaum, Olsen, Koch, & Berman, 2005; Davis & Johnsrude, 2007; Kalm, Davis, & Norris, 2012). Given evidence that both anterior and posterior temporal regions are additionally engaged during degraded speech comprehension (Davis & Johnsrude, 2003; Davis et al., 2011; Obleser et al., 2007), it is plausible that both of these forms of memory are recruited during ACs. However, studies that separate ACs on the basis of their impact on articulatory vs. echoic memory are thus far lacking (however, see Frankish, 2008, for one example of how such studies might be performed).

Perceptual learning

In the previous sections, we discussed the immediate impact of ACs on recognition and on domain-general processes—perceptual selection, attention, and memory—that are engaged during effortful listening situations. We can contrast these processes with later, postrecognition processes which, rather than attempting to achieve optimal comprehension of the ongoing signal, serve to adjust the perceptual system in order to achieve better comprehension of subsequent utterances.

These are instances of perceptual learning, defined by Goldstone (1998, p. 586) as "relatively long-lasting changes to an organism's perceptual system that improve its ability to respond to its environment and are caused by this environment". Perceptual

learning can be construed as a knowledge-driven process, similar to that described during failure of recognition. However, there are two salient differences between perceptual learning and online processes recruited in difficult listening situations: (1) Perceptual learning only concerns the effect of higher-level knowledge on the recognition of subsequent, rather than current input. Hence, typical paradigms for investigating perceptual learning involve assessing the impact of prior exposure to ambiguous or degraded speech on subsequent perception (cf. Norris, McQueen, & Cutler, 2003; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998); (2) Perceptual learning seems to occur more often after successful than unsuccessful recognition—as shown by various demonstrations that perceptual learning is enhanced in the presence of higher-level lexical information, or external feedback concerning speech content (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Norris, et al., 2003).

To some extent, then, it is likely that perceptual learning operates in most of the ACs reviewed in this article. Indeed, dramatic changes in performance can occur over a very small number of practice trials at the start of a typical experiment. We refer to Samuel and Kraljic (2009) for a review of perceptual learning in speech. In the present context, we note that there are surprisingly few ACs in which perceptual learning is entirely absent. Exceptions would include cases in which either the communicative channel or the target utterance changes trial-by-trial so as to prevent effective retuning. For instance, perceptual learning will be less effective for ACs in which speech is degraded by an unstructured or unpredictable masker which cannot easily be learned from one trial to the next (Peelle & Wingfield, 2005) or if key characteristics of the distortion change during testing (e.g., Hervais-Adelman, Davis, Taylor, Johnsrude, & Carlyon, 2011). Perceptual learning will also be diminished if it is unclear which perceptual hypotheses are to be reinforced, e.g., if entirely unintelligible speech is presented or external feedback on speech content is absent (Davis et al., 2005; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008).

WHAT CAN ADVERSE CONDITIONS TELL US ABOUT THE HUMAN SPEECH RECOGNISER?

To the extent that most of the ACs reviewed above constitute a listener's daily auditory experience, our claim is that *speech recognition under ACs* is, by and large, synonymous with *speech recognition* per se. Even when ACs do not lead to obvious surface manifestations such as reduced intelligibility, they often elicit a re-weighing of basic processes and the development of compensatory strategies. Modulations under ACs can give an insight into the relative weights of those strategies in optimal conditions (e.g., Fernandes et al., 2010; Fernandes, Ventura, & Kolinsky, 2007; Mattys, 2004; Mattys, White, & Melhorn, 2005; Newman, Sawusch, & Wunnenberg, 2011; Van Engen & Bradlow, 2007). ACs can also provide a more sensitive testing ground for some of the key issues in speech science and psycholinguistics (some of these are discussed in Dahan and Magnuson, 2006, and McQueen, 2007), as summarised below.

Content and format of lexical representations

Human listeners are good at recognising speech despite large variations in how sounds and words are realised (Luce & McLennan, 2005). This has led researchers to ask how much (if any) of those variations is encoded in our long-term memory and activated during lexical access, as opposed to being discarded early on in the signal-to-representations

mapping process (e.g., Goldinger, 1998; McQueen, Cutler, & Norris, 2006; Pisoni, 1997). AC-related theorising on this topic has focused almost exclusively on source variability (e.g., talker, style, and accent variation; disordered speech) and receiver limitations (e.g., non-native knowledge). Collectively, this research suggests that: (1) A certain amount of episodic information is stored in long-term memory and is retrieved during subsequent speech processing (e.g., Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Maye et al., 2008; Nygaard, Sommers, & Pisoni, 1994), (2) The exact nature of this episodic information and the part of the memory system in which it resides are disputed (e.g., Pisoni & Levi, 2007; Sumner & Samuel, 2009), (3) Episodic traces, wherever they are stored, are likely to coexist with abstract representations (e.g., Eisner & McQueen, 2005; Jesse, McQueen, & Page, 2007), and (4) The merging of episodic and abstract information could be a matter of processing time-course, with abstract representations accessed first and episodic details later (Luce, McLennan, & Charles-Luce, 2003; Mattys & Liss, 2008; McLennan & Luce, 2005).

In contrast, the contribution of environmental/transmission degradation (e.g., additive noise) to the episodic/abstract debate has been largely ignored. Particularly critical in that category is the distinction between degradation with energetic masking (presence of a competing signal) and degradation without energetic masking (no competing signal). It is conceivable, for instance, that the inclusion of episodic information in long-term memory is limited to cases in which the source of degradation can *not* be separated from the target, that is, when there is no physical interfering masker (e.g., band-pass filtered speech, poor room acoustics). In contrast, degradation due to a segregable masker (e.g., background noise or competing talker) would not leave any episodic traces in memory, as the "noise" could first be stripped away by a process of signal separation.

Time-course of lexical access

Research carried out with idealised speech stimuli has led us to believe that the time-course of lexical access strictly shadows the unfolding of the speech signal and, consequently, that word recognition often occurs before the entire word has been heard. While there is little doubt that clearly enunciated words can lead to early and precise lexical activation (e.g., Allopena, Magnuson, & Tanenhaus, 1998; Van Petten, Coulson, Rubin, Plante, & Parks, 1999) and pre-offset recognition (e.g., Grosjean, 1980; Marslen-Wilson, 1984), early identification in ACs is a lot less tractable. For example, Radeau, Morais, Mousty, and Bertelson (2000) found that pre-offset recognition is not observed when isolated words are played at an average conversational speech rate rather than at the conventionally slower citation rate. Likewise, Bard and colleagues (Bard, Shillcock, & Altmann, 1988; Bard, Sotillo, Kelly, & Aylett, 2001) showed that, contrary to the sequential activation claim, people listening to conversational speech often need to hear substantial portions of the speech following the offset of a word in order to cope with its hypo-articulated characteristics. Pre-offset identification is also less apparent for spoken words heard in noise (Orfanidou, Davis, Ford, & Marslen-Wilson, 2011).

ACs also show that the time-course of lexical identification may not have a single end-point. For example, as mentioned earlier, abstract lexical representations and the episodic details associated with them, e.g., voices, noise, distortions, seem to be accessed at different points in time (Luce et al., 2003; McLennan & Luce, 2005). At the same time, eye-tracking research indicates that when such episodic details are explicitly helpful for a task—for instance, when knowledge of dialectal variants can facilitate

lexical disambiguation—these are taken into account surprisingly early (Dahan, Drucker, & Scarborough, 2008). ACs due to a cognitive load, too, impact on the time-course of speech recognition: Using a Ganong-type phoneme-categorisation task (Ganong, 1980), Mattys and Wiget (2011) found that divided attention delayed reliance on fine phonetic detail but not lexical access. In sum, ACs reveal not only that different sources of information enter the recognition process at different times but also that the relative timing of these sources of information can be notably altered by ambient listening conditions.

Word segmentation

The segmentation of connected speech into words is thought to result from both sublexical cues (e.g., acoustic-phonetic, segmental, prosodic) and knowledge-driven inferences (lexical, sentential). However, the relative weights assigned by listeners to those sources of information are highly dependent on the listening conditions. Prosodic cues such as stress and *F*0 movements are resilient to high levels of noise or articulatory imprecision (e.g., Liss, Spitzer, Caviness, Adler, & Edwards, 1998; Mattys, 2004; Mattys et al., 2005; Smith, Cutler, Butterfield, & Nimmo-Smith, 1989; Welby, 2007), whereas coarticulatory cues and transitional probabilities show greater vulnerability (e.g., Fernandes et al., 2007; Mattys et al., 2005). Among the latter cues, sensitivity to transitional probability survives noise better than do acoustic-phonetic cues (Fernandes et al., 2007), even though acoustic-phonetic cues are highly effective in intact speech (Newman et al., 2011).

The role of lexical-semantic information in segmenting speech under ACs is more difficult to assess. While it is generally accepted that listeners benefit from constraining lexical and semantic context when the signal is acoustically degraded (Miller et al., 1951; Obleser & Kotz, 2011), use of contextual information depends heavily on the informativeness of the lexical and contextual information as well as the strength of the alternative cues. For instance, Mattys et al. (2009, Mattys et al. (2010)) found that acoustic cues can outweigh the contribution of lexical information to word segmentation when these acoustic cues (e.g., glottalisation, aspiration) can be glimpsed through background noise. In contrast, lexical information outweighs acoustic cues under cognitive load (e.g., divided attention, short-term-memory load), which Mattys and Wiget (2011) claim is likely to be a cascaded effect of impoverished sensory encoding under cognitive load. In sum, the study of speech segmentation under ACs has broadened our understanding of the functional architecture of the speech system, especially with regard to the flexibility of cue ranking in everyday speech segmentation.

Feed-forward vs. feed-back effects during segment and word recognition

The question of whether speech recognition is an exclusively feed-forward process (perception leading to recognition) or a mixture of feed-forward and feed-back flows (perception can be modified by recognition or by surrounding context) is deeply rooted in research on listening in ACs. Phoneme restoration, for instance, relied on perceptual illusions elicited by energetic masking (superimposed cough, tone, buzz, white noise, Samuel, 1981; Warren & Obusek, 1971) to suggest that lexical knowledge affects phoneme perception. A similar assertion was made by Connine and Clifton (1987), who showed that lexical knowledge could influence the identification of ambiguous phonemes (cf. Ganong, 1980). Segment-report data also provide evidence

for enhanced segment identification in noise when degraded speech contains familiar words (Boothroyd & Nittrouer, 1988).

The reason why ACs are at the core of the feed-forward/feed-back debate is that any higher-level effect on perception is more likely to occur—and be measurable—if the input is too impoverished to support lexical access on its own. However, even striking demonstrations that perceptual experience of degraded speech is modified by prior knowledge need not imply top-down influences rather than late integration of low- and higher-level information. The extent to which higher-order knowledge modulates perception per se or simply biases its output is still a matter for debate (McClelland, Mirman, & Holt, 2006; McQueen, Norris, & Cutler, 2006). One innovative method for distinguishing these two explanations comes from Frankish (2008), who showed that, while the subjective clarity of degraded speech is enhanced by the presentation of matching text, this effect does not lead to the recency effects that are specific to spoken input. Thus, written context provides the perceptual experience of clear speech without restoring perceptual processes per se (such as echoic memory). Functional imaging data using similar paradigms are valuable, and results with both fMRI (Wild, Davis, & Johnsrude, 2012) and EEG (Hannemann, Obleser, & Eulitz, 2007) have been suggested to show top-down effects when prior knowledge is used to support perception of degraded speech.

However, while there is disagreement concerning the role of feedback in explaining the immediate perception of ambiguous or degraded speech segments, there is considerable evidence that top-down mechanisms play a critical role in supporting perceptual learning when speech is degraded by vocoding (Dahan & Mead, 2010; Davis et al., 2005; Hervais-Adelman et al., 2008) or inclusion of ambiguous speech segments (Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Norris et al., 2003). In all these cases, research has demonstrated rapid and long-lasting perceptual learning that is enhanced by external feedback, combined with constraining lexical information. Such results can only be explained by invoking top-down feedback that allows higher-level information to support changes to lower-level perceptual processing. While this result does not imply that top-down feedback is also used in support of on-line perceptual processing (Norris et al., 2003), an argument in favour of on-line top-down feedback has, however, been made on the grounds of parsimony (Davis & Johnsrude, 2007).

Regardless, few of the available results show that feedback operates during the recognition of carefully read and unambiguous speech. This does not mean that feedback, if indeed present in the speech system, is switched on or off depending on listening quality, but rather that the relative time-course of feed-forward and feed-back processes might only allows feed-back to be noticeable when recognition is delayed by ACs and feed-forward processes are consequently slowed down or unsuccessful.

Feed-forward vs. feed-back effects during lexical-contextual integration

Just as there has been controversy concerning the degree to which segment perception is modified by constraining lexical information, there has been similar controversy regarding the influence of sentential semantic and syntactic context on word recognition (e.g., Marslen-Wilson & Tyler, 1980). In fact, Boothroyd and Nittrouer (1988) have shown that a single mathematical framework can provide an elegant

method of quantifying effects of both lexical context on segment identification and sentence context on lexical identification.

Yet, while there is ample empirical evidence for contextual benefits on lexical identification in ACs (e.g., Kalikow et al., 1977; Miller et al., 1951; Miller & Isard, 1963), the critical issue remains whether sentential information constrains lexical access on-line or modifies the processing demands associated with later sentence-level integration. Proponents of off-line lexical-contextual integration claim that lexical activation is context-free and solely sensory-driven, and that contextual information can only have a biasing effect on the recognition output (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979; Zwitserlood, 1989). In contrast, proponents of on-line lexical-contextual integration claim that a sentential context constrains lexical activation at an early stage by inhibiting the activation of lexical candidates that are semantically incompatible with the foregoing context (e.g., Friederici, Steinhauer, & Frisch, 1999; Mattys, Pleydell-Pearce, Melhorn, & Whitecross, 2005; Van Petten et al., 1999).

While the consensus seems to favour the early integration view (Borsky, Tuller, & Shapiro, 1998; Haggort & van Berkum, 2007), ACs place interesting constraints on this mechanism, especially ACs leading to acoustic degradation. Indeed, on the one hand, target degradation is bound to increase the relative contribution of sentential context as a way of compensating for the incomplete sensory input and the resulting diffuse lexical activation. On the other hand, if we assume that ACs affect both the target and the context, the constraining effect of the (degraded) context might itself be attenuated—or, at best, delayed. Electrophysiological and neuroimaging research does indeed show that the balance between sensory-driven and context-driven processes depends on both the level of signal degradation and the strength of the sentential context, with greater effort involved in recognising degraded words when the sentential context is only moderately useful and, independently, greater lexically driven facilitation when the signal is severely degraded (Obleser & Kotz, 2010, 2011; Obleser, Wise, Dresner, & Scott, 2007). However, information on the relative timing of higher-level and lower-level processes is required to distinguish between top-down feedback and late integration accounts. Time-resolved fMRI data have not yet provided unambiguous evidence for top-down processes (Davis et al., 2011) and further evidence from methods such as MEG combining moderate spatial and millisecond temporal resolution is required.

Interface between speech and cognition

A great deal of research on spoken-word recognition has been carried out without much consideration for its potential interplay with nonlinguistic, cognitive resources. ACs are an ideal ground for examining speech and cognitive processes in combination. For example, "Auditory cognitive science" (Holt & Lotto, 2008) and "Cognitive hearing science" (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2010) are emerging cross-disciplinary fields focusing on the effects of ACs (arising from hearing impairment, aging, and non-native knowledge) on short-term memory recruitment, attention, and cross-modal integration during speech perception. Specifically, within cognitive hearing sciences, the Ease of Language Understanding model (Rönnberg, Rudner, Foo, & Lunner, 2008, see also Rönnberg et al., 2010) suggests that, in cases of segmental-lexical mismatches due to a degraded input, working memory is a key predictor of intelligibility, owing to its role in retrospectively and prospectively reconstructing missing information. Thus, in that conceptualisation, it is possible to envisage individual differences in

perception of degraded speech as a manifestation of individual differences in memory functions. Likewise, research on compensation (or failure to compensate) for low-level hearing deficits has often limited its scope to the contribution of other sources of linguistic information, such as lexical or syntactic knowledge. However, explicit links to nonlinguistic functions and cognitive faculties (e.g., executive functions, attention, speed of processing, IQ) have proved helpful in refining models of speech perception in hearing-impaired individuals (e.g., Akeroyd, 2008), older adults (e.g., Pichora-Fuller & Singh, 2006; Schneider, Daneman, & Murphy, 2005), and even in spoken language technology (e.g., Moore, 2010). Relatedly, ACs can provide an insight into the degree to which various processes involved in speech recognition are subject to active attentional control or, instead, automatic. For instance, speech tasks requiring active inhibition (e.g., ignoring voice characteristics) or selection (e.g., choosing "bat" rather than "pat" when hearing "?at") are shown to be particularly sensitive to divided attention and working memory load (e.g., Nusbaum & Schwab, 1986). In contrast, processes involving memory retrieval based on passive familiarity, e.g., recognition of familiar words, are less resource-demanding (e.g., Jacoby, 1991). This distinction is consistent with recent data by Mattys and Wiget (2011), who found that processing fine phonetic detail towards phoneme identification (e.g., deciding if "?ift" starts with a /g/ or a /k/) was more sensitive to cognitive load than relying on lexical knowledge (e.g., favouring a /g/ answer because "gift" is a word whereas "kift" is not).

Neural responses to speech in ACs

Functional neuroimaging and electrophysiological methods have been widely used to explore the cognitive and neural processes that support the perception of speech under ACs. Readers interested in the basic anatomical foundations of intelligible speech perception are referred to review articles by Scott and colleagues (Rauschecker & Scott, 2009; Scott & Johnsrude, 2003), Hickok and Poeppel (2007), and Davis and colleagues (Davis & Gaskell, 2009; Davis & Johnsrude, 2007).

Two practical challenges arise in assessing neural responses associated with processing speech in ACs. The first challenge is the level of scanner noise during data acquisition in a typical fMRI experiment; this in itself provides an extremely disruptive AC (Peelle et al., 2010). To address this problem, a majority of speech-perception fMRI studies use a form of sparse imaging in which speech is played in silent intervals between MRI scans that measure the neural response to the preceding stimulus (see, e.g., Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999). The second challenge is to ensure that the contrast of clear speech and speech in AC is not confounded by differences in intelligibility. For this reason, studies often use correlational designs, compare two or more forms of distorted speech that are equated for intelligibility, or look for interactions between speech content (such as predictability or semantic coherence) and signal degradation.

An fMRI study reported by Davis and Johnsrude (2003) combined a correlational design with three forms of distortion that were matched for intelligibility (vocoded, interrupted, and energetically masked sentences). The study revealed an increased response to all three forms of distorted speech in a large region of left inferior frontal and premotor cortex. One controversial proposal is thus that the recruitment of premotor regions typically associated with speech production reflects a form of analysis-by-synthesis specific to degraded speech conditions (Davis & Johnsrude, 2007; Poeppel, Idsardi, & van Wassenhove, 2008). In addition, the bilateral posterior superior and middle temporal gyri (STG/MTG) showed an overall increase in

response to degraded speech but further differentiated the three types of distortion through an increased response to speech interrupted with noise. Subsequent work has suggested that these responses in bilateral regions may contribute to the perceived continuity of interrupted speech (Heinrich, Carlyon, Davis, & Johnsrude, 2008; Shahin, Miller & Bishop, 2009). Finally, a region of the left inferior frontal gyrus (IFG) and insula responded specifically to vocoded speech, a finding that is perhaps related to the auditory learning responses observed by Giraud et al. (2004) and Eisner, McGettigan, Faulkner, Rosen, and Scott (2010) for vocoded speech.

Several studies have more systematically explored how lower-level acoustic and higher-level linguistic processes combine in compensating for distorted speech. A PET study by Scott, Rosen, Wickham, and Wise (2004) contrasted neural responses to speech in speech-spectrum noise (energetic masking) with speech against a competing talker (informational masking), while varying signal-to-noise ratios to equate intelligibility. Neural responses in the supplementary motor area (SMA) and ventral IFG were correlated with the level of an energetic masker, whereas level-independent effects of the energetic masker were seen in the frontal pole, left dorsolateral prefrontal cortex and right posterior parietal cortex. Thus, there is widespread recruitment of frontal-parietal regions when listening to speech in noise, including responses in regions that contribute to articulation (SMA) and higher-order semantic processes (ventral IFG). In contrast, informational masking revealed only level-independent effects in bilateral STG, consistent with informational masking being associated with competition for linguistic processes that are engaged by both target and masking speech (Iyer, Brungart, & Simpson, 2010). A further study of informational masking showed differential responses in the left and right STG to sentences masked with speech compared to sentences masked with spectrally rotated speech (which is of similar acoustical complexity but entirely unintelligible). Whereas neural responses in the left STG were enhanced specifically for masking speech (Scott, Rosen, Beaman, Davis, & Wise, 2009), responses in the right STG were increased when either speech or spectrally rotated speech was used as a masker. This finding was interpreted by Scott et al. as indicating that both acoustic and linguistic factors contribute to informational masking and with differential contributions from the two hemispheres. Whereas masking with complex acoustic signals increases activation in the right STG, only masking with intelligible speech leads to an increased response in the left STG. This finding is attributed to the left lateralisation of neural responses to intelligible speech, both when it is the target signal and when it is the masker.

A further set of studies, exemplified by work from Obleser and colleagues, focussed on whether and how systems involved in higher-level semantic and syntactic processing contribute to the perception of speech in degraded conditions. An initial study by Obleser et al. (2007) explored neural correlates of the effect of sentence context on the intelligibility of vocoded speech. They found that predictable sentences were more accurately reported and evoked greater activity in the left medial prefrontal cortex, ventral prefrontal cortex, inferior parietal lobe, and posterior cingulate. Further work also showed effects of close probability (Obleser & Kotz, 2010) and syntactic complexity (Obleser, Meyer, & Friederici, 2011) in response to degraded speech in inferior frontal and posterior temporal regions. While these findings are proposed to show top-down processes, with fronto-parietal regions guiding lower-level perceptual processes in the temporal lobe (Obleser et al., 2007), this remains controversial. Evidence of the relative timing of higher- and lower-level regions—or connectivity analyses showing top-down causal influences—is required to demonstrate the direction of information flow. Such analyses are challenging for conventional

sparse fMRI or PET data which do not provide suitable temporal resolution. Schwarzbauer, Davis, Rodd, and Johnsrude (2006) describe a form of sparse imaging that provides the resolution required to assess the relative timing of frontal and temporal responses. In a recent experiment using this method, Davis et al. (2011) showed interactions between sentence content and signal quality in both frontal and temporal responses to degraded speech. However, the relative timing of these responses was contra to the predictions of a top-down account with frontal responses lagging rather than leading temporal lobe regions. One interpretation of these findings is that neural combination of acoustic and linguistic information operates through lower-level processes maintaining information until it can be satisfactorily integrated by higher-level sentential or semantic information. These findings illustrate the potential value of combining ACs and functional imaging in assessing long-standing cognitive issues in spoken-language comprehension.

Finally, while a review of the neurobiology of auditory attention is beyond the scope of this article, a topic for further research concerns the effect of degradation-free ACs (e.g., memory load, divided attention, intoxication) on neural responses to speech. In previous sections, we noted that cognitive load led to a reduction in perceptual acuity (Mattys & Wiget, 2011) and greater interference from auditory distractors (Francis, 2010). Identifying the brain regions involved in those effects could help pinpoint the time-course of attentional effects on the recognition process and the extent to which such effects are under listener control. fMRI studies that have assessed neural responses to a speech signal that listeners are instructed to ignore have shown considerable activation of auditory regions (e.g., Heinrich, Carlyon, Davis, & Johnsrude, 2011; Scott et al., 2004), though higher-level processes (such as semantic processing) are less robust in lightly sedated compared to fully attentive volunteers (Davis et al., 2007).

CONCLUSION

The goal of this article was to provide a review of the various types of ACs that listeners encounter in everyday environments. ACs were categorised based on their origin (speaker, environment, listener) and their effect (recognition failure, perceptual interference, attentional/memory load). Among the effects, we also included perceptual learning, because we believe that difficult listening situations are a natural ground for compensatory mechanisms leading to perceptual recalibration. Our contention is that perceptual learning occurs quickly and automatically whenever ACs are encountered.

A detailed understanding of how the speech system operates under ACs is important for theory not only because it can improve the external validity of speech-recognition models and refine our knowledge of the interaction between language processing and cognition, but also because the coping strategies used by listeners under ACs can reveal mechanisms that might not emerge as clearly within the highly controlled constraints of clear laboratory speech. Our section on what ACs tell us about the human speech recogniser highlighted several phenomena in which ACs lead to a magnification of known processes and striking reweighing or trade-offs between these processes.

In addition, research on speech recognition in ACs can provide valuable input for disciplines concerned with the more practical aspects of verbal communication. For example, the challenge of surface variation (e.g., conversational, accented, non-native,

disordered speech) and the interaction between speech recognition and cognitive factors are of direct relevance for human-factor speech applications and ergonomics (e.g., flight-deck communication, driving safety, human-machine interaction, intelligent systems). The debate about ACs should hopefully give speech engineers theoretical insights for designing robust and realistic automated speech-recognition systems (e.g., Scharenborg, 2007). Likewise, we hope that AC-oriented speech research can provide hearing scientists and speech clinicians more cognitively informed foundations for intervention in hearing-impaired individuals and individuals with learning difficulties (see, e.g., Liss, 2007, for a compelling stance of potential cross-fertilisation). The same logic applies to clinical interventions for individuals prone to cognitive overload or suffering from attentional disorders, for whom models that specify how cognitive functions and acoustic processing interact could prove highly beneficial.

REFERENCES

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Familiarity with a regional accent facilitates comprehension of that accent in noise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 520–529.

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*, S53–S71.

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B*, *273*, 1339–1345.

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

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529–555.

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology*, *50*, 371–384.

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In S. Greenberg & W. Ainsworth (Eds.), *The auditory basis of speech perception* (pp. 231–308). Berlin: Springer.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 8, pp. 47–89). New York: Academic Press.

Badecker, W. (2005). Speech perception following focal brain injury. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 524–545). Oxford: Blackwell Publishing.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, *44*, 395–408.

Bard, E. G., Sotillo, C., Kelly, M. L., & Aylett, M. P. (2001). Taking the hit: Leaving some lexical competition to be resolved post-lexically. *Language and Cognitive Processes*, *16*, 731–737.

Blumstein, S. E. (2007). Word recognition in aphasia. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 141–155). New York: Oxford University Press.

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*, *84*, 101–114.

Borden, G., Harris, K., & Raphael, L. (2003). *Speech science primer: Physiology, acoustics, and perception of speech* (4th ed.). Baltimore: Lippincott, Williams & Wilkins.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America*, *103*, 2670–2676.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*, 2339–2349.

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

Bradlow, A. R, Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*, 255–272.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274–296.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, *109*, 1101–1109.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G., Jr. (2005). Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *Journal of the Acoustical Society of America*, *117*, 292–304.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*, 687–697.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73–111.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning dialects. *Language and Speech*, *47*, 207–239.

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

Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, *119*, 1562–1573.

Cooke, M. P., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail effect party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*, 414–427.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*, 365–373.

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

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

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

Dahan, D., & Magnuson, J. S. (2006). Spoken-word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249–283). Amsterdam: Academic Press.

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 704–728.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, *12*, 246–269.

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society of London B*, *363*, 1011–1021.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., Owen, A. M., & Menon, D. K. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of the USA*, *104*, 16032–16037.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top-down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*, 3914–3932.

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

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

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*, 132–147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222–241.

Dick, F., Bates, E., Wulfeck, B., Aydelott Utman, J., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localisation and grammar: Evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review*, *108*, 759–788.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*, 89–97.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*, 7179–7186.

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

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *112*, 259–271.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). Cognitive noise is also noise: The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, Psychophysics*, *72*, 1522–1532.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics*, *69*, 856–864.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception SRT for impaired and normal hearing. *Journal of the Acoustical Society of America*, *88*, 1725–1736.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1276–1293.

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception, Psychophysics*, *72*, 501–516.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*, 1360–1374.

Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language*, *58*, 815–836.

Freyman, R. L., Balakrishnan, U., & Helfer, K. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America*, *115*, 2246–2256.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition*, *27*, 438–453.

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

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

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

Gaskell, M. G., Quinlan, P. T., Tamminen, J., & Cleland, A. A. (2008). The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*, *137*, 282–302.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, *14*, 247–255.

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

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

Gosselin, P. A., & Gagné, J. -P. (2010). Use of dual-task paradigm to measure listening effort. *Canadian Journal of Speech-Language Pathology and Audiology*, *34*, 43–51.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*, 267–283.

Haggort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B*, *362*, 801–811.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*, 213–223.

Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research*, *1153*, 134–143.

Harding, A., & Grunwell, P. (1996). Characteristics of cleft palate speech. *European Journal of Disorders of Communication*, *31*, 331–357.

Hazan, V., & Baker, R. (2011) Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, *130*, 2139–2152.

Heinrich, A., Carlyon, R., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*, 1737–1752.

Heinrich, A., Carlyon, R., Davis, M. H., & Johnsrude, I. S. (2011). The continuity illusion does not depend on attentional state: fMRI evidence from illusory vowels. *Journal of Cognitive Neuroscience*, *23*, 2675–2689.

Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *Journal of Speech and Hearing Research*, *37*, 429–438.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 460–474.

Hervais-Adelman, A., Davis, M. H., Taylor, K., Johnsrude, I. S., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 283–295.

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

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*, 42–46.

Huckvale, M., & Frasi, D. (2010). Measuring the effect of noise reduction on listening effort. *Audio engineering society 39th conference on audio forensics*. Copenhagen, Denmark.

Huggins, A. W. F. (1975). Temporally segmented speech. *Perception & Psychophysics*, *18*, 149–157.

Iyer, N., Brungart, D. S., & Simpson, B. D. (2010). Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task. *Journal of the Acoustical Society of America*, *128*, 2998–3010.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Jacquemot, C., Dupoux, E., Decouche, O., & Bachoud-Lévi, A.-C. (2006). Misperception in sentences but not in words: Speech perception and the phonological buffer. *Cognitive Neuropsychology*, *23*, 949–971.

Jesse, A., McQueen, J. M., & Page, M. (2007). The locus of talker-specific effects in spoken word recognition. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th international congress of phonetic sciences* (pp. 1921–1924). Dudweiler: Pirrot.

Jiang, J., Chen, M., & Alwan, A. (2006). On the perception of voicing in syllable-initial plosives in noise. *Journal of the Acoustical Society of America*, *119*, 1092–1105.

Juang, B. H. (1991). Speech recognition in adverse environments. *Computer Speech and Language*, *5*, 275–294.

Junqua, J.-C., & Haton, J.-P. (1995). *Robustness in automatic speech recognition: Fundamentals and applications*. Norwell, MA: Kluwer Academic Publishers.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*, 1337–1351.

Kalm, K., Davis, M. H., Norris, D. (2012) Neural mechanisms underlying the temporal grouping effect in short-term memory. *Human Brain Mapping*, *33*, 1634–1647.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, *54*, 482–499.

Kidd, G. Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational masking. In W. Yost (Ed.), *Springer Handbook of Auditory Research, 29: Auditory Perception of Sound Sources* (pp. 143–190). New York: Springer.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 54–81.

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

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America*, *112*, 2165–2172.

Leek, M. R. (1987). Directed attention in complex sound perception. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 278–288). Erlbaum, Hillsdale, NJ.

Leek, M. R., & Watson, C. S. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical Society of America*, *76*, 1037–1044.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*, 447–454.

Liss, J. M. (2007). The role of speech perception in motor speech disorders. In G. Weismer (Ed.), *Motor speech disorders* (pp. 187–219). San Diego: Plural Publishing.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America*, *104*, 2457–2466.

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx*, *37*, 101–119.

Luce, P. A., & McLennan, C. T. (2005). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 591–609). Oxford: Blackwell Publishing.

Luce, P. A., McLennan, C. T., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). New York: Oxford University Press.

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

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X. Control of language processes*. Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Mattys, S. L. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 397–408.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243.

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*, 887–899.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as from artificial normality. *Perception & Psychophysics*, *70*, 1235–1242.

Mattys, S. L., Pleydell-Pearce, C. W., Melhorn, J. F., & Whitecross, S. E. (2005). Detecting silent pauses in speech: A new tool for measuring on-line lexical and semantic processing. *Psychological Science*, *16*, 958–964.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477–500.

Mattys, S. L., & Wiget, L. (2011). Effect of cognitive load on speech recognition. *Journal of Memory and Language*, *65*, 145–160.

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

Mayo, L. H., Florentine, M., & Buss, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*, 686–693.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*, 363–369.

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

McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.

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

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*, 533.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*, 329–335.

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*, 217–228.

Miller, A. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, *27*, 167–173.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science*, *32*, 398–417.

Mitterer, H. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103.

Moore, R. K. (2010). Cognitive approaches to spoken language technology. In F. Chen & K. Jokinen (Eds.), *Speech technology: Theory and applications* (pp. 89–103). New York: Springer.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Nábelek, A. K. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America*, *84*, 476–484.

Nábelek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America*, *75*, 632–634.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

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

Nilsson, M., & Kleijn, W. B. (2001). Avoiding overestimation in bandwidth extension of telephony speech. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, *2*, 869–872.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395.

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

Nusbaum, H., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 109–132). San Diego, CA: Academic Press.

Nusbaum, H., & Morin, T. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 66–94). Tokyo: IOS Press.

Nusbaum, H. C., & Schwab, E. X. (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. I. Speech perception* (pp. 113–157). San Diego: Academic Press.

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

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, *28*, 8116–8123.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633–640.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*, 713–723.

Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *Neuroimage*, *56*, 2310–2320.

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse conditions. *Journal of Neuroscience*, *27*, 2283–2289.

Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *Quarterly Journal of Experimental Psychology*, *64*, 96–121.

Osberger, M. J., & McGarr, N. S. (1982). Speech production characteristics of the hearing-impaired. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice*, Vol. 8 (pp. 221–284). New York: Academic Press.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, *26*, 844–851.

Parikh, G., & Loizou, P. (2005). The influence of noise on vowel and consonant cues. *Journal of the Acoustical Society of America*, *118*, 3874–3888.

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage*, *52*, 1410–1419.

Peelle, J. E., & Wingfield, A. (2005). Dissociable components of perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1315–1330.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility difference between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434–446.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, *32*, 600–603.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608.

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiological rehabilitation. *Trends in Amplification*, *10*, 29–59.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Pisoni, D. B., & Levi, S. V. (2007). Some observations on representations and representational specificity in speech perception and spoken word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of Psycholinguistics* (pp. 3–18). Oxford: Oxford University Press.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*, 1071–1086.

Prigatano, G. P., & Schacter, D. L. (1991). *Awareness of deficit after brain injury*. New York: Oxford University Press.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, *20*, 241–248.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory and Language*, *42*, 406–422.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*, 182–188.

Rogers, C. L., Lister, J. J., Febor, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*, 465–485.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*, S171–S177.

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health*, *12*, 263–269.

Rostolland, D. (1982). Acoustic features of shouted voice. *Acustica*, *50*, 118–125.

Rostolland, D. (1985). Intelligibility of shouted voice. *Acustica*, *57*, 103–121.

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

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, Psychophysics*, *71*, 1207–1218.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*, 1230–1240.

Scharenborg, O. (2007). Reaching over the gap: A review of efforts to link human and automatic speech recognition research. *Speech Communication*, *49*, 336–347.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging*, *20*, 261–271.

Schwarzbauer, C., Davis, M. H., Rodd, J. M., & Johnsrude, I. S. (2006). Sparse imaging with interleaved, silent steady state (ISSS). *Neuroimage*, *25*, 774–782.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J., & Wise, R. J. S. (2009). The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *Journal of the Acoustical Society of America*, *125*, 1737–1743.

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America*, *115*, 813–821.

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage*, *44*, 1133–1143.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies* (Unpublished doctoral dissertation). University of California, Berkeley.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. D. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*, 3306–3316.

Simpson, S., & Cooke, M. P. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America*, *118*, 2775–2778.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass*, *3*, 236–264.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research*, *32*, 912–920.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, *84*, 917–928.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*, 487–501.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.

Tanenhaus, M. K., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*, 427–440.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). The consequences of diverting attention within and across sensory modalities on statistical learning. *Cognition*, *97*, B25–B34.

Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 207–235). Malden, MA: Blackwell Publishers.

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, *121*, 519–526.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Timecourse of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 394–417.

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

Warren, R. M., & Obusek, C. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, *9*, 358–363.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, *49*, 28–48.

Wild, C., Davis, M. H., & Johnsrude, I. S. (2012). The perceptual clarity of speech modulates activity in primary auditory cortex: fMRI evidence of interactive processes in speech perception. *Neuroimage*, *60*, 1490–1502.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature*, *352*, 236–238.

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