Patti Adank, 25 July 2018

Neural Mechanisms Governing Perception of Speech under Adverse Listening Conditions

Patti Adank

Department of Speech, Hearing and Phonetic Sciences,

University College London,

Chandler House,

2 Wakefield Street, London, WC1N 1PF, United Kingdom

p.adank@ucl.ac.uk

Short title: Neural Substrates of Distorted Speech Perception

Number of Words: 5897

Number of Figures: 1

Number of Tables: 1

Abstract

Listeners are able to understand each other in a wide variety of adverse listening conditions. Listening conditions that present a challenge to speech perception can be attributed to environmental and/or source-related distortions. Environmental distortions originate from outside the speaker, and include background sounds such as noise (energetic masking) or competing speakers (informational masking). For source distortions, degradation originates from the speaker's speech style or voice (e.g., an unfamiliar accent). This chapter integrates results from neuroimaging (e.g., functional Magnetic Resonance Imaging) and neurostimulation (e.g., Transcranial Magnetic Stimulation) studies focusing on the cognitive and neural mechanisms governing listening under adverse listening conditions. Neuroimaging studies indicate that the neural substrates for processing speech in adverse listening conditions compared to speech in quiet are distributed across temporal, frontal and medial areas. Informational masking tends to recruit a network of areas associated with auditory processing (particularly, superior temporal cortex), while energetic masking and source distortions recruit additional areas, including motor and premotor regions. Neurostimulation studies suggest that premotor cortex is crucial for processing speech in energetic maskers. However, these studies have almost exclusively targeted frontal (premotor and motor) areas, and future studies using a combination of both types of methods can further elucidate the precise neural mechanisms involved in understanding speech under distinct adverse listening conditions.

1. Introduction

Perceiving speech in everyday situations seems effortless. We are able to understand each other in a wide variety of ecological situations. This ability of human listeners to perceive speech in demanding circumstances demonstrates the robustness and flexibility of the human spoken language comprehension system. Speech perception is defined here in the broadest sense: as all auditory, cognitive, and neural processes required to classify, understand and interpret spoken utterances at all linguistic levels, from phoneme to discourse. Most of everyday speech perception in fact occurs in adverse listening conditions, and it is fairly rare that a conversation occurs under ideal listening conditions, i.e., in quiet, with our full attention on the conversation, and speaking to someone whose voice and speaking style is familiar. Speech perception in adverse listening conditions is often slower and less efficient than under less challenging conditions. Adverse conditions that present a challenge to speech perception can be classified into environmental and source distortions (Mattys, Davis, Bradlow, & Scott, 2012). First, the speech signal can be masked by distortions originating from the speaker's environment, such as background noise, or competing speakers (Figure 1). Second, the distortions can originate from the source, i.e., directly from the speaker's speech production, e.g., a hoarse voice, or an unfamiliar regional or foreign accent.

Environmental distortions can be further classified into two main types: energetic and informational (Mattys et al., 2012). Energetic distortions are defined as variation sources masking the target speech spectrally and temporally, e.g., simultaneous background noise. The presence of background noise tends to decrease the intelligibility of the speech signal. It has been possible since the 1950s to predict the relative intelligibility of the speech signal based on its signal-to-noise ratio (SNR). Lower SNRs decrease the intelligibility of the speech signal for speech-shaped noise (i.e., noise with the long-term spectral characteristics of speech). Informational distortions are generally defined as competing speech signals, e.g., the presence of one or more background speakers. As energetic maskers tend to completely block the target speech spectrotemporally and they compete with the speech signal at the level of the cochlea.

Informational masking can be defined as the acoustic consequences of the informational distortion after all acoustic consequences of the energetic masking are accounted for (Mattys et al., 2012). Effects of informational masking on intelligibility are less straightforward to pinpoint than those of energetic masking, because informational masking signals often allow listeners to glimpse parts of the target due to the fluctuating spectral amplitude of the masking signal (Cooke, 2003). Moreover, in contrast with energetic masking, the extent to which informational masking affects speech perception is dependent on the segmental and lexical familiarity of the listener with the masker. Speech perception is more perturbed by informational maskers containing semantically observable information, e.g., with babble noise constructed from intelligible speakers.

Source distortions originate from the speaker's style of speech (e.g., regional or foreign accent, fast or slow speech rate, sloppy or formal speaking style), or voice (e.g., hoarse voice, noise-vocoded speech). Listeners tend to show less efficient perception for speech in an unfamiliar regional accent, specifically when combined with an environmental masker (Adank, Evans, Stuart-Smith, & Scott, 2009), for fast speech (Dupoux & Green, 1997), and for noise-vocoded speech (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005). Fast speech is generally generated using artificial time-compression, using a manipulation that reduces the utterance duration without affecting its fundamental frequency. Noise-vocoded speech is created by passing the original speech signal through a channel noise vocoder. Noise-vocoded speech sounds like a harsh, rough, whisper, yet is largely intelligible (depending on the number of channels used, >six channels is intelligible), but the harmonic structure is no longer intact, so the intonation pattern is disrupted.

--- Insert Figure 1 about here ---

This chapter will provide an overview of how processing of adverse listening conditions has been investigated using functional neuroimaging methods, specifically functional Magnetic Resonance Imaging (fMRI) and Position Emission Tomography (PET), and brain stimulation methods, i.e., Transcranial Magnetic Stimulation (TMS). Functional MRI and PET are ideally suited for outlining the network of brain areas involved in speech processing in adverse listening conditions. However, it remains unclear to which extent any brain areas active during processing of adverse listening conditions are causally involved, as neuroimaging methods can only establish a correlative link between activation of a brain area and task performance. Neurostimulation methods, such as TMS, involve direct stimulation of neural tissues using a pulse delivered through the scalp noninvasively. Specifically, TMS can be used in two main ways: first, unlike neuroimaging methods, TMS can establish causal links by temporarily disrupting neural functioning in a target brain area and measuring task performance before and after stimulation. If task performance is affected post-stimulation, then a causal link can be assumed. Second, TMS can be used to determine the extent to which primary motor cortex (M1) is facilitated during task performance or perception by measuring Motor Evoked Potentials (MEPs). MEPs are comparable to fMRI/PET in terms of explanative power, as they are also used to show correlational links between behaviour and brain activation (Adank, Nuttall, & Kennedy-Higgins, 2016). This chapter discusses neuroimaging and neurostimulation studies related to environmental and source distortions with the aim of elucidating the neural mechanisms associated with processing speech in adverse listening conditions in general.

2. Neuroimaging

2.1 Environmental: Energetic

Several fMRI studies scanned participants while listening to speech target stimuli in the presence of energetic maskers. Osnes, Hugdahl, & Specht (2011) presented participants with CV (/da/ and /ta/) syllables in quiet and in seven SNRs of white noise, in a sparse sampling design. They also presented participants with non-speech sounds and musical sounds (piano or guitar chords), and participants were to identify the stimuli as speech, noise, or music. Osnes et al. report a graded increase in activation in left superior temporal sulcus (STS) for decreasing SNRs. Premotor cortex activity was present at intermediate SNRs, when the syllables were identifiable but still distorted. Premotor activity was not reported for syllables in the most favourable SNRs.

Participants in Du, Buchsbaum, Grady, & Alain (2014) identified the initial phoneme in four CV syllables (/ba/, /ma/, /da/, or /ta/) presented in six SNRs (-2, -9, -6, -2, 8dB, and in quiet). Du et al. tested the hypothesis that speech production motor areas contribute to categorical speech perception under adverse, but not quiet, listening conditions. A negative correlation was observed between neural activity and perceptual accuracy in left premotor cortex, which contributed to phoneme categorisation specifically at moderate-to-adverse SNRs.

Wong, Uppanda, Parrish, & Dhar (2008) presented participants with words in quiet, in moderately loud noise (+20dB SNR), and in loud noise (-5dB SNR). Wong et al. used a sparse-temporal scanning paradigm, thus ensuring that the stimuli were presented in relative silence. The noise was multi-talker babble noise, classified here as an energetic masker. They report increased activation in posterior superior temporal gyrus (STG) and left anterior insula for the words presented in -5dB SNR noise compared to +20dB SNR noise.

Adank, Davis, & Hagoort (2012a) scanned listeners while they performed a semantic verification task for sentences in quiet and background noise (+2dB SNR). Compared to sentences in quiet, listening to sentences in noise was associated with increased activation in left inferior frontal gyrus (IFG) and frontal operculum (FO), medial areas including anterior cingulate gyrus (ACC), parahippocampal gyrus, and caudate nucleus.

Zekveld, Heslenfeld, Festen, & Schoonhoven (2006) presented participants with sentences in increasing noise levels; the SNR was varied in 144 steps between +5dB and -35dB SNR. Higher activation was found in left middle frontal gyrus (MFG), left IFG, and bilateral temporal areas for increasing noise levels.

Finally, Hwang, Wu, Chen, & Liu (2006) measured neural responses while participants heard stories in quiet, or mixed with white noise at +5dB SNR. They report reduced activation in left superior and middle temporal gyri, parahippocampal gyrus, cuneus and thalamus for the +5dB condition relative to speech in quiet. They also report reductions in right lingual gyrus, anterior and middle STG, uncus, fusiform gyrus, and right IFG.

2.2 Environmental: Informational

Several fMRI and PET studies scanned participants while listening to speech target stimuli in the presence of informational maskers, or directly compared the neural networks associated with processing speech in the presence of informational or energetic maskers. Dole, Meuneir, & Hoen (2014) investigated neural correlates of speech-in-speech perception (informational masking) in neurotypical controls and participants with dyslexia (not discussed here) using fMRI. Listeners performed a subjective intelligibility rating test with single words played against concurrent maskers consisting of babble noise from four speakers. In the condition designed to maximise informational masking, target words were presented to the right ear, whereas babble noise was presented to the left ear at equal intensity. The authors argue that a second condition maximised energetic masking, as both the target word and noise were presented to the right ear only at a signal-to-noise ratio SNR of 0dB. In this condition, both signals were to be encoded in the same cochlea, thus maximising energetic masking (albeit using a noise signal that is classified here as an informational masker). The informational masking minus energetic masking contrast showed increases in the Blood-Oxygenated-Level Depended (BOLD) response in right STG, while the reverse contrast showed increased activity in right IFG, left MFG, left STG, and left supplementary motor area (SMA).

Scott, Rosen, Beaman, Davis, & Wise (2009) examined neural effects of masking from speech and two additional maskers derived from the original speech, while participants listened passively to sentences, using PET. The first additional maskers consisted of spectrally rotated versions of the sentences, while the second consisted of speech-modulated noise. Rotated speech represents a spectral inversion of the original speech signal, in which the spectrum of low-pass filtered speech is inverted around a centre frequency. It has a similar temporal and spectral structure to the original speech signal, but is not intelligible. Three sets of stimuli were presented to participants: speech-in-speech, speech-in-rotated-speech, and speech-in-speech-modulated-noise (energetic masking baseline). The speech-in-speech masker was linked to increased bilateral STG activation, compared to the speech-modulated-noise baseline, and masking speech with spectrally rotated speech was related only to right STG activation relative to the baseline. Scott et al. argue that informational masking links to two main asymmetrically distributed neural loci, one related to linguistic processes engaging the left superior temporal sulcus (STS)/STG and the other involving right STG, reflecting signal segregation processes related to separating out the signal and masking signals.

Nakai, Kato, & Natsuo (2005) measured the BOLD response while participants listened to a story narrated by a female speaker that was masked by speech from a male or female (same as the narrator) speaker. Bilateral increases in the BOLD-response were reported in STG for the male talker blocks compared to the unmasked baseline condition. However, the masked condition with the female (same) speaker resulted in greater activation in a network spanning the bilateral temporal lobes and the prefrontal and parietal lobes. A direct contrast of the same speaker and different speaker masked conditions showed increases in the BOLD-response in pre-SMA, left pre-central gyrus (PCG) and bilateral IFG, right FO, and right SMG.

2.2 Conclusions

Energetic and informational maskers appear to recruit a similar network of cortical areas in frontal, temporal and medial regions (Table 1). However, there are subtle differences between the activation patterns associated with both types of maskers, and these may point to different neural strategies. While both types of maskers recruit bilateral areas in STS/STG, informational maskers seem to recruit this area more than energetic maskers. Moreover, energetic maskers appear to recruit a wider network of areas, notably including premotor and motor areas. It has been suggested that processing a speech target that is completely masked spectrally and temporally leads listeners to rely to a greater extent on top-down processes, and may be related to increased reliance on executive processes including working memory and attention (Mattys et al., 2012). Further studies that directly contrast energetic and informational maskers, ideally using different types of informational maskers (e.g., overlapping in semantic or syntactic content/structure as well as speaker-specific aspects), will further elucidate the question to which extent the neural mechanisms for both types of maskers are similar.

--- Insert Table 1 about here ---

2.3 Source: Unfamiliar Accent

Adank, Noordzij, & Hagoort (2012b) presented listeners in an fMRI study with sentences spoken in familiar and unfamiliar accents. Compared to the familiar accent, increased activation was found for the unfamiliar accent in frontal (bilateral FO and insulas), temporal (left middle temporal gyrus, MTG, bilateral STG), and parietal regions (supramarginal gyrus, SMG). In Adank, Davis, & Hagoort (2012a), listeners were again exposed to sentences spoken in both accents, while performing a speeded semantic verification task. Compared to the familiar accent, listening to sentences in the unfamiliar accent was associated with increased activation in left STG/STS. Yi et al. (2014) tested participants in an fMRI study while they listened to native- and Korean-accented English sentences. They report that foreign-accented speech evoked greater activity in bilateral STG/STS and IFG.

2.4 Source: Fast Speech

Poldrack et al. (2001) presented participants with sentences compressed to 60%, 45%, 30% and 15% of their original duration. They report compression related increases in BOLD in left middle frontal gyrus, right inferior frontal gyrus, anterior cingulate and the striatum. Peelle, McMillan, Moore, Grossman, & Wingfield, (2004) presented listeners with sentences that were time-compressed to 80, 65, and 50% their duration. Processing speech at higher compression rates recruited areas in bilateral ACC, left striatum, right caudate nucleus, but also in bilateral premotor areas. Participants in Adank & Devlin (2010) listened to sentences at their original speech rate and compressed to 45%. Compression-related increases were found in bilateral anterior and posterior STG/STS, pre-SMA, cingulate sulcus, and bilateral frontal operculi. Processing fast sentences thus seems to recruit a network comprising of bilateral temporal areas, midline areas including the anterior cingulate, pre-SMA, striatum, caudate nucleus, and a set of frontal areas including left IFG and the bilateral FOs.

2.5 Source: Noise-vocoded Speech

Hervais-Adelman, Carlyon, Johnsrude, & Davis (2012) scanned participants while they listened to six-channel noise-vocoded words, clear words, and non-speech stimuli and performed a non-speech target detection task. In comparison with clear words, noise-vocoded words were associated with increases in the BOLD-response in frontal areas, including left

IFG, pre-central gyrus, and left insula. Erb, Henry, Eisner, & Obleser (2013) presented participants in an fMRI experiment with spoken sentences in three conditions: four-band vocoded sentences clear (non-vocoded) sentences (clear speech), and trials lacking any auditory stimulation (silent trials). An increase in the BOLD signal was reported in left SMA, left ACC, anterior insula, and bilateral caudate nucleus for degraded relative to clear speech.

2.6 Conclusions

The extended network for processing source-related distortions recruits areas in bilateral STG/STS, FO and insula, left MFG, left MTG, bilateral IFG, ACC, anterior insula, bilateral PMv, striatum, caudate nucleus, pre-SMA and SMA, left SMG, and pre-central gyrus (Table 1). It is not straightforward to determine to which extent the networks associated with processing different source-related distortions differ from each other and also how the overall network for source-related distortions differs from the network recruited for environmental distortions. Most studies report strong involvement of bilateral STS/STG in processing source distorted speech relative to clear speech, and it seems likely that the neural mechanisms for processing this type of adverse condition are predominantly auditory in nature, as is probably also the case for informational maskers.

3. Neurostimulation

3.1 MEPs

Several neuroimaging studies assessing speech perception in adverse listening conditions discussed earlier report the involvement of (pre-)motor areas (Du et al., 2014; Nakai et al., 2005; Osnes et al., 2011). If has been suggested that (pre-)motor areas, particularly lip and tongue areas of M1, play an active role in supporting speech perception. This is thought to be specifically the case if the incoming speech signal is distorted or unclear. Articulatory M1 is thought to support speech perception using an analysis-by-synthesis approach, in which articulatory motor patterns are used to 'fill in' the missing parts during speech perception (e.g., (Skipper, Devlin, & Lametti, 2017). Several MEP studies tested this specific hypothesis by testing if lip M1 is activated to a greater degree when listening to speech in challenging conditions compared to speech in less challenging conditions.

3.1.1 Environmental: Energetic

Murakami et al. (2011) recorded MEPs after stimulation to the lip area of M1, while participants listened syllables embedded in quiet and several noise levels. Lip MEPs were enhanced for perceiving syllables-in-noise relative to perceiving clear syllables (experiment 4). This result was interpreted to reflect increased excitability of articulatory lip motor representations when listening to speech in noise.

Nuttall, Kennedy-Higgins, Devlin, & Adank (2017) recorded MEPs to test if lip M1 shows differential sensitivity depending on distortion type (motor-distorted or noise, Experiment 1) and quantity (two levels of syllables in noise, Experiment 2), and if lip M1 excitability relates to individual hearing ability. For experiment 1, larger lip M1 MEPs were reported during perception of motor-distorted speech that had been produced using a tongue depressor, and during perception of speech presented in background noise, relative to natural speech in quiet. However, no difference was reported between both distortion types. Experiment 2 did not find evidence of motor system facilitation when speech was presented in noise at SNR levels ratios where speech intelligibility for individual listeners was at 50% (harder) or 75% (easier). However, there was a significant interaction between noise condition and hearing ability, which indicated that when speech stimuli were correctly classified at 50%, speech motor facilitation was observed in individuals with better hearing. Individuals with relatively worse but still normal hearing showed more activation of lip M1 during perception of clear speech. Taken together, these results indicate that articulatory M1 is activated more during perception of speech under adverse conditions, thus supporting claims suggesting a role for M1 in processing distorted speech signals (Skipper et al., 2017). Moreover, results from Nuttall et al. (2017) indicate that M1 becomes more activated whenever the speech signal is more difficult to process, irrespective of whether the distortion is environmental or source-related.

3.1.2 Environmental and Source: Energetic, Motor-distorted

Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank (2016) recorded MEPs from lip and hand (control site) M1 while participants listened to clearly articulated syllables (clear) or syllables articulated while the speaker held a tongue depressor in their mouth (tongue-depressed). Participants passively listened to clear and tongue-depressed VCV syllables (/apa/, /aba/, /ata/, /ada/) in separate blocks, while hand and lip MEPs were collected. After MEP collection was completed, participants performed an identification task for the tongue-depressed stimuli. The results showed facilitation for lip MEPs for tongue-depressed compared to clear stimuli. Moreover, this facilitation was increased for stimuli containing a lip-articulated consonant (/apa/ and /aba/) compared to a tongue-articulated consonant (/ata/ and /ada/). Finally, participants who performed best on the identification task showed the greatest amount of facilitation for lip MEPs.

3.2 TMS

3.2.1 Environmental: Energetic

Meister et al. (2007) tested the causal role of left STG and left ventral premotor cortex (PMv) in perception of CV syllables embedded in white noise and of simple tones. Participants received 15 minutes of 1Hz repetitive TMS to either target site. The study aimed to establish the role of left PMv and left STG in processing speech in noise. Participants performed either a phoneme or tone identification task, or a colour identification control task. Repetitive TMS to left PMv only impaired phoneme discrimination, thus demonstrating a causal effect of TMS to speech perception, but had no effect on tone or colour discrimination tasks. TMS to left STG impaired tone discrimination, but had no effect on phoneme or colour discrimination tasks. Meister et al. argue that the lack of an inhibitory effect of TMS to left STG during syllable discrimination can be attributed to recruitment of a more extensive, bilateral, neural network for speech processing than for tone perception. Speech perception is arguably a more complex process than tone perception, as it encompasses a basic auditory signal processing stage as well as higher-level phonetic and phonological processing stages, which tend to recruit areas in bilateral temporal areas.

Participants in D'Ausilio et al. (2009) performed a phoneme identification task for CV syllables in which the consonant was articulated using either the lips (/pœ/ and /bœ/) or tongue (/tœ/ and /dœ/) embedded in white noise. Participants received TMS pulses to the left lip or tongue area of M1 in an online TMS design. Responses were also collected for when no TMS pulse was given (baseline). The results showed a double dissociation between stimulation site (lip or tongue) and discrimination performance between the primary articulator of the stimuli (lips or tongue). Participants were faster to classify a tongue sound following TMS to tongue M1, and slower to classify a lips sound following a TMS pulse to tongue M1, and vice versa. This pattern in the results was not replicated when the stimuli were presented in quiet, thus showing that the causal role of articulatory M1 was specific to noisy syllables. The results from the virtual lesion TMS studies discussed here demonstrate that articulatory M1 plays a causal role in perception of speech masked by environmental maskers, thus further supporting the proposed role of M1 in perception of distorted speech.

3.2.2 Source: Motor-distorted Speech

Nuttall, Kennedy-Higgins, Devlin, & Adank (in press) examined the connection between left PMv and left lip M1 during challenging speech perception in two experiments that combined collection of MEPs with virtual lesion TMS. Experiment 1 tested intra-hemispheric connectivity between left PMv and left M1 lip perception during comprehension of speech under clear and distorted listening conditions. TMS was applied to left PMv. Next, participants performed a speeded sentence verification task on motor-distorted and clear speech, whilst also undergoing stimulation of left lip M1 to elicit MEPs. Experiment 2 aimed to clarify the role of inter-hemispheric connectivity between right hemisphere PMv and left hemisphere M1 lip area. Dual-coil transcranial magnetic stimulation was applied to right PMv and left lip M1. The results from both experiments indicated that disruption of PMv during speech perception affected comprehension of distorted speech specifically, and listening to distorted speech was found to modulate the balance of intra- and inter-hemispheric interactions, with a larger sensorimotor network implicated during comprehension of distorted speech than when speech perception is optimal.

3.3 Conclusions

Only three TMS studies thus far examined the causal role of cortical areas in processing speech in adverse listening conditions. The results from the three studies clearly support a causal role for (pre-)motor regions in perception of motor-distorted speech and speech in the presence of an energetic masker, thus supporting accounts that propose a supporting role for speech production substrates in speech perception in challenging listening conditions (Skipper et al. 2017). Note that only a single study (Meister et al., 2007) examined the causal role of an area in the temporal lobe (left STG). Yet, Meister et al. did not report a causal role of this area in processing syllables in noise (but reported a causal role of left STG in tone discrimination).

Due to the inherent limitations of TMS, it is not possible to stimulate more medial target areas but there is a clear lack of studies directly targeting, accessible, lateral cortical areas, specifically in STG/STS, while participants process distorted speech at pre-lexical or lexical levels.

4. General Conclusions

This chapter discussed neuroimaging (fMRI/PET) and neurostimulation (MEP/TMS) studies aiming to further our understanding of how the brain processes speech under environmental and source-related adverse listening conditions. The overview of neurostimulation studies in Table 1 displays a different picture from the neuroimaging results. While neuroimaging studies report the involvement of cortical areas in frontal, temporal, and parietal lobes as well as an extended network of medial areas, neurostimulation studies seems have mostly focused on frontal areas including articulatory M1 and left PMv. For MEP studies, M1 is the obvious target, since it is not straightforward (if not impossible) to elicit MEPs from cortical areas outside (pre-)motor areas of the brain. There is a clear lack of neurostimulation studies examining the role of (bilateral) temporal areas in processing speech in adverse conditions. It is not possible to collect MEPs from areas outside the (pre-)motor areas, but it is surprising that only a single virtual lesion TMS study (Meister et al., 2007), examined the role of STS/STG in processing distorted speech signals, but this study did not confirm a causal role for STG. It may not be straightforward to establish a clear causal effect for temporal regions, presumably due to possible inter-hemispheric compensation during speech processing. Inter-hemispheric compensation can occur especially in so-called offline TMS paradigms, where the application of pulses occurs several minutes before task performance, allowing for online reorganisation or compensation by the non-targeted hemisphere. Future TMS studies might therefore explore either the use of online TMS (where the TMS pulse is delivered during stimulus presentation), or target areas in both temporal lobes simultaneously in either an online or offline paradigm, to limit compensation mechanisms.

This chapter aimed to outline the neural mechanisms associated with processing different types of distortions. The results discussed here can be summarised as that informational maskers tend to recruit a network of areas associated with auditory processing in STGS/STG, while energetic maskers and source distortions also recruit areas outside STS/STG, including motor and (pre-)motor regions. Premotor cortex appears to be crucial for processing speech in energetic maskers. Yet, the precise neural mechanisms associated with each type of distortion remain largely unclear and it is suggested that future studies exploit the respective strengths of neuroimaging and neurostimulation methods to further elucidate these mechanisms. For example, future studies might systematically link fMRI and TMS by first identifying the relevant nodes and second by establishing their causal role in processing speech under adverse listening conditions using a variety of speech stimuli and environmental and source distortions.

Figure 1. Overview of the types of adverse listening conditions discussed in this chapter.

Table I. Overview of studies contrasting speech perception under adverse listening conditions versus easier listening conditions. ACC: anterior cingulate cortex; fMRI: Functional Magnetic Resonance Imaging, PET: Positron Emission Tomography; IFG: inferior frontal gyrus; MEP: Motor Evoked Potential; MFG: middle frontal gyrus; M1: primary motor cortex; MTG: middle temporal gyrus; PMv: ventral premotor cortex, STG: superior temporal gyrus; FO: frontal operculum; STS: superior temporal sulcus, SMA: supplementary motor area; SMG: supramarginal gyrus; TMS: Transcranial Magnetic Stimulation.

Study	**Adverse condition**	**Stimuli**	**Method**	**Areas**
Neuroimaging studies				
Osnes et al. (2011)	Environmental (energetic)	Syllables in noise	fMRI	Bilateral STS, left PMv
Du et al. (2014)	Environmental (energetic)	Syllables in noise	fMRI	Left PMv
Wong et al. (2008)	Environmental (energetic)	Words in noise	fMRI	Bilateral STG and left insula
Adank et al. (2012a)	Environmental (energetic)	Sentences in noise	fMRI	Left IFG, left FO, ACC, parahippocampal gyrus, caudate nucleus
Zekveld et al. (2006)	Environmental (energetic)	Sentences in noise	fMRI	Left MFG, left IFG, bilateral STG/STS

Hwang et al. (2006)	Environmental (energetic)	Stories in noise	fMRI	Left STG/MTG, parahippocampal gyrus, cuneus, thalamus
Dole et al. (2014)	Environmental (energetic)	Words in monaural, binaural, or dichotic conditions	fMRI	Right IFG, left STG, left MFG, left SMA
Dole et al. (2014)	Environmental (informational)	Words in monaural, binaural, or dichotic conditions	fMRI	Right STG
Scott et al. (2009)	Environmental (informational)	Sentences masked by noise-vocoded or spectrally-rotated maskers	PET	Bilateral STG
Nakai et al. (2005)	Environmental (informational)	Stories masked by same > different speaker	fMRI	Pre-SMA, left pre-central gyrus (PCG), bilateral IFG, right FO, right SMG
Adank et al. (2012b)	Source (accent)	Sentences in unfamiliar accent	fMRI	Bilateral FO, bilateral insula, left MTG, bilateral STG, SMG

Adank et al. (2012a)	Source (accent)	Sentences in unfamiliar accent	fMRI	Left STG/STS
Yi et al. (2014)	Source (accent)	Sentences in unfamiliar accent	fMRI	Bilateral STG/STS, bilateral IFG
Poldrack et al. (2001)	Source (time-compression)	Time-compressed sentences	fMRI	Left MFG, right IFG, ACC, striatum
Peelle et al. (2004)	Source (time-compression)	Time-compressed sentences	fMRI	Bilateral ACC, left striatum, right caudate nucleus, bilateral premotor areas
Adank & Devlin (2010)	Source (time-compression)	Time-compressed sentences	fMRI	Bilateral STG/STS, ACC, pre-SMA, striatum, caudate nucleus, left IFG, bilateral FO
Hervais-Adelman, et al. (2012)	Source (noise-vocoding)	Noise-vocoded words	fMRI	Left IFG, left pre-central gyrus, left insula
Erb et al. (2013)	Source (noise-vocoding)	Noise-vocoded sentences	fMRI	Left SMA, left ACC, anterior insula, bilateral caudate nuclei
Neurostimulation studies				
D'Ausilio et al. (2009)	Environmental, energetic	Syllables in noise	MEP	Tongue M1

Murakami et al. (2011)	Environmental (energetic)	Syllables in noise	MEP	Lip M1
Nuttall et al. (2017)	Environmental (energetic)	Syllables in noise	MEP	Lip M1
Nuttall et al. (2016)	Source (motoric)	Syllables, tongue-depressed	MEP	Lip M1
Meister et al. (2007)	Environmental (energetic)	Syllables in noise	TMS	Left PMv
Nuttall et al. (in press)	Source (motoric)	Tongue-depressed sentences	TMS	Right PMv and left lip M1

References:

Adank, P., Davis, M., & Hagoort, P. (2012a). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia, 50*(1), 77-84. doi:10.1016/j.neuropsychologia.2011.10.024

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

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception and Performance, 35*(2), 520-529. doi:10.1037/a0013552

Adank, P., Noordzij, M. L., & Hagoort, P. (2012b). The role of Planum Temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping, 33*(2), 360-372. doi:10.1002/hbm.21218

Adank, P., Nuttall, H. E., & Kennedy-Higgins, D. (2016). Transcranial Magnetic Stimulation (TMS) and Motor Evoked Potentials (MEPs) in Speech Perception Research. *Language, Cognition & Neuroscience*, 1-10. doi:10.1080/23273798.2016.1257816

Cooke, M. (2003). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America, 119*(3), 1562-1573. doi:10.1121/1.2166600

D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*(5), 381-385. doi:10.1016/j.cub.2009.01.017

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., 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*(2), 222-241.

Dole, M., Meunier, F., & Hoen, M. (2014). Functional correlates of the speech-in-noise perception impairment in dyslexia: An MRI study. *Neuropsychologia, 60*, 103-114. doi:10.1016/j.neuropsychologia.2014.05.016

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences, 111*, 7126-7131. doi:10.1073/pnas.1318738111

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914-927.

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

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

Hwang, J. H., Wu, C. W., Chen, J. H., & Liu, T. C. (2006). The effects of masking on the activation of auditory-associated cortex during speech listening in white noise. *Acta Oto-Laryngologica, 126*(9), 916-920. doi:10.1080/00016480500546375

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7/8), 953-978. doi:10.1080/01690965.2012.705006

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

Murakami, T., Restle, J., & Ziemann, U. (2011). Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. *Neuropsychologia, 49*(7), 2045-2054. doi:10.1016/j.neuropsychologia.2011.03.034

Nakai, T., Kato, C., & Matsuo, K. (2005). An fMRI study to investigate auditory attention: a model of the cocktail party phenomenon. *Magnetic Resonance in Medical Sciences, 4*(2), 75-82. doi:10.2463/mrms.4.75

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (2017). The role of hearing ability and speech distortion in the facilitation of articulatory motor cortex. *Neuropsychologia, 94*(8), 13-22. doi:10.1016/j.neuropsychologia.2016.11.016

Nuttall, H. E., Kennedy-Higgins, D., Devlin, J. T., & Adank, P. (in press). Modulation of intra- and inter-hemispheric connectivity between primary and premotor cortex during speech perception. *Brain & Language*. doi:10.1016/j.bandl.2017.12.002

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., & Adank, P. (2016). The effect of speech distortion on the excitability of articulatory motor cortex *NeuroImage, 128*, 218-226. doi:10.1016/j.neuroimage.2015.12.038

Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage, 54*(3), 2437-2445. doi:10.1016/j.neuroimage.2010.09.078

Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language, 91*, 315-325. doi:10.1016/j.bandl.2004.05.007

Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. E. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience, 13*(5), 687-697. doi:10.1162/089892901750363235

Scott, S. K., Rosen, S., Beaman, P., Davis, J. P., & 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). doi:10.1121/1.3050255

Skipper, J., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language, 164*, 77-105. doi:10.1016/j.bandl.2016.10.004

Wong, P. C. M., Uppanda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *Journal of Speech, Hearing and Language Research, 51*, 1026-1041. doi:10.1044/1092-4388(2008/075)

Yi, H., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience, 8*, 1-12. doi:10.3389/fnhum.2014.00768

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage, 32*(4), 1826-1836. doi:10.1016/j.neuroimage.2006.04.199

Index Keywords: speech, perception, neuroimaging, neurostimulation