
Patti Adank
School of Psychological Sciences, University of Manchester, Manchester, United Kingdom

Short title: Mechanisms of Effortful Speech Understanding

Correspondence:
Patti Adank
School of Psychological Sciences
University of Manchester
Zochonis Building
Brunswick Street
M20 6HL
Manchester, UK
Email: patti.adank@manchester.ac.uk
Phone: 0044-161-2752693

Keywords: perception, auditory, speech, articulation, production, fMRI, PET, neuroimaging, meta-analysis.
Abstract

The role of speech production mechanisms in difficult speech comprehension is the subject of on-going debate in speech science. Two Activation Likelihood Estimation (ALE) analyses were conducted on neuroimaging studies investigating difficult speech comprehension or speech production. Meta-analysis 1 included 10 studies contrasting comprehension of less intelligible/distorted speech with more intelligible speech. Meta-analysis 2 (21 studies) identified areas associated with speech production. The results indicate that difficult comprehension involves increased reliance of cortical regions in which comprehension and production overlapped (bilateral anterior Superior Temporal Sulcus (STS) and anterior Supplementary Motor Area (pre-SMA)) and in an area associated with intelligibility processing (left posterior MTG), and second involves increased reliance on cortical areas associated with general executive processes (bilateral anterior insulae). Comprehension of distorted speech may be supported by a hybrid neural mechanism combining increased involvement of areas associated with general executive processing and areas shared between comprehension and production.
1. Introduction

The human speech comprehension system is remarkable in its ability to quickly extract the linguistic message from a transient acoustic signal. Much of everyday processing occurs under listening conditions that are less than ideal, due to background noise, regional accents, or speech rate differences, to name a few common everyday variations in the speech signal. Listeners are generally able to successfully comprehend speech under such adverse - or difficult - listening conditions. Nevertheless, speech comprehension is often slower and less efficient than under less difficult conditions (see Mattys, Brooks, & Cooke, 2009, for an overview). For instance, when performing a semantic verification task (i.e., reporting whether a sentence such as ‘dogs have four ears’ is true or false) spoken in an unfamiliar regional accent, listeners show slower response times and higher error scores (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009). However, the neural mechanisms supporting the intrinsic robustness of the speech comprehension system are largely unclear.

Cognitive neuroscience models of the cortical organisation of spoken language processing (e.g., Hickok & Poeppel, 2007) have not made explicit predictions regarding the neural locus of processing of distorted speech signals. However, this neural locus has been discussed in more detail in various papers investigating difficult speech processing (Davis & Johnsrude, 2003; Peelle, Johnsrude, & Davis, 2010). Davis & Johnsrude and Peelle et al. propose a critical role for left Inferior Frontal Gyrus in processing of distorted speech signals. This proposed role of IFG is motivated by its frequent activation during speech perception and comprehension tasks (e.g., Crinion & Price, 2005; Obleser, Wise, Dresner, & Scott, 2007). Second, it is argued that IFG’s anatomical connectivity to auditory belt and parabelt regions
(e.g., Hackett, Stepniewska, & Kaas, 1999) makes it well-positioned to affect processing in primary auditory and association areas traditionally associated with speech perception. Finally, Peelle et al. argue that Davis and Johnsrude (2003) provide direct evidence a role of IFG in processing distorted speech signals by showing that activity in left IFG was elevated for distorted (but still intelligible) speech compared to both clear speech and unintelligible noise.

In speech science, three general behavioural/neural mechanisms have been suggested to support difficult speech comprehension. First, it has been hypothesised that comprehension relies predominantly on auditory processes and associated brain areas (Holt & Lotto, 2008). Processing distortions of the speech signal is predicted to be governed through involvement of general cognitive processes, such as working memory and/or attention. Second, it has been proposed that comprehension of distorted speech signals recruits neural mechanisms associated with speech production (Hickok & Poeppel, 2007; Pickering & Garrod, 2007; Skipper, Nusbaum, & Small, 2006). This line of reasoning proposes that speech processing in the absence of external distortions relies predominantly on auditory processes, while speech production processes are selectively active when listening conditions deteriorate. Third, it has been put forward that auditory speech processing relies almost entirely on speech motor mechanisms, with only a minor role for auditory (or general cognitive) mechanisms (Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Whalen et al., 2006). Here, the auditory signal is relayed through speech production mechanisms to achieve successful comprehension regardless of the listening conditions.

A number of functional Magnetic Resonance Imaging (fMRI) studies support the view that comprehension of distorted speech relies on the recruitment of speech
production mechanisms. Functional MRI studies investigating processing of distorted speech commonly present listeners with speech stimuli perturbed by adding background noise or multi-speaker babble (Indefrey & Levelt, 2004; Stowe et al., 1998), by passing the speech signal through a noise-vocoder (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), artificially time-compressing the signal (Dupoux & Green, 1997), or by using speech stimuli that have been spoken in an unfamiliar foreign or regional accent (Adank, et al., 2009; Munro & Derwing, 1995). Neural responses associated with processing distorted speech stimuli are then contrasted with more intelligible – undistorted – speech stimuli. Usually, activity related to processing distorted speech is found across a variety of cortical areas, including posterior superior and middle temporal areas bilaterally, left inferior frontal areas, left and right frontal opercula, precentral gyrus bilaterally extending to the Supplementary Motor Area (SMA), and subcortical areas including Thalamus. For instance, Davis & Johnsrude (2003) report activations in temporal and inferior frontal areas, as well as in premotor areas in their Figure 7. The activations in IFG and motor areas including precentral gyrus and SMA are also associated with speech production (Alario, Chainay, Lehericy, & Cohen, 2006; Levelt, 1989; Wilson, Saygin, Sereno, & Iacoboni, 2004), have been reported to be active during speech comprehension (Adank & Devlin, 2010; Wilson, et al., 2004). Furthermore, these areas are considered to be an integral part of the speech production network, as demonstrated in a previous meta-analysis on Positron Emission Tomography (PET) studies on single word reading (Turkeltaub, Eden, Jones, & Zeffiro, 2002). However, it is unclear to which degree specific speech motor areas are consistently activated during comprehension of distorted speech.
The present study aims to, first, identify the network of areas involved in processing distorted speech signals. Second, it aims to determine whether the network for difficult speech comprehension includes areas involved in speech production. If difficult speech comprehension consistently activates areas also involved in speech production, then this would support for the hypothesis that speech production areas are selectively recruited during speech comprehension.

This paper presents two meta-analyses using the Activation Likelihood Estimation (ALE) method (Laird et al., 2005; Turkeltaub, et al., 2002). ALE has been used to determine the overlap between coordinates obtained from neuroimaging studies by modelling them as probability distributions that are centred at the reported coordinates. The first meta-analysis applies ALE to coordinates extracted from studies contrasting difficult speech comprehension with comprehension of intelligible speech under less adverse listening conditions. This analysis aims to identify the areas consistently activated during successful but difficult speech comprehension. The second meta-analysis applies ALE to coordinates extracted from studies that include conditions in which participants produce speech at pre-lexical (e.g., individual speech sounds, syllables) and post-lexical levels (e.g., words, sentences) that are contrasted with conditions in which participants do not produce speech. The aim of the second analysis is to identify areas consistently activated during speech production. This second meta-analysis is intended to represent an extension to the meta-analysis described in Turkeltaub et al. (2002). It was decided to perform a new meta-analysis on the neural network involved in speech production, as Turkeltaub et al. included only PET studies. Recent neuroimaging studies predominantly use fMRI for studying auditory processing of speech, but also for speech production. In addition, Turkeltaub et al. focused on studies that addressed single word reading. The present research also
includes studies that investigate production of pre-lexical linguistic elements, to identify the network of speech production involved in the articulation of pre- and post-lexical speech stimuli, and to avoid skewing the results of the ALE towards areas involved in production of linguistically meaningful elements only.

2. Method

2.1.1 Selection of literature studies for meta-analysis 1

Neuroimaging studies were included that investigated comprehension of distorted (yet intelligible) speech at post-lexical levels. The PubMed online database for was searched for studies using the keywords: “distorted”, “degraded”, “dialect”, “accent”, “sine wave”, “synthetic”, “noise”, “time-compressed”, “noise-vocoded”, “speech”, “intelligibility”, “intelligible”, “comprehension”, “fMRI”, “post-lexical”, “narrative”, “word”, “PET”, “neuroimaging”, and appropriate combinations of these keywords. Additional papers were collected by searching for prominent researchers in the field. Papers were selected from January 1 2001 onwards, including papers in press or in advance online publication.

2.1.2 Inclusion and exclusion criteria for meta-analysis 1

Papers were included that fulfilled the following criteria: i) neural responses were collected using fMRI or PET, ii) only healthy, adult, neurotypical subjects with intact hearing and no known neurological or psychiatric disorders were tested, iii) the experiments contained conditions in which less intelligible speech as well as conditions in which more intelligible speech was presented, iv) speech stimuli were words, sentences, or narratives; v) stimuli were naturally spoken utterances and not synthetic utterances, vi) results were reported at a group-level in a stereotactic 3-coordinate system. In addition, the following criteria were used to exclude papers
from the analysis: single subject studies and studies that report only the results from a pre-specified region-of-interest (ROI). The selected studies are listed in Table I.

A single study included two different distortions (Adank, Davis, & Hagoort, in press) – speech in an unfamiliar accent and speech in a familiar accent with added background noise – and contrasted these with undistorted speech in quiet in a familiar accent. It was decided to include only the coordinate from the contrast involving speech in an unfamiliar accent, as only one other study (Adank, Noordzij, & Hagoort, 2012) used speech in an unfamiliar accent, while two other studies (Davis & Johnsrude, 2003; Wong, Uppanda, Parrish, & Dhar, 2008) used added background noise to distort the speech stimuli (note that Davis et al. also used noise-vocoded and noise-segmented speech).

2.2.1 Selection of literature studies for meta-analysis 2

Neuroimaging studies were included that investigated speech production using PubMed. The PubMed online database for was searched for studies using the keywords: “speech”, “production”, “articulation”, “syllable”, “phoneme”, “word”, “fmri”, “PET”, “neuroimaging”, and appropriate combinations of these keywords. In addition, papers were identified by searching for prominent researchers in the field. As for analysis 1, papers were selected from January 1 2001 onwards, including papers in press or in advance online publication.

2.2.2 Inclusion and exclusion criteria for meta-analysis 1

Papers were included that fulfilled the following criteria: i) neural responses were collected using fMRI or PET, ii) only healthy, adult, right-handed neurotypical subjects with intact hearing and no known neurological or psychiatric disorders were tested, iii) the experiments contained conditions in which participants produced phonemes, syllables, words, sentences, or narratives; iv) results were reported at a
group-level in a stereotactic 3-coordinate system. Again, single subject studies and studies that report only the results from a pre-specified region-of-interest (ROI) were excluded. The selected studies are listed in Table III.

2.3 ALE methods

The ALE analysis was implemented using GingerALE 2.04 (www.brainmap.org). This version of the likelihood estimation algorithm was selected as it has been shown to be more precise than previous versions, while it retains comparable sensitivity (Eickhoff, Laird, Grefkes, Wang, Zilles, et al, 2009). Coordinates collected from studies that reported coordinates in Talairach space were converted to MNI space using the tal2icbm_spm algorithm implemented in the GingerALE software (www.brainmap.org/ale).

In GingerALE, first, modelled activation maps are computed for each set of foci per included study. All foci were modelled as Gaussian distributions and merged into a single 3-dimensional volume. GingerALE uses an uncertainty modelling algorithm to empirically estimate the between-subjects and between-templates variability of all included foci sets. Second, ALE values are computed on a voxel-to-voxel basis by taking the values that are common to the individual modelled activation maps. GingerALE constrains the limits of this analysis to a grey matter mask that was used to define the outer limits of MNI coordinate space, which excludes most white-matter structures (Eickhoff, Heim, Zilles, & Amunts, 2009). The analysis was corrected for multiple comparisons using the FDR (false discovery rate) method at $q < 0.01$, voxel wise, (default = 0.05), using a cluster extent of $400\text{mm}^3$ (default = $200\text{mm}^3$). The Mango software package (http://ric.uthscsa.edu/mango/) was used to view the
resulting activation maps and all results were overlaid on a single MNI template available in Mango (Colin27_T1_seg_MNI.nii).

3. Results

3.1 Meta-analysis 1

Meta-analysis 1 was based on the results of 10 experiments (Table I), 116 participants, and 75 foci that were published in 10 papers. One used a PET design, nine used fMRI, five used a sparse scanning paradigm, and four used a continuous paradigm. In all experiments, a condition in which listeners were required to comprehend distorted speech was included. Several types of distortions were used: added background noise (e.g., in Wong, et al., 2008), speech in an unfamiliar accent (Adank, et al., 2012), artificially time-compressed speech (see Dupoux & Green, 1997, for a description of this specific distortion type and Poldrack et al., 2001, for a study including time-compressed speech). Some used noise-vocoded speech (for example Sharp et al., 2010), a single study included segmented speech (Davis & Johnsrude, 2003, see Schroeder, 1968, for a description of this type of distortion). Finally, one study (Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010) contrasted continuous scanning using standard EPI noise with a “quiet” EPI sequence. This “quiet” sequence minimises the acoustic disturbance associated with traditional EPI using the same imaging parameters (Schmitter et al., 2008). Stimuli were either words or sentences. Nine studies used an experimental task, and one study used passive listening plus an after-task (Adank, et al., 2012). The following experimental tasks were employed: speaker judgment, semantic decision, intelligibility judgment, gender decision, grammatical decision, rhyme decision, target matching, and target to picture matching.
The meta-analysis resulted in eight clusters at the selected significance level (Table II). These clusters show a similar pattern across both hemispheres, but the pattern of ALE clusters was more widespread on the left (3840mm$^3$, excluding clusters 3 and 7, as they are centred at x = 0) than on the right (1760mm$^3$, excluding clusters 3 and 7). The highest ALE score was found for a cluster in left STS, located just anterior to Heschl’s Gyrus. A second cluster was found in posterior left Middle Temporal Gyrus (MTG). Subsequent clusters were found in pre-SMA, and in left anterior insula. Clusters were also found in right anterior insula and right anterior STS, just anterior to Heschl’s Gyrus. The two final clusters were located in pre-SMA and in right anterior MTG. All clusters were driven by at least two studies.

The network of areas activated for difficult speech comprehension appears remarkably symmetrical in both hemispheres and includes temporal, insular, and medial frontal areas (cf. Figure 1).

Meta-analysis 2 was based on the results of 21 experiments (Table III), 116 participants, and 473 foci published in 21 papers. One used PET, 20 used fMRI. Of the fMRI studies, nine used a sparse scanning paradigm, and 11 used a continuous paradigm. All experiments included a condition in which participants were required to produce speech and contrasted with a wide variety of baseline or other control conditions. Baseline stimuli included rest in the presence of scanner noise, rest in the absence of scanner noise, reading, covert speaking, observing (audiovisual) speech, listening to pink noise, and listening to speech. Ten of the 21 studies required participants to produce stimuli at sublexical levels (vowels, syllables or series of...
syllables, or pseudowords), while the remaining 11 required participants to produce speech stimuli at post-lexical levels (words, sentences, narratives/poems). Note that one study (Shuster, 2009) included contrasts for words > pink noise (Table 3) and pseudowords > pink noise (Table 2). Only one set of coordinates, i.e., pseudowords > pink noise, was included as not to over-represent foci from a single group of subjects and also to equalize the number of foci from studies investigating pre-lexical and post-lexical speech production as much as possible. The majority of studies used an event-related design (14), and a small number used a block design (seven). Finally, various tasks were used, including repeating words after auditory presentation, repeating phonemes after auditory presentation, responding to queries about personal experience or cite nursery rhymes or repeat a word list, producing syllables, reading aloud *Beowulf*, reading words after visual presentation, citing the months of the year, reading aloud pseudowords, reading aloud sentences, or performing a phonological verbal fluency task.

--- Insert Table III about here ---

The meta-analysis resulted in 12 clusters at the selected significance level (Table IV and Figure 1). As for meta-analysis 1, the pattern of ALE clusters was considerably more widespread on the left (14,672mm$^3$) than on the right (7,016mm$^3$). The first cluster was located in left pre-SMA and extended into SMA, while a second cluster was located in left Precentral Gyrus. A third cluster was found in right Lentiform Nucleus, extending medially into right Thalamus. Two right-lateralised clusters were in posterior STG/MTG and right Precentral Gyrus. A left-lateralised cluster was found in Lentiform Nucleus. Clusters were also found in left Thalamus, left Dentate Gyrus, left anterior STG, left Heschl’s Gyrus, left Precentral Gyrus, and finally in left anterior Insula extending laterally into left IFG (pars opercularis).
The network for speech production was thus largely located in the left hemisphere, and includes frontal and temporal cortical regions, as well as subcortical regions.

--- Insert Table IV about here ---

3.3 Overlap between speech comprehension and speech production

Figures 1 depicts the extent to which comprehension and production overlap (in purple). Overlap was found in left anterior STS, right anterior STS, pre-SMA and SMA. Meta-analysis 2 was repeated for the 10 studies in Table III involving the production of pre-lexical stimuli (vowels, syllables, pseudowords), and for the 11 studies involving the production of stimuli at post-lexical levels (words, sentences, poems/stories/narratives). These two sub-analyses were performed to ascertain whether the overlap between comprehension and production was between comprehension and the network for producing speech stimuli at post-lexical or at pre-lexical levels.

The ALE analysis (FDR, \( q<0.01 \), voxel wise, cluster extent 400mm\(^3\)) on the 10 pre-lexical production studies showed consistent activations in a network of eight areas, including left Lentiform Nucleus/Putamen, bilateral Thalamus, SMA, right Precentral Gyrus, left Cerebellum, and left Transverse Temporal Gyrus. The network for production of pre-lexical items thus includes mostly subcortical areas such as the Thalamus, Lentiform Nucleus, and Putamen, and well as the Cerebellum. Of these areas, only a very minor part of the cluster in SMA was also present (i.e., cluster #3 in Table II) in the network for difficult speech comprehension.

The ALE analysis on the 11 post-production studies showed consistent activations in a network of three areas, including pre-SMA, left STS, and right precentral gyrus. The network for producing post-lexical speech stimuli was less
extended than the network associated with producing pre-lexical stimuli, and also involved only cortical areas. Of these three areas, the clusters in pre-SMA (cluster #7 in Table II) and left STS (cluster #1 in Table II) were also present in the network for difficult speech comprehension. In sum, it appears that the network for speech comprehension shows overlap with the network for speech production. These results illustrate that the network for difficult speech comprehension includes areas also active during speech production at predominantly post-lexical levels.

--- Insert Table V and Figure 2 about here ---

4. Discussion
The present study used Activation Likelihood Analysis to localize the network of areas involved in difficult speech comprehension (meta-analysis 1) and in speech production (meta-analysis 2). Second, the study aimed to determine whether and to which extent the network for processing distorted speech overlaps with the speech production network.

4.1 Neural locus of comprehension of distorted speech (meta-analysis 1)
Meta-analysis 1 resulted in a description of areas that are consistently activated for difficult comprehension of intelligible speech (Figure 1). The results revealed a network that was remarkably symmetrical across both hemispheres, yet more substantial on the left. The network for difficult speech processing included anterior STS bilaterally, left posterior MTG, pre-SMA, the bilateral anterior insulae, and right posterior MTG.

It is unclear to what extent the network for difficult speech processing overlaps with the network of areas associated with comprehension of intelligible speech signals. Neuroimaging studies on processing (undistorted) intelligible speech report
activations in the majority of areas in Table II. Activity in left anterior STS has been widely reported (Adank & Devlin, 2010; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Dick, Saygin, Galati, Pitzalis, Bentrovato, D’Amico, et al., 2007; Friederici, Kotz, Scott, & Oleser, 2010; Leff et al., 2009; Oleser & Kotz, 2010; Oleser, Meyer, & Friederici, 2011; Oleser, et al., 2007; Rodd, Lange, Randall, & Tyler, 2010; Schon et al., 2010; Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006; Stevenson & James, 2009; Wilson, Molnar-Szakacs, & Iacoboni, 2008) and it has been proposed that the neural focus point of speech intelligibility processing is placed in left STS (Narain et al., 2003; Rauschecker & Scott, 2009; Scott, et al., 2000). Despite this claim, most of the studies reporting activity in left anterior STS also report involvement of right anterior STS (Adank & Devlin, 2010; Crinion, et al., 2003; Friederici, et al., 2010; Oleser & Kotz, 2010; Oleser, et al., 2007; Rodd, Davis, & Johnsrude, 2005; Rodd, et al., 2010; Stevenson & James, 2009; Wilson, et al., 2008). Finally, activity in left posterior MTG is also frequently reported in relation to intelligibility processing (Binder, Swanson, Hammeke, & Sabsevitz, 2008; Davis, Ford, Kherif, & Johnsrude, 2011; Davis & Johnsrude, 2003; Dick, Saygin, Galati, Pitzalis, Bentrovato, D’Amico, et al., 2007; Gonzalez-Castillo & Talavage, 2011; Oleser, Eisner, & Kotz, 2008; Okada et al., 2010; Rodd, et al., 2005).

Activations in pre-SMA have been reported in several studies that included an intelligibility contrast (Aleman et al., 2005; Binder, et al., 2008; Gonzalez-Castillo & Talavage, 2011; Jardri et al., 2007; Tyler et al., 2010; Wildgruber et al., 2004). It does not seem plausible that the activations in pre-SMA in these studies are due to task-related aspects (and associated button-presses), as three studies employed passive listening (Binder, et al., 2008; Gonzalez-Castillo & Talavage, 2011; Jardri, et al.,
2007), two used a task both in the speech condition and in the non-speech condition (Aleman, et al., 2005; Tyler, et al., 2010). Only Wildgruber et al. (2004) contrasted a speech condition with a task with a non-speech condition in which no task was used. It thus seems likely that the activation in pre-SMA is also part of the speech intelligibility processing network. SMA and pre-SMA have previously predominantly been associated with various aspects of the speech production process, including lexical selection, linear sequence encoding, and control of motor output (Alario, et al., 2006). Alario et al. proposed that this region is parcellated according to a rostrocaudal gradient, with lexical selection (Seifritz et al., 2006) in the most rostral/anterior aspect, and motor control in the caudal/posterior aspect. Therefore, it may be the case that increased activation of pre-SMA in noisy listening conditions reflects increased reliance on lexical selection processes. Few studies on intelligibility processing report activation in the left anterior insula (Binder, et al., 2008; Obleser, et al., 2011), the right anterior insula (Binder, et al., 2008; Ischebeck, Friederici, & Alter, 2008) or in right posterior MTG (Okada, et al., 2010; Rimol, Specht, & Hugdahl, 2006; Rodd, et al., 2005). It seems likely that both anterior insulae and bilateral posterior MTG are not part of a core network for processing intelligibility and represent areas additionally recruited under difficult challenging listening conditions.

The network for difficult speech comprehension partially overlaps with the network for pre-lexical speech processing as described in Turkeltaub & Coslett (2010). Turkeltaub & Coslett report activations in left posterior STG, left STG/STS, right MTG/STS and pre-SMA for the speech vs. non-speech contrast in the ALE-analysis listed in their Table 2. The network for difficult processing thus recruits temporal (bilateral STS) and frontal areas (pre-SMA) also involved in (pre-lexical) speech perception. However, Turkeltaub & Coslett do not report the activations in
(deep) frontal areas in the anterior insulae reported in the present analysis. It seems plausible that these activations are associated with increased attentional and/or working memory processes as proposed in a recent meta-analysis (Vigneau et al., 2011). Yet, the fact that these areas are also present in the network of difficult speech comprehension suggests that difficult comprehension relies in part on increased involvement of general cognitive processes, as proposed by Holt and Lotto (2008).

4.2 Neural locus of speech production (meta-analysis 2)

The network consistently activated for speech production appears more extensive on the left, and includes frontal and temporal cortical regions including pre-SMA and SMA, right posterior STG/MTG, left Precentral Gyrus, left Heschl’s Gyrus, and left IFG (part opercularis), as well as subcortical regions including right Thalamus, and Cerebellum. Two sub-analyses showed that the subcortical activations in the network may be driven mostly by the inclusion of studies in which participants produced pre-lexical speech stimuli, whereas producing post-lexical stimuli activates areas in left STS, Precentral Gyrus, and pre-SMA and SMA.

The network for speech production reported in the present study shows considerable overlap with the network for single word production in Turkeltaub et al. (2002). Turkeltaub et al. report ALE clusters in bilateral Precentral Gyrus, bilateral STS (left anterior and posterior, right posterior), posterior STG, left Fusiform Gyrus, left Thalamus, (right) pre-SMA, and bilateral Cerebellum. The sub-analysis on the production of post-lexical stimuli found clusters in pre-SMA, left STS and right Precentral Gyrus. Methodological and statistical (such as the present’s paper strict significance levels) differences most likely underlie any differences between the two meta-analyses. Note that Turkeltaub et al. included only PET studies on single-word reading, whereas the present study on the 11 papers that used post-lexical stimuli
included 10 fMRI studies and one PET study and comprised of studies in which participants produced a wider range of speech stimuli, also including sentences and longer stretches of speech. Nevertheless, results for the present study together with Turkeltaub et al.’s results converge on a core network for post-lexical speech production that includes pre-SMA, Precentral Gyrus, and anterior STS. Further study is required to determine the effects of neuroimaging technique and stimulus material on the inclusion or exclusion of specific brain areas outside the core network.

It seems unlikely that activation in left STS related to producing speech can be entirely explained by the presence of auditory feedback during speech production, as no activation was found in anterior temporal regions when pre-lexical speech production was assessed separately. Instead, it appears that producing intelligible speech involves access to semantic processing, as does comprehension of intelligible speech in the absence and presence of distortions of the acoustic signal.

4.3 Neural overlap between speech production and speech comprehension

Difficult speech comprehension and speech production overlapped in bilateral anterior STS and pre-SMA. Repeating meta-analysis 2 for studies using pre-lexical stimuli and those using post-lexical stimuli revealed that activations related to difficult speech processing overlapped predominantly with the network associated with post-lexical speech production. In section 4.1 it was argued that pre-SMA and bilateral anterior STS are involved in intelligibility processing. This implies that difficult speech comprehension, intelligibility processing, and speech production all activate a small network of frontal and temporal regions, indicating that perception and production of speech - at least in part - rely on a shared network of areas.

4.4 Implications for speech processing models
Three mechanisms for effective processing of distorted speech have been proposed: difficult comprehension is resolved by general auditory mechanisms with the involvement of general cognitive mechanisms (Holt & Lotto, 2008), difficult comprehension relies on auditory mechanisms and especially recruited speech production mechanisms (Hickok & Poeppel, 2007; Pickering & Garrod, 2007; Skipper, et al., 2006), and difficult speech comprehension relies nearly entirely on speech motor mechanisms (Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Whalen, et al., 2006). The results of the present study indicate that difficult speech comprehension first leads to increased reliance of cortical regions involved in production and comprehension processes (bilateral anterior STS and pre-SMA), increased activation in an area associated with speech intelligibility processing (left posterior MTG) and second involves increased reliance on cortical areas associated with general executive processes - such as working memory - (bilateral anterior insulae). The results therefore support a hybrid neural mechanism for processing distorted speech that combines elements from the general auditory approaches (Holt & Lotto, 2008) and speech motor involvement (e.g., (Skipper, et al., 2006).

Yet, the results do not support the proposed critical role of left IFG in processing distorted speech (Davis & Johnsrude, 2003; Peelle, Johnsrude, et al., 2010). No evidence was found of involvement of left IFG in processing distorted speech signals in meta-analysis 1. Left IFG played a (small) role in speech production, but was not found to be one of the cortical areas displaying overlap between comprehension and production. Left IFG has frequently been associated with effective speech comprehension. For instance, patient studies show that left IFG lesion have been associated with decreased ability to understand distorted speech (Moineau, Dronkers, & Bates, 2005) and with compromised word recognition (Utman, Blumstein, &
Sullivan, 2001). Nevertheless, speech processing in individuals with a lesion may not be representative of speech processing in the healthy individuals included in the present meta-analyses. Also, lesions associated with left inferior frontal areas tend to be quite large and may extend to superior frontal, temporal, and parietal lobes (e.g., Dronkers, Redfern, & Knight, 2000). Finally, a recent study found no evidence that lesions in left Broca’s area (i.e., left Brodmann Areas 44 and 45) negatively impact language comprehension (Dronkers, Wilkins, Van Valin Jr., Redfern, & Jaeger, 2004).

Prominent models for speech processing do generally not propose neural mechanisms subserving effective processing of distorted speech signals (cf. Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Models for the neural architecture of speech processing could account for difficult speech processing by incorporating the following mechanism. First, comprehension of distorted speech leads behaviourally to more effortful processing and increased associated cognitive load (cf. Adank, et al., 2009). This higher load leads to increased activation in areas associated with speech intelligibility processing, specifically bilateral anterior STS and left posterior MTG. Second, the higher cognitive load may also lead to increased activation in areas associated with general cognitive processing, such as the bilateral anterior insulae. Finally, processing distorted speech may lead to increased activation in areas shared between comprehension and production processes, such as bilateral anterior STS and pre-SMA. Note that the presents results cannot inform about causal or hierarchical relationships between aforementioned cortical areas. This last issue could be approached using functional and structural connectivity studies on degraded speech signals, using an approach used by Saur, Schelte, Schnell, Kratochvil, Küpper, et al. (2010).
4.5 Conclusion

Analysis of the results from the two meta-analyses and their overlap leads to the conclusion that processing of distorted speech specifically recruits areas involved in general cognitive processing, such as the anterior insulae, and areas involved in speech production, such as pre-SMA and bilateral anterior STS, but does not involve left IFG. This suggests that the mechanism governing the successful understanding of others in difficult listening conditions, including background noise, signal degradation, or accented speech, combines an increased reliance on general cognitive processing with increased involvement of resources shared between speech comprehension and speech production.

References


Crescentini, C., Shallice, T., & Macaluso, E. Item retrieval and competition in noun and verb generation: an FMRI study. *Journal of Cognitive Neuroscience, 22*(6), 1140-1157.


syllabic sequences in the speech production network. *NeuroImage*, 50(2), 626-638.


Figure captions

Figure 1. Areas activated by processing of distorted speech signals (meta-analysis 1, in blue) with areas activated during speech production (meta-analysis 2, in red), and their overlap (mauve).

Figure 2. Areas activated by processing of distorted speech signals (meta-analysis 1, in blue) with areas activated during pre-lexical speech production (meta-analysis 2, in red), areas activated during post-lexical speech production (green), and overlap between speech comprehension and post-lexical speech production (turquoise).

Patti Adank

School of Psychological Sciences, University of Manchester, Manchester, United Kingdom

Short title: Mechanisms of Effortful Speech Understanding

Correspondence:
Patti Adank
School of Psychological Sciences
University of Manchester
Zochonis Building
Brunswick Street
M20 6HL
Manchester, UK
Email: patti.adank@manchester.ac.uk
Phone: 0044-161-2752693

Keywords: perception, auditory, speech, articulation, production, fMRI, PET, neuroimaging, meta-analysis.
Abstract

The role of speech production mechanisms in difficult speech comprehension is the subject of on-going debate in speech science. Two Activation Likelihood Estimation (ALE) analyses were conducted on neuroimaging studies investigating difficult speech comprehension or speech production. Meta-analysis 1 included 10 studies contrasting comprehension of less intelligible/distorted speech with more intelligible speech. Meta-analysis 2 (21 studies) identified areas associated with speech production. The results indicate that difficult comprehension involves increased reliance of cortical regions in which comprehension and production overlapped (bilateral anterior Superior Temporal Sulcus (STS) and anterior Supplementary Motor Area (pre-SMA)) and in an area associated with intelligibility processing (left posterior MTG), and second involves increased reliance on cortical areas associated with general executive processes (bilateral anterior insulae). Comprehension of distorted speech may be supported by a hybrid neural mechanism combining increased involvement of areas associated with general executive processing and areas shared between comprehension and production.
1. Introduction
The human speech comprehension system is remarkable in its ability to quickly extract the linguistic message from a transient acoustic signal. Much of everyday processing occurs under listening conditions that are less than ideal, due to background noise, regional accents, or speech rate differences, to name a few common everyday variations in the speech signal. Listeners are generally able to successfully comprehend speech under such adverse - or difficult - listening conditions. Nevertheless, speech comprehension is often slower and less efficient than under less difficult conditions (see Mattys, Brooks, & Cooke, 2009, for an overview). For instance, when performing a semantic verification task (i.e., reporting whether a sentence such as ‘dogs have four ears’ is true or false) spoken in an unfamiliar regional accent, listeners show slower response times and higher error scores (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009). However, the neural mechanisms supporting the intrinsic robustness of the speech comprehension system are largely unclear.

Cognitive neuroscience models of the cortical organisation of spoken language processing (e.g., Hickok & Poeppel, 2007) have not made explicit predictions regarding the neural locus of processing of distorted speech signals. However, this neural locus has been discussed in more detail in various papers investigating difficult speech processing (Davis & Johnsrude, 2003; Peelle, Johnsrude, & Davis, 2010). Davis & Johnsrude and Peelle et al. propose a critical role for left Inferior Frontal Gyrus in processing of distorted speech signals. This proposed role of IFG is motivated by its frequent activation during speech perception and comprehension tasks (e.g., Crinion & Price, 2005; Obleser, Wise, Dresner, & Scott, 2007). Second, it is argued that IFG’s anatomical connectivity to auditory belt and parabelt regions
(e.g., Hackett, Stepniewska, & Kaas, 1999) makes it well-positioned to affect processing in primary auditory and association areas traditionally associated with speech perception. Finally, Peelle et al. argue that Davis and Johnsrude (2003) provide direct evidence a role of IFG in processing distorted speech signals by showing that activity in left IFG was elevated for distorted (but still intelligible) speech compared to both clear speech and unintelligible noise.

In speech science, three general behavioural/neural mechanisms have been suggested to support difficult speech comprehension. First, it has been hypothesised that comprehension relies predominantly on auditory processes and associated brain areas (Holt & Lotto, 2008). Processing distortions of the speech signal is predicted to be governed through involvement of general cognitive processes, such as working memory and/or attention. Second, it has been proposed that comprehension of distorted speech signals recruits neural mechanisms associated with speech production (Hickok & Poeppel, 2007; Pickering & Garrod, 2007; Skipper, Nusbaum, & Small, 2006). This line of reasoning proposes that speech processing in the absence of external distortions relies predominantly on auditory processes, while speech production processes are selectively active when listening conditions deteriorate. Third, it has been put forward that auditory speech processing relies almost entirely on speech motor mechanisms, with only a minor role for auditory (or general cognitive) mechanisms (Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Whalen et al., 2006). Here, the auditory signal is relayed through speech production mechanisms to achieve successful comprehension regardless of the listening conditions.

A number of functional Magnetic Resonance Imaging (fMRI) studies support the view that comprehension of distorted speech relies on the recruitment of speech
production mechanisms. Functional MRI studies investigating processing of distorted speech commonly present listeners with speech stimuli perturbed by adding background noise or multi-speaker babble (Indefrey & Levelt, 2004; Stowe et al., 1998), by passing the speech signal through a noise-vocoder (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), artificially time-compressing the signal (Dupoux & Green, 1997), or by using speech stimuli that have been spoken in an unfamiliar foreign or regional accent (Adank, et al., 2009; Munro & Derwing, 1995). Neural responses associated with processing distorted speech stimuli are then contrasted with more intelligible – undistorted – speech stimuli. Usually, activity related to processing distorted speech is found across a variety of cortical areas, including posterior superior and middle temporal areas bilaterally, left inferior frontal areas, left and right frontal opercula, precentral gyrus bilaterally extending to the Supplementary Motor Area (SMA), and subcortical areas including Thalamus. For instance, Davis & Johnsrude (2003) report activations in temporal and inferior frontal areas, as well as in premotor areas in their Figure 7. The activations in IFG and motor areas including precentral gyrus and SMA are also associated with speech production (Alario, Chainay, Lehericy, & Cohen, 2006; Levelt, 1989; Wilson, Saygin, Sereno, & Iacoboni, 2004), have been reported to be active during speech comprehension (Adank & Devlin, 2010; Wilson, et al., 2004). Furthermore, these areas are considered to be an integral part of the speech production network, as demonstrated in a previous meta-analysis on Positron Emission Tomography (PET) studies on single word reading (Turkeltaub, Eden, Jones, & Zeffiro, 2002). However, it is unclear to which degree specific speech motor areas are consistently activated during comprehension of distorted speech.
The present study aims to, first, identify the network of areas involved in processing distorted speech signals. Second, it aims to determine whether the network for difficult speech comprehension includes areas involved in speech production. If difficult speech comprehension consistently activates areas also involved in speech production, then this would support for the hypothesis that speech production areas are selectively recruited during speech comprehension.

This paper presents two meta-analyses using the Activation Likelihood Estimation (ALE) method (Laird et al., 2005; Turkeltaub, et al., 2002). ALE has been used to determine the overlap between coordinates obtained from neuroimaging studies by modelling them as probability distributions that are centred at the reported coordinates. The first meta-analysis applies ALE to coordinates extracted from studies contrasting difficult speech comprehension with comprehension of intelligible speech under less adverse listening conditions. This analysis aims to identify the areas consistently activated during successful but difficult speech comprehension. The second meta-analysis applies ALE to coordinates extracted from studies that include conditions in which participants produce speech at pre-lexical (e.g., individual speech sounds, syllables) and post-lexical levels (e.g., words, sentences) that are contrasted with conditions in which participants do not produce speech. The aim of the second analysis is to identify areas consistently activated during speech production. This second meta-analysis is intended to represent an extension to the meta-analysis described in Turkeltaub et al. (2002). It was decided to perform a new meta-analysis on the neural network involved in speech production, as Turkeltaub et al. included only PET studies. Recent neuroimaging studies predominantly use fMRI for studying auditory processing of speech, but also for speech production. In addition, Turkeltaub et al. focused on studies that addressed single word reading. The present research also
includes studies that investigate production of pre-lexical linguistic elements, to identify the network of speech production involved in the articulation of pre- and post-lexical speech stimuli, and to avoid skewing the results of the ALE towards areas involved in production of linguistically meaningful elements only.

2. Method

2.1.1 Selection of literature studies for meta-analysis 1

Neuroimaging studies were included that investigated comprehension of distorted (yet intelligible) speech at post-lexical levels. The PubMed online database for was searched for studies using the keywords: “distorted”, “degraded”, “dialect”, “accent”, “sine wave”, “synthetic”, “noise”, “time-compressed”, “noise-vocoded”, “speech”, “intelligibility”, “intelligible”, “comprehension”, “fMRI”, “post-lexical”, “narrative”, “word”, “PET”, “neuroimaging”, and appropriate combinations of these keywords. Additional papers were collected by searching for prominent researchers in the field. Papers were selected from January 1 2001 onwards, including papers in press or in advance online publication.

2.1.2 Inclusion and exclusion criteria for meta-analysis 1

Papers were included that fulfilled the following criteria: i) neural responses were collected using fMRI or PET, ii) only healthy, adult, neurotypical subjects with intact hearing and no known neurological or psychiatric disorders were tested, iii) the experiments contained conditions in which less intelligible speech as well as conditions in which more intelligible speech was presented, iv) speech stimuli were words, sentences, or narratives; v) stimuli were naturally spoken utterances and not synthetic utterances, vi) results were reported at a group-level in a stereotactic 3-coordinate system. In addition, the following criteria were used to exclude papers
from the analysis: single subject studies and studies that report only the results from a pre-specified region-of-interest (ROI). The selected studies are listed in Table I.

A single study included two different distortions (Adank, Davis, & Hagoort, in press) – speech in an unfamiliar accent and speech in a familiar accent with added background noise – and contrasted these with undistorted speech in quiet in a familiar accent. It was decided to include only the coordinate from the contrast involving speech in an unfamiliar accent, as only one other study (Adank, Noordzij, & Hagoort, 2012) used speech in an unfamiliar accent, while two other studies (Davis & Johnsrude, 2003; Wong, Uppanda, Parrish, & Dhar, 2008) used added background noise to distort the speech stimuli (note that Davis et al. also used noise-vocoded and noise-segmented speech).

2.2.1 Selection of literature studies for meta-analysis 2

Neuroimaging studies were included that investigated speech production using PubMed. The PubMed online database for was searched for studies using the keywords: “speech”, “production”, “articulation”, “syllable”, “phoneme”, “word”, “fmri”, “PET”, “neuroimaging”, and appropriate combinations of these keywords. In addition, papers were identified by searching for prominent researchers in the field. As for analysis 1, papers were selected from January 1 2001 onwards, including papers in press or in advance online publication.

2.2.2 Inclusion and exclusion criteria for meta-analysis 1

Papers were included that fulfilled the following criteria: i) neural responses were collected using fMRI or PET, ii) only healthy, adult, right-handed neurotypical subjects with intact hearing and no known neurological or psychiatric disorders were tested, iii) the experiments contained conditions in which participants produced phonemes, syllables, words, sentences, or narratives; iv) results were reported at a
group-level in a stereotactic 3-coordinate system. Again, single subject studies and studies that report only the results from a pre-specified region-of-interest (ROI) were excluded. The selected studies are listed in Table III.

2.3 ALE methods

The ALE analysis was implemented using GingerALE 2.04 (www.brainmap.org). This version of the likelihood estimation algorithm was selected as it has been shown to be more precise than previous versions, while it retains comparable sensitivity (Eickhoff, Laird, Grefkes, Wang, Zilles, et al, 2009). Coordinates collected from studies that reported coordinates in Talairach space were converted to MNI space using the tal2icbm_spm algorithm implemented in the GingerALE software (www.brainmap.org/ale).

In GingerALE, first, modelled activation maps are computed for each set of foci per included study. All foci were modelled as Gaussian distributions and merged into a single 3-dimensional volume. GingerALE uses an uncertainty modelling algorithm to empirically estimate the between-subjects and between-templates variability of all included foci sets. Second, ALE values are computed on a voxel-to-voxel basis by taking the values that are common to the individual modelled activation maps. GingerALE constrains the limits of this analysis to a grey matter mask that was used to define the outer limits of MNI coordinate space, which excludes most white-matter structures (Eickhoff, Heim, Zilles, & Amunts, 2009). The analysis was corrected for multiple comparisons using the FDR (false discovery rate) method at $q < 0.01$, voxel wise, (default = 0.05), using a cluster extent of 400mm$^3$ (default = 200mm$^3$). The Mango software package (http://ric.uthscsa.edu/mango/) was used to view the
resulting activation maps and all results were overlaid on a single MNI template available in Mango (Colin27_T1_seg_MNI.nii).

3. Results

3.1 Meta-analysis 1

Meta-analysis 1 was based on the results of 10 experiments (Table I), 116 participants, and 75 foci that were published in 10 papers. One used a PET design, nine used fMRI, five used a sparse scanning paradigm, and four used a continuous paradigm. In all experiments, a condition in which listeners were required to comprehend distorted speech was included. Several types of distortions were used: added background noise (e.g., in Wong, et al., 2008), speech in an unfamiliar accent (Adank, et al., 2012), artificially time-compressed speech (see Dupoux & Green, 1997, for a description of this specific distortion type and Poldrack et al., 2001, for a study including time-compressed speech). Some used noise-vocoded speech (for example Sharp et al., 2010), a single study included segmented speech (Davis & Johnsrude, 2003, see Schroeder, 1968, for a description of this type of distortion). Finally, one study (Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010) contrasted continuous scanning using standard EPI noise with a “quiet” EPI sequence. This “quiet” sequence minimises the acoustic disturbance associated with traditional EPI using the same imaging parameters (Schmitter et al., 2008). Stimuli were either words or sentences. Nine studies used an experimental task, and one study used passive listening plus an after-task (Adank, et al., 2012). The following experimental tasks were employed: speaker judgment, semantic decision, intelligibility judgment, gender decision, grammatical decision, rhyme decision, target matching, and target to picture matching.
The meta-analysis resulted in eight clusters at the selected significance level (Table II). These clusters show a similar pattern across both hemispheres, but the pattern of ALE clusters was more widespread on the left (3840mm$^3$, excluding clusters 3 and 7, as they are centred at x = 0) than on the right (1760mm$^3$, excluding clusters 3 and 7). The highest ALE score was found for a cluster in left STS, located just anterior to Heschl’s Gyrus. A second cluster was found in posterior left Middle Temporal Gyrus (MTG). Subsequent clusters were found in pre-SMA, and in left anterior insula. Clusters were also found in right anterior insula and right anterior STS, just anterior to Heschl’s Gyrus. The two final clusters were located in pre-SMA and in right anterior MTG. All clusters were driven by at least two studies.

The network of areas activated for difficult speech comprehension appears remarkably symmetrical in both hemispheres and includes temporal, insular, and medial frontal areas (cf. Figure 1).

3.2 Meta-analysis 2

Meta-analysis 2 was based on the results of 21 experiments (Table III), 116 participants, and 473 foci published in 21 papers. One used PET, 20 used fMRI. Of the fMRI studies, nine used a sparse scanning paradigm, and 11 used a continuous paradigm. All experiments included a condition in which participants were required to produce speech and contrasted with a wide variety of baseline or other control conditions. Baseline stimuli included rest in the presence of scanner noise, rest in the absence of scanner noise, reading, covert speaking, observing (audiovisual) speech, listening to pink noise, and listening to speech. Ten of the 21 studies required participants to produce stimuli at sublexical levels (vowels, syllables or series of
syllables, or pseudowords), while the remaining 11 required participants to produce speech stimuli at post-lexical levels (words, sentences, narratives/poems). Note that one study (Shuster, 2009) included contrasts for words > pink noise (Table 3) and pseudowords > pink noise (Table 2). Only one set of coordinates, i.e., pseudowords > pink noise, was included as not to over-represent foci from a single group of subjects and also to equalize the number of foci from studies investigating pre-lexical and post-lexical speech production as much as possible. The majority of studies used an event-related design (14), and a small number used a block design (seven). Finally, various tasks were used, including repeating words after auditory presentation, repeating phonemes after auditory presentation, responding to queries about personal experience or cite nursery rhymes or repeat a word list, producing syllables, reading aloud *Beowulf*, reading words after visual presentation, citing the months of the year, reading aloud pseudowords, reading aloud sentences, or performing a phonological verbal fluency task.

--- Insert Table III about here ---

The meta-analysis resulted in 12 clusters at the selected significance level (Table IV and Figure 1). As for meta-analysis 1, the pattern of ALE clusters was considerably more widespread on the left (14,672mm$^3$) than on the right (7,016mm$^3$). The first cluster was located in left pre-SMA and extended into SMA, while a second cluster was located in left Precentral Gyrus. A third cluster was found in right Lentiform Nucleus, extending medially into right Thalamus. Two right-lateralised clusters were in posterior STG/MTG and right Precentral Gyrus. A left-lateralised cluster was found in Lentiform Nucleus. Clusters were also found in left Thalamus, left Dentate Gyrus, left anterior STG, left Heschl’s Gyrus, left Precentral Gyrus, and finally in left anterior Insula extending laterally into left IFG (pars opercularis).
The network for speech production was thus largely located in the left hemisphere, and includes frontal and temporal cortical regions, as well as subcortical regions.

--- Insert Table IV about here ---

3.3 Overlap between speech comprehension and speech production

Figures 1 depicts the extent to which comprehension and production overlap (in purple). Overlap was found in left anterior STS, right anterior STS, pre-SMA and SMA. Meta-analysis 2 was repeated for the 10 studies in Table III involving the production of pre-lexical stimuli (vowels, syllables, pseudowords), and for the 11 studies involving the production of stimuli at post-lexical levels (words, sentences, poems/stories/narratives). These two sub-analyses were performed to ascertain whether the overlap between comprehension and production was between comprehension and the network for producing speech stimuli at post-lexical or at pre-lexical levels.

The ALE analysis (FDR, q<0.01, voxel wise, cluster extent 400mm$^3$) on the 10 pre-lexical production studies showed consistent activations in a network of eight areas, including left Lentiform Nucleus/Putamen, bilateral Thalamus, SMA, right Precentral Gyrus, left Cerebellum, and left Transverse Temporal Gyrus. The network for production of pre-lexical items thus includes mostly subcortical areas such as the Thalamus, Lentiform Nucleus, and Putamen, and well as the Cerebellum. Of these areas, only a very minor part of the cluster in SMA was also present (i.e., cluster #3 in Table II) in the network for difficult speech comprehension.

The ALE analysis on the 11 post-production studies showed consistent activations in a network of three areas, including pre-SMA, left STS, and right precentral gyrus. The network for producing post-lexical speech stimuli was less
extended than the network associated with producing pre-lexical stimuli, and also involved only cortical areas. Of these three areas, the clusters in pre-SMA (cluster #7 in Table II) and left STS (cluster #1 in Table II) were also present in the network for difficult speech comprehension. In sum, it appears that the network for speech comprehension shows overlap with the network for speech production. These results illustrate that the network for difficult speech comprehension includes areas also active during speech production at predominantly post-lexical levels.

--- Insert Table V and Figure 2 about here ---

4. Discussion

The present study used Activation Likelihood Analysis to localize the network of areas involved in difficult speech comprehension (meta-analysis 1) and in speech production (meta-analysis 2). Second, the study aimed to determine whether and to which extent the network for processing distorted speech overlaps with the speech production network.

4.1 Neural locus of comprehension of distorted speech (meta-analysis 1)

Meta-analysis 1 resulted in a description of areas that are consistently activated for difficult comprehension of intelligible speech (Figure 1). The results revealed a network that was remarkably symmetrical across both hemispheres, yet more substantial on the left. The network for difficult speech processing included anterior STS bilaterally, left posterior MTG, pre-SMA, the bilateral anterior insulae, and right posterior MTG.

It is unclear to what extent the network for difficult speech processing overlaps with the network of areas associated with comprehension of intelligible speech signals. Neuroimaging studies on processing (undistorted) intelligible speech report
activations in the majority of areas in Table II. Activity in left anterior STS has been widely reported (Adank & Devlin, 2010; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Dick, Saygin, Galati, Pitzalis, Bentrovato, D’Amico, et al., 2007; Friederici, Kotz, Scott, & Obleser, 2010; Leff et al., 2009; Obleser & Kotz, 2010; Obleser, Meyer, & Friederici, 2011; Obleser, et al., 2007; Rodd, Longe, Randall, & Tyler, 2010; Schon et al., 2010; Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006; Stevenson & James, 2009; Wilson, Molnar-Szakacs, & Iacoboni, 2008) and it has been proposed that the neural focus point of speech intelligibility processing is placed in left STS (Narain et al., 2003; Rauschecker & Scott, 2009; Scott, et al., 2000). Despite this claim, most of the studies reporting activity in left anterior STS also report involvement of right anterior STS (Adank & Devlin, 2010; Crinion, et al., 2003; Friederici, et al., 2010; Obleser & Kotz, 2010; Obleser, et al., 2007; Rodd, Davis, & Johnsruede, 2005; Rodd, et al., 2010; Stevenson & James, 2009; Wilson, et al., 2008). Finally, activity in left posterior MTG is also frequently reported in relation to intelligibility processing (Binder, Swanson, Hammeke, & Sabsevitz, 2008; Davis, Ford, Kherif, & Johnsruede, 2011; Davis & Johnsruede, 2003; Dick, Saygin, Galati, Pitzalis, Bentrovato, D’Amico, et al., 2007; Gonzalez-Castillo & Talavage, 2011; Obleser, Eisner, & Kotz, 2008; Okada et al., 2010; Rodd, et al., 2005).

Activations in pre-SMA have been reported in several studies that included an intelligibility contrast (Aleman et al., 2005; Binder, et al., 2008; Gonzalez-Castillo & Talavage, 2011; Jardri et al., 2007; Tyler et al., 2010; Wildgruber et al., 2004). It does not seem plausible that the activations in pre-SMA in these studies are due to task-related aspects (and associated button-presses), as three studies employed passive listening (Binder, et al., 2008; Gonzalez-Castillo & Talavage, 2011; Jardri, et al.,
2007), two used a task both in the speech condition and in the non-speech condition (Aleman, et al., 2005; Tyler, et al., 2010). Only Wildgruber et al. (2004) contrasted a speech condition with a task with a non-speech condition in which no task was used. It thus seems likely that the activation in pre-SMA is also part of the speech intelligibility processing network. SMA and pre-SMA have previously predominantly been associated with various aspects of the speech production process, including lexical selection, linear sequence encoding, and control of motor output (Alario, et al., 2006). Alario et al. proposed that this region is parcellated according to a rostrocaudal gradient, with lexical selection (Seifritz et al., 2006) in the most rostral/anterior aspect, and motor control in the caudal/posterior aspect. Therefore, it may be the case that increased activation of pre-SMA in noisy listening conditions reflects increased reliance on lexical selection processes. Few studies on intelligibility processing report activation in the left anterior insula (Binder, et al., 2008; Obleser, et al., 2011), the right anterior insula (Binder, et al., 2008; Ischebeck, Friederici, & Alter, 2008) or in right posterior MTG (Okada, et al., 2010; Rimol, Specht, & Hugdahl, 2006; Rodd, et al., 2005). It seems likely that both anterior insulae and bilateral posterior MTG are not part of a core network for processing intelligibility and represent areas additionally recruited under difficult challenging listening conditions.

The network for difficult speech comprehension partially overlaps with the network for pre-lexical speech processing as described in Turkeltaub & Coslett (2010). Turkeltaub & Coslett report activations in left posterior STG, left STG/STS, right MTG/STS and pre-SMA for the speech vs. non-speech contrast in the ALE-analysis listed in their Table 2. The network for difficult processing thus recruits temporal (bilateral STS) and frontal areas (pre-SMA) also involved in (pre-lexical) speech perception. However, Turkeltaub & Coslett do not report the activations in
(deep) frontal areas in the anterior insulae reported in the present analysis. It seems plausible that these activations are associated with increased attentional and/or working memory processes as proposed in a recent meta-analysis (Vigneau et al., 2011). Yet, the fact that these areas are also present in the network of difficult speech comprehension suggests that difficult comprehension relies in part on increased involvement of general cognitive processes, as proposed by Holt and Lotto (2008).

4.2 Neural locus of speech production (meta-analysis 2)

The network consistently activated for speech production appears more extensive on the left, and includes frontal and temporal cortical regions including pre-SMA and SMA, right posterior STG/MTG, left Precentral Gyrus, left Heschl’s Gyrus, and left IFG (part opercularis), as well as subcortical regions including right Thalamus, and Cerebellum. Two sub-analyses showed that the subcortical activations in the network may be driven mostly by the inclusion of studies in which participants produced pre-lexical speech stimuli, whereas producing post-lexical stimuli activates areas in left STS, Precentral Gyrus, and pre-SMA and SMA.

The network for speech production reported in the present study shows considerable overlap with the network for single word production in Turkeltaub et al. (2002). Turkeltaub et al. report ALE clusters in bilateral Precentral Gyrus, bilateral STS (left anterior and posterior, right posterior), posterior STG, left Fusiform Gyrus, left Thalamus, (right) pre-SMA, and bilateral Cerebellum. The sub-analysis on the production of post-lexical stimuli found clusters in pre-SMA, left STS and right Precentral Gyrus. Methodological and statistical (such as the present’s paper strict significance levels) differences most likely underlie any differences between the two meta-analyses. Note that Turkeltaub et al. included only PET studies on single-word reading, whereas the present study on the 11 papers that used post-lexical stimuli
included 10 fMRI studies and one PET study and comprised of studies in which participants produced a wider range of speech stimuli, also including sentences and longer stretches of speech. Nevertheless, results for the present study together with Turkeltaub et al.’s results converge on a core network for post-lexical speech production that includes pre-SMA, Precentral Gyrus, and anterior STS. Further study is required to determine the effects of neuroimaging technique and stimulus material on the inclusion or exclusion of specific brain areas outside the core network.

It seems unlikely that activation in left STS related to producing speech can be entirely explained by the presence of auditory feedback during speech production, as no activation was found in anterior temporal regions when pre-lexical speech production was assessed separately. Instead, it appears that producing intelligible speech involves access to semantic processing, as does comprehension of intelligible speech in the absence and presence of distortions of the acoustic signal.

4.3 Neural overlap between speech production and speech comprehension

Difficult speech comprehension and speech production overlapped in bilateral anterior STS and pre-SMA. Repeating meta-analysis 2 for studies using pre-lexical stimuli and those using post-lexical stimuli revealed that activations related to difficult speech processing overlapped predominantly with the network associated with post-lexical speech production. In section 4.1 it was argued that pre-SMA and bilateral anterior STS are involved in intelligibility processing. This implies that difficult speech comprehension, intelligibility processing, and speech production all activate a small network of frontal and temporal regions, indicating that perception and production of speech - at least in part - rely on a shared network of areas.

4.4 Implications for speech processing models
Three mechanisms for effective processing of distorted speech have been proposed: difficult comprehension is resolved by general auditory mechanisms with the involvement of general cognitive mechanisms (Holt & Lotto, 2008), difficult comprehension relies on auditory mechanisms and especially recruited speech production mechanisms (Hickok & Poeppel, 2007; Pickering & Garrod, 2007; Skipper, et al., 2006), and difficult speech comprehension relies nearly entirely on speech motor mechanisms (Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Whalen, et al., 2006). The results of the present study indicate that difficult speech comprehension first leads to increased reliance of cortical regions involved in production and comprehension processes (bilateral anterior STS and pre-SMA), increased activation in an area associated with speech intelligibility processing (left posterior MTG) and second involves increased reliance on cortical areas associated with general executive processes - such as working memory - (bilateral anterior insulae). The results therefore support a hybrid neural mechanism for processing distorted speech that combines elements from the general auditory approaches (Holt & Lotto, 2008) and speech motor involvement (e.g., (Skipper, et al., 2006).

Yet, the results do not support the proposed critical role of left IFG in processing distorted speech (Davis & Johnsrude, 2003; Peelle, Johnsrude, et al., 2010). No evidence was found of involvement of left IFG in processing distorted speech signals in meta-analysis 1. Left IFG played a (small) role in speech production, but was not found to be one of the cortical areas displaying overlap between comprehension and production. Left IFG has frequently been associated with effective speech comprehension. For instance, patient studies show that left IFG lesion have been associated with decreased ability to understand distorted speech (Moineau, Dronkers, & Bates, 2005) and with compromised word recognition (Utman, Blumstein, &
Sullivan, 2001). Nevertheless, speech processing in individuals with a lesion may not be representative of speech processing in the healthy individuals included in the present meta-analyses. Also, lesions associated with left inferior frontal areas tend to be quite large and may extend to superior frontal, temporal, and parietal lobes (e.g., Dronkers, Redfern, & Knight, 2000). Finally, a recent study found no evidence that lesions in left Broca’s area (i.e., left Brodmann Areas 44 and 45) negatively impact language comprehension (Dronkers, Wilkins, Van Valin Jr., Redfern, & Jaeger, 2004).

Prominent models for speech processing do generally not propose neural mechanisms subserving effective processing of distorted speech signals (cf. Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Models for the neural architecture of speech processing could account for difficult speech processing by incorporating the following mechanism. First, comprehension of distorted speech leads behaviourally to more effortful processing and increased associated cognitive load (cf. Adank, et al., 2009). This higher load leads to increased activation in areas associated with speech intelligibility processing, specifically bilateral anterior STS and left posterior MTG. Second, the higher cognitive load may also lead to increased activation in areas associated with general cognitive processing, such as the bilateral anterior insulae. Finally, processing distorted speech may lead to increased activation in areas shared between comprehension and production processes, such as bilateral anterior STS and pre-SMA. Note that the presents results cannot inform about causal or hierarchical relationships between aforementioned cortical areas. This last issue could be approached using functional and structural connectivity studies on degraded speech signals, using an approach used by Saur, Schelte, Schnell, Kratochvil, Küpper, et al. (2010).
4.5 Conclusion

Analysis of the results from the two meta-analyses and their overlap leads to the conclusion that processing of distorted speech specifically recruits areas involved in general cognitive processing, such as the anterior insulae, and areas involved in speech production, such as pre-SMA and bilateral anterior STS, but does not involve left IFG. This suggests that the mechanism governing the successful understanding of others in difficult listening conditions, including background noise, signal degradation, or accented speech, combines an increased reliance on general cognitive processing with increased involvement of resources shared between speech comprehension and speech production.

References


Crescentini, C., Shallice, T., & Macaluso, E. Item retrieval and competition in noun and verb generation: an FMRI study. *Journal of Cognitive Neuroscience, 22*(6), 1140-1157.


Figure captions

Figure 1. Areas activated by processing of distorted speech signals (meta-analysis 1, in blue) with areas activated during speech production (meta-analysis 2, in red), and their overlap (mauve).

Figure 2. Areas activated by processing of distorted speech signals (meta-analysis 1, in blue) with areas activated during pre-lexical speech production (meta-analysis 2, in red), areas activated during post-lexical speech production (green), and overlap between speech comprehension and post-lexical speech production (turquoise).
Table I. Experiments included in meta-analysis 1 (difficult speech comprehension). All included studies contrasted comprehension of less intelligible speech with more intelligible speech. FWHM: Full-Width Half Maximum in millimeters (mm).

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Participants</th>
<th>Language</th>
<th>Stimuli</th>
<th>Task</th>
<th>Contrast</th>
<th>Paradigm</th>
<th>Distortion</th>
<th>Smooth (FWHM in mm)</th>
<th>Foci</th>
<th>Source</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Adank &amp; Devlin, 2010)</td>
<td>fMRI</td>
<td>18</td>
<td>British-English sentences</td>
<td>semantic verification</td>
<td>time-compressed &gt; normal speed sentences</td>
<td>continuous</td>
<td>artificial time-compression</td>
<td>6</td>
<td>12</td>
<td>Table I</td>
<td>block</td>
<td></td>
</tr>
<tr>
<td>(Adank, Davis, &amp; Hagoort, in press)</td>
<td>fMRI</td>
<td>26</td>
<td>Dutch sentences</td>
<td>semantic decision</td>
<td>unfamiliar accent &gt; familiar accent</td>
<td>sparse</td>
<td>unfamiliar accent</td>
<td>8</td>
<td>10</td>
<td>Table III</td>
<td>event-related</td>
<td></td>
</tr>
<tr>
<td>(Adank, Noordzij, &amp; Hagoort, 2012)</td>
<td>fMRI</td>
<td>20</td>
<td>Dutch sentences</td>
<td>passive listening + after task</td>
<td>unfamiliar accent &gt; familiar accent</td>
<td>continuous</td>
<td>unfamiliar accent</td>
<td>8</td>
<td>1</td>
<td>Table III</td>
<td>repetition</td>
<td></td>
</tr>
<tr>
<td>(Allen et al., 2005)</td>
<td>fMRI</td>
<td>11</td>
<td>British-English words</td>
<td>voice recognition</td>
<td>pitch-shifted words &gt; normal words</td>
<td>sparse</td>
<td>pitch-shifting</td>
<td>7.2</td>
<td>3</td>
<td>Table III</td>
<td>event-related</td>
<td></td>
</tr>
<tr>
<td>(Davis &amp; Johnsrude, 2003)</td>
<td>fMRI</td>
<td>12</td>
<td>British-English sentences</td>
<td>intelligibility judgment</td>
<td>three types of distorted speech &gt; normal speech and SCN</td>
<td>sparse</td>
<td>background noise, noise-vocoded, segmented</td>
<td>12</td>
<td>13</td>
<td>Table I</td>
<td>event-related</td>
<td></td>
</tr>
<tr>
<td>(Peelle, Eason, Schmitter, Schwarzbauer, &amp; Davis, 2010)</td>
<td>fMRI</td>
<td>6</td>
<td>American English sentences</td>
<td>target matching</td>
<td>Standard &gt; Quiet EPI sequence</td>
<td>n/a</td>
<td>different scanning sequence</td>
<td>10</td>
<td>8</td>
<td>Table 4</td>
<td>block</td>
<td></td>
</tr>
<tr>
<td>(Peelle, 2010)</td>
<td>fMRI</td>
<td>8</td>
<td>American sentences</td>
<td>grammatical</td>
<td>time-compressed &gt; normal speed sentences</td>
<td>continuous</td>
<td>artificial time-compression</td>
<td>8</td>
<td>7</td>
<td>Table 3</td>
<td>event-related</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Method</td>
<td>Participants</td>
<td>Task</td>
<td>Condition</td>
<td>Event-related</td>
<td>Continuous</td>
<td>Baseline (related)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>--------</td>
<td>--------------</td>
<td>------</td>
<td>---------------------------------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMillan, Moore, Grossman, &amp; Wingfield, 2004</td>
<td>fMRI</td>
<td>American Englishes</td>
<td>English decision</td>
<td>speed sentences</td>
<td>compression</td>
<td>Baseline)</td>
<td>related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Poldrack et al., 2001)</td>
<td>fMRI</td>
<td>American English sentences</td>
<td>rhyme decision</td>
<td>compression-related increases in BOLD</td>
<td>continuous</td>
<td>artificial time-compression</td>
<td>6</td>
<td>5</td>
<td>Table I block</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sharp et al., 2010)</td>
<td>PET</td>
<td>British-English words</td>
<td>semantic decision (on probe after triplets)</td>
<td>noise-vocoded words &gt; non-degraded normal words</td>
<td>n/a</td>
<td>noise-vocoding</td>
<td>16</td>
<td>2</td>
<td>Table II block</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Wong, Uppanda, Parrish, &amp; Dhar, 2008)</td>
<td>fMRI</td>
<td>British-English words</td>
<td>target to picture matching</td>
<td>speech in noise &gt; speech in quiet</td>
<td>sparse</td>
<td>background noise</td>
<td>6</td>
<td>20</td>
<td>Table I block</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table II. Meta-analysis 1 (difficult speech comprehension): activated clusters for all included studies, including number of contributing foci ([1]).

MTG: Middle Temporal Gyrus; pre-SMA: anterior Supplementary Motor Cortex, STS: Superior Temporal Sulcus.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Location</th>
<th>$mm^3$</th>
<th>ALE</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BA</th>
<th>Contributing studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Left anterior STS</td>
<td>1824</td>
<td>0.022</td>
<td>-60</td>
<td>-14</td>
<td>-2</td>
<td>22</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Davis &amp; Johnsrude (2003) [2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (in press) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wong, et al. (2008) [1]</td>
</tr>
<tr>
<td>2</td>
<td>Left posterior MTG</td>
<td>1136</td>
<td>0.021</td>
<td>-58</td>
<td>-46</td>
<td>4</td>
<td>22</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Davis &amp; Johnsrude (2003) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td>3</td>
<td>Pre-SMA</td>
<td>968</td>
<td>0.019</td>
<td>0</td>
<td>22</td>
<td>44</td>
<td>6</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Davis &amp; Johnsrude (2003) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wong, et al. (2008) [1]</td>
</tr>
<tr>
<td>4</td>
<td>Left anterior Insula</td>
<td>880</td>
<td>0.022</td>
<td>-36</td>
<td>24</td>
<td>-4</td>
<td>13</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allen, et al. (2005) [1]</td>
</tr>
<tr>
<td>5</td>
<td>Right anterior Insula</td>
<td>720</td>
<td>0.018</td>
<td>36</td>
<td>26</td>
<td>2</td>
<td>13</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Poldrack et al. (2001) [1]</td>
</tr>
<tr>
<td>6</td>
<td>Right anterior STS</td>
<td>592</td>
<td>0.017</td>
<td>64</td>
<td>-14</td>
<td>0</td>
<td>22</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td>7</td>
<td>Pre-SMA</td>
<td>472</td>
<td>0.016</td>
<td>0</td>
<td>12</td>
<td>60</td>
<td>6</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
<tr>
<td>8</td>
<td>Right posterior MTG</td>
<td>456</td>
<td>0.016</td>
<td>56</td>
<td>-32</td>
<td>4</td>
<td>22</td>
<td>Adank &amp; Devlin (2010) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adank, et al. (2012) [1]</td>
</tr>
</tbody>
</table>
Table III. Experiments included in meta-analysis 2 (speech production). All included studies contrasted speech production with a condition in which no speech was produced. FWHM: Full-Width Half Maximum in millimeters (mm).

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Participants</th>
<th>Language</th>
<th>Stimuli</th>
<th>Task</th>
<th>Contrast</th>
<th>Paradigm</th>
<th>Smooth (FWHM in mm)</th>
<th>Foci</th>
<th>Source</th>
<th>Design</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Soros et al., 2006)</td>
<td>fMRI</td>
<td>9</td>
<td>Canadian</td>
<td>vowels</td>
<td>repeat speech sounds</td>
<td>speech production &gt; rest</td>
<td>sparse</td>
<td>5</td>
<td>29</td>
<td>Table 1</td>
<td>event-related</td>
<td>rest (silence)</td>
</tr>
<tr>
<td>(Bohland &amp; Guenther, 2006)</td>
<td>fMRI</td>
<td>13</td>
<td>American</td>
<td>syllables</td>
<td>produce three syllables in sequence</td>
<td>speech production &gt; rest</td>
<td>sparse</td>
<td>8</td>
<td>40</td>
<td>Table 1</td>
<td>event-related</td>
<td>rest (silence)</td>
</tr>
<tr>
<td>(Fridriksson et al., 2009)</td>
<td>fMRI</td>
<td>13</td>
<td>American</td>
<td>syllables</td>
<td>read nonsense syllables</td>
<td>speech production &gt; observation speech video</td>
<td>sparse</td>
<td>8</td>
<td>5</td>
<td>Table 1</td>
<td>event-related</td>
<td>observing speech (audiovisual)</td>
</tr>
<tr>
<td>(Ghosh, Tourville, &amp; Guenther, 2008)</td>
<td>fMRI</td>
<td>10</td>
<td>American</td>
<td>syllables</td>
<td>produce isolated monosyllables</td>
<td>speech production &gt; rest</td>
<td>sparse</td>
<td>12</td>
<td>61</td>
<td>Table 3</td>
<td>event-related</td>
<td>rest (silence)</td>
</tr>
<tr>
<td>(Riecker, Wildgruber, Dogil, Grodd, &amp; Ackermann, 2002)</td>
<td>fMRI</td>
<td>12</td>
<td>German</td>
<td>syllables</td>
<td>produce three syllables</td>
<td>speech production &gt; speech perception</td>
<td>continuous</td>
<td>10</td>
<td>13</td>
<td>Table 1</td>
<td>event-related</td>
<td>speech perception</td>
</tr>
<tr>
<td>(Wilson, Saygin, Sereno, &amp; Iacoboni, 2004)</td>
<td>fMRI</td>
<td>10</td>
<td>American</td>
<td>syllables</td>
<td>produce syllable</td>
<td>speech production &gt; rest</td>
<td>continuous</td>
<td>4</td>
<td>5</td>
<td>Table 5</td>
<td>block</td>
<td>rest (scanner noise)</td>
</tr>
<tr>
<td>(Zheng, Munhall, &amp; Yang, 2011)</td>
<td>fMRI</td>
<td>21</td>
<td>Canadian</td>
<td>syllable</td>
<td>produce syllable</td>
<td>speech production &gt; rest</td>
<td>sparse</td>
<td>10</td>
<td>7</td>
<td>Table 5</td>
<td>event-related</td>
<td>speech</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>N</td>
<td>Language</td>
<td>Task Description</td>
<td>Event-related</td>
<td>Table</td>
<td>Perception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------</td>
<td>----</td>
<td>---------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>-------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnsrude, 2009</td>
<td>fMRI</td>
<td>13</td>
<td>American English</td>
<td>Pseudowords read pseudo words, speech production &gt; rest</td>
<td>sparse</td>
<td>8</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Golfinopoulos et al., 2011</td>
<td>fMRI</td>
<td>18</td>
<td>American English</td>
<td>Pseudowords reading bisyllabic pseudowords, speech production &gt; rest (viewing strings of ‘XXXX’)</td>
<td>continuous</td>
<td>12</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shuster, 2009</td>
<td>fMRI</td>
<td>14</td>
<td>American English</td>
<td>Pseudowords produce pseudowords, speech production &gt; listen to pink noise stimuli</td>
<td>continuous</td>
<td>6</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alario, Chainay, Lehericy, &amp; Cohen, 2006</td>
<td>fMRI</td>
<td>10</td>
<td>French words</td>
<td>Repeat words after auditory presentation, speech production &gt; rest</td>
<td>continuous</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blank, Scott, Murphy, Warburton, &amp; Wise, 2002</td>
<td>PET</td>
<td>8</td>
<td>British-English words</td>
<td>Respond to queries about personal experience, cite nursery rhymes, repeat word list, speech production &gt; rest</td>
<td>n/a</td>
<td>10</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crescentini, Shallice, &amp; Macaluso</td>
<td>fMRI</td>
<td>14</td>
<td>Italian words</td>
<td>Read aloud words, speech production &gt; reading</td>
<td>continuous</td>
<td>8</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dogil et al., 2002</td>
<td>fMRI</td>
<td>9</td>
<td>German words</td>
<td>Repeat months of the year, overt &gt; covert speaking</td>
<td>continuous</td>
<td>10</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heim, Friederici, Schiller, Ruschemeyer, &amp;</td>
<td>fMRI</td>
<td>16</td>
<td>German words</td>
<td>Picture naming, picture naming &gt; rest</td>
<td>sparse</td>
<td>8</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>fMRI</td>
<td>N</td>
<td>language</td>
<td>task</td>
<td>contrast</td>
<td>materials</td>
<td>event-related</td>
<td>notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------</td>
<td>---</td>
<td>----------</td>
<td>-----------------------</td>
<td>-----------------------------------</td>
<td>--------------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amunts, 2009</td>
<td>tMRI</td>
<td>10</td>
<td>American English</td>
<td>words produce words overt speaking &gt; covert speaking</td>
<td>sparse</td>
<td>4</td>
<td>Table 1 (“Over &gt; Covert”)</td>
<td>covert speaking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Shuster &amp; Lemieux, 2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Tremblay &amp; Small, 2011)</td>
<td>tMRI</td>
<td>20</td>
<td>American English</td>
<td>words read words speech production &gt; observation speech video</td>
<td>continuous</td>
<td>6</td>
<td>Table 1 (“Production &gt; Perception”)</td>
<td>observing speech (audiovisual)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Turkeltaub, Eden, Jones, &amp; Zeffiro, 2002)</td>
<td>tMRI</td>
<td>32</td>
<td>American English</td>
<td>words reading words speech production &gt; reading</td>
<td>sparse</td>
<td>9</td>
<td>Table 2</td>
<td>block reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Whitney et al., 2009)</td>
<td>tMRI</td>
<td>18</td>
<td>German</td>
<td>words phonological verbal fluency speech production &gt; reading</td>
<td>continuous</td>
<td>10</td>
<td>Table 3 (“PVF &gt; Repeat”)</td>
<td>reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Kell, Morillon, Kouneiher, &amp; Giraud, 2011)</td>
<td>tMRI</td>
<td>26</td>
<td>French</td>
<td>sentences read sentences speech production &gt; reading</td>
<td>continuous</td>
<td>8</td>
<td>Table 1 (“Execution of overt vs. covert reading”)</td>
<td>reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Brown et al., 2009)</td>
<td>tMRI</td>
<td>16</td>
<td>Canadian English</td>
<td>narrative/poem read aloud Beowulf speech production &gt; rest</td>
<td>continuous</td>
<td>8</td>
<td>Table 1 (“Speech production &gt; speech viewing”)</td>
<td>block scanner noise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table IV. Meta-analysis 2: activated clusters for all included studies, including number of contributing foci ([]). IFG/PO: Inferior Frontal Gyrus/Pars Opercularis; (pre-)SMA: Supplementary Motor Area, STG: Superior Temporal Gyrus; STS: Superior Temporal Sulcus.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Location</th>
<th>mm³</th>
<th>ALE</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BA</th>
<th>Contributing studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heim, et al. 2008 [1]</td>
</tr>
<tr>
<td></td>
<td>Left Precentral Gyrus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1888</td>
<td>0.024</td>
<td>58</td>
<td>-4</td>
<td>20</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.024</td>
<td>58</td>
<td>-8</td>
<td>26</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.018</td>
<td>56</td>
<td>-4</td>
<td>44</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.029</td>
<td>-12</td>
<td>-20</td>
<td>0</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.032</td>
<td>-16</td>
<td>-64</td>
<td>-22</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.028</td>
<td>-60</td>
<td>-12</td>
<td>-2</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.031</td>
<td>-40</td>
<td>-28</td>
<td>12</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.021</td>
<td>44</td>
<td>-8</td>
<td>34</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.019</td>
<td>-50</td>
<td>12</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**7** Left Thalamus 1584 0.029 -12 -20 0 -

**8** Left Dentate Gyrus 1424 0.032 -16 -64 -22 -

**9** Left anterior STG 1328 0.028 -60 -12 -2 22

**10** Left Heschl's Gyrus 1128 0.031 -40 -28 12 41

**11** Left Precentral Gyrus 504 0.021 44 -8 34 6

**12** Left anterior Insula 456 0.019 -50 12 4 13

**Left IFG/PO** 0.019 -52 10 12 44
Table V. Meta-analysis 2 for studies using production of pre-lexical and post-lexical speech items separately: activated clusters for all included studies, including number of contributing foci ([ ]). SMA: Supplementary Motor Area, TTG: Transverse Temporal Gyrus.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Location</th>
<th>mm(^{+})</th>
<th>ALE</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>BA</th>
<th>Contributing studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-lexical</td>
</tr>
<tr>
<td>1</td>
<td>Left Putamen</td>
<td>2064</td>
<td>0.024</td>
<td>-2</td>
<td>6</td>
<td>0.024</td>
<td>Bohland &amp; Guenther (2006) [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Peeva et al 2010 [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ghosh et al 2008 [4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shuster et al 2009 [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fridriksson et al 2009 [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Golfinopoulos et al 2011 [2]</td>
</tr>
<tr>
<td>2</td>
<td>Left Lentiform Nucleus</td>
<td>1464</td>
<td>0.024</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>0.024</td>
<td>Bohland &amp; Guenther 2006) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shuster et al (2009) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Golfinopoulos et al (2011) [1]</td>
</tr>
<tr>
<td>3</td>
<td>Left Thalamus</td>
<td>1192</td>
<td>0.022</td>
<td>-20</td>
<td>-2</td>
<td>0.022</td>
<td>Bohland &amp; Guenther 2006) [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shuster et al (2009) [1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Golfinopoulos et al (2011) [2]</td>
</tr>
<tr>
<td>4</td>
<td>Right Thalamus</td>
<td>1120</td>
<td>0.021</td>
<td>12</td>
<td>-12</td>
<td>8</td>
<td>0.021</td>
<td>Bohland &amp; Guenther 2006) [3]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Golfinopoulos et al (2011) [1]</td>
</tr>
<tr>
<td>5</td>
<td>SMA</td>
<td>1056</td>
<td>0.022</td>
<td>-2</td>
<td>66</td>
<td>0.022</td>
<td>Bohland &amp; Guenther (2006) [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Golfinopoulos et al (2011) [1]</td>
</tr>
<tr>
<td></td>
<td>Hemisphere</td>
<td>Area</td>
<td>MNI Coordinates</td>
<td>p-value</td>
<td>References</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>---------</td>
<td>------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
                            |               |                | 0.019 | Fridriksson et al (2009) [1]  
                            |               |                |       | Golfinopoulos et al (2011) [1]  
|7 | Right      | Precentral Gyrus | 736             | 0.014   | Bohland & Guenther (2006) [1]  
                            |               |                | 0.014 | Bohland & Guenther (2006) [1]  
                            |               |                |       | Golfinopoulos et al (2011) [1]  
| 8 | Left       | Cerebellum     | 520             | 0.019   | Ghosh et al (2008) [2]  
                            |               |                | 0.019 | Bohland & Guenther 2006) [1]  
|   | Left       | TTG            | 520             | 0.019   | Bohland & Guenther 2006) [1]  
                            |               |                | 0.019 | Soros et al (2006) [1]  |

*Post-lexical*

<table>
<thead>
<tr>
<th></th>
<th>Left pre-SMA</th>
<th>MNI Coordinates</th>
<th>p-value</th>
<th>References</th>
</tr>
</thead>
</table>
|1 | Left pre-SMA | 888             | 0.024   | Alario et al (2006) [1]  
|   | Right pre-SMA| 0.014           | 6       | Crescentini et al (2010) [1]  
                            |               | 18     | 54     | 6     | Whitney et al (2010) [1]  
                            |               |       |        |       | Dogil et al (2002) [1]  |
|2 | Left anterior STS | 744       | 0.023   | Brown et al (2009) [1]  
                            |               | 21     | -60    | -12   | -4     | Kell et al (2010) [1]  
                            |               |       |        |       | Turkeltaub et al (2002) [1]  |
|---|------------------------|---|---|---|---|------------------------|
| 3 | 416                    | 0.017 | 6 | 58 | -4 | 20                      |
|   |                        |       |   |   |   | Turkeltaub et al (2002) [1] |