

The Neural Bases of Difficult Speech Comprehension and Speech Production: Two
Activation Likelihood Estimation (ALE) Meta-analyses.

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 (Lieberman & 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 (Pelle, 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.

--- *Insert Table I about here* ---

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).

--- *Insert Table II and Figure 1 about here* ---

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 ($14,672\text{mm}^3$) than on the right ($7,016\text{mm}^3$). 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

Figure 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, Bontrovato, 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, & 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, Bontrovato, 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 (Lieberman & Mattingly, 1985; Lieberman & 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

- Adank, P., Davis, M., & Hagoort, P. (in press). Neural dissociation in comprehension of distorted speech. *Neuropsychologia*.
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, *49*(1), 1124-1132.
- Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception and Performance*, *35*(2), 520-529.
- Adank, P., Noordzij, M. L., & Hagoort, P. (2012). The role of Planum Temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping*.

- Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research, 1076*(1), 129-143.
- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex, 15*(2), 221-228.
- Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., et al. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping, 26*(44-53).
- Binder, J. R., Swanson, S. J., Hammeke, T. A., & Sabsevitz, D. S. (2008). A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia, 49*(12), 1980-1997.
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125*(8), 1829-1838.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage, 32*(2), 821-841.
- Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain and Cognition, 70*(1), 31-41.
- 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.
- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. S. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain, 126*, 1193-1201.

- Crinion, J. T., & Price, C. J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*, *128*(Pt 12), 2858-2871.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914-3932.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423-3431.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Benvivuto, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799-816.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Benvivuto, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional Magnetic Resonance Imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799-816.
- Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., et al. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*, 59-90.
- Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 949-958). Cambridge, MA: MIT Press.

- Dronkers, N. F., Wilkins, D. P., Van Valin Jr., R., Redfern, B. B., & Jaeger, J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145-177.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914-927.
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, *367*(1896), 2399-2421.
- Fridriksson, J., Moser, D., Ryalls, J., Bonilla, L., Rorden, C., & Baylis, G. (2009). Modulation of Frontal Lobe Speech Areas Associated With the Production and Perception of Speech Movements. *Journal of Speech, Hearing and Language Research*, *52*, 812-819.
- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, *31*(3), 448-457.
- Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech Language and Hearing Research*, *51*(5), 1183-1202.
- Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*, 1324-1338.

- Gonzalez-Castillo, J., & Talavage, T. (2011). Reproducibility of fMRI activations associated with auditory sentence comprehension. *NeuroImage*, *54*, 2138-2155.
- Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, *817*, 45-58.
- Heim, S., Friederici, A. D., Schiller, N. O., Ruschemeyer, S. A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human Brain Mapping*, *30*(3), 928-940.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393-402.
- Holt, L., & Lotto, A. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42-46.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101-144.
- Ischebeck, A. K., Friederici, A. D., & Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cerebral Cortex*, *18*(3), 541-552.
- Jardri, R., Pins, D., Bubrovsky, M., Desprez, P., Pruvo, J. P., Steinling, M., et al. (2007). Self awareness and speech processing: an fMRI study. *NeuroImage*, *35*(4), 1645-1653.
- Kell, C. A., Morillon, B., Kouneiher, F., & Giraud, A. (2011). Lateralization of Speech Production Starts in Sensory Cortices—A Possible Sensory Origin of Cerebral Left Dominance for Speech. *Cerebral Cortex*, *21*, 932-937.
- Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., et al. (2005). ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*, 155-164.

- Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2009). The cortical dynamics of intelligible speech. *Journal of Neuroscience*, 28(49), 13209-13215.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn Sci*, 4(5), 187-196.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59, 203-243.
- Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the Processing Continuum of Single-Word Comprehension in Aphasia. *Journal of Speech, Hearing and Language Research*, 48, 884-896.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12), 1362-1368.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, 28(32), 8116-8123.

- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633-640.
- Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *NeuroImage*, *56*(4), 2310-2320.
- Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions *Journal of Neuroscience*, *27*, 2283-2289.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486-2495.
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage*, *52*(4), 1410-1419.
- Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, *4*, 1-3.
- Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language*, *91*, 315-325.
- Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., et al. Distinct representations of phonemes, syllables, and supra-

- syllabic sequences in the speech production network. *NeuroImage*, 50(2), 626-638.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105-110.
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the Neural Bases of Dynamic Auditory Processing and Phonological Processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, 13(5), 687-697.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724.
- Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. *NeuroImage*, 16(1), 169-176.
- Rimol, L. M., Specht, K., & Hugdahl, K. (2006). Controlling for individual differences in fMRI brain activation to tones, syllables, and words. *NeuroImage*, 30(2), 554-562.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261-1269.
- Rodd, J. M., Longe, O. L., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48, 1324-1335.

- Saur, D., Schelte, B., Schnell, S., Kratochvil, D., Küpper, D., Kellmeyer, P., et al. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension *NeuroImage*, *49*(3187–3197).
- Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory fMRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *21*(317-325).
- Schon, D., Gordon, R., Campagne, A., Magne, C., Astesano, C., Anton, J. L., et al. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450-461.
- Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America*, *44*, 1735-1736.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123 Pt 12*, 2400-2406.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *Journal of the Acoustical Society of America*, *120*(2), 1075-1083.
- Seifritz, E., Di Salle, F., Esposito, F., Herdener, M., Neuhoff, J. G., & Scheffler, K. (2006). Enhancing BOLD response in the auditory system by neuro-physiologically tuned fMRI sequence. *Neuroimage*, *29*, 1013-1022.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303-304.
- Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., Vigliocco, G., & Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Human Brain Mapping*, *31*(3), 365-377.

- Shuster, L. I. (2009). The effect of sublexical and lexical frequency on speech production: An fMRI investigation. *Brain Lang*, *111*(1), 66-72.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*(1), 20-31.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 250-285). Cambridge: Cambridge University Press.
- Soros, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *NeuroImage*, *32*(1), 376-387.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*, 1210-1223.
- Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., et al. (1998). Localizing components of a complex task: sentence processing and working memory. *Neuroreport*, *9*(2995-2999).
- Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*.
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*, 1-15.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta_analysis of the functional neuroanatomy of single word reading: Method and Validation. *NeuroImage*, *16*, 765-780.

- Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex*, *20*(2), 352-364.
- Utman, J., Blumstein, S. E., & Sullivan, K. (2001). Mapping from Sound to Meaning: Reduced Lexical Activation in Broca's Aphasics. *Brain and Language*, *79*, 444-472.
- Vigneau, M., Van Deuren, V., Herve, P., Jobard, G., Petit, L., Crivello, F., et al. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*, 577-593.
- Whalen, D., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America*, *119*(1), 575-581.
- Whitney, C., Weis, S., Krings, T., Huber, W., Grossman, M., & Kircher, T. (2009). Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *Journal of Cognitive Neuroscience*, *21*(4), 697-712.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384-1389.
- Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230-242.

- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701-702.
- Wong, P. C. M., Uppanda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical Mechanisms of Speech Perception in Noise. *Journal of Speech, Hearing and Language Research*, *51*, 1026-1041.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770-1781.

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).

MARKED REVISION

The Neural Bases of Difficult Speech Comprehension and Speech Production: Two
Activation Likelihood Estimation (ALE) Meta-analyses.

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 (Lieberman & 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 (Pelle, 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.

--- *Insert Table I about here* ---

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).

--- *Insert Table II and Figure 1 about here* ---

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 ($14,672\text{mm}^3$) than on the right ($7,016\text{mm}^3$). 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

Figure 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, Bontrovato, 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, & 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, Bontrovato, 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 (Lieberman & 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

- Adank, P., Davis, M., & Hagoort, P. (in press). Neural dissociation in comprehension of distorted speech. *Neuropsychologia*.
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, *49*(1), 1124-1132.
- Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception and Performance*, *35*(2), 520-529.
- Adank, P., Noordzij, M. L., & Hagoort, P. (2012). The role of Planum Temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping*.

- Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research, 1076*(1), 129-143.
- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex, 15*(2), 221-228.
- Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., et al. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping, 26*(44-53).
- Binder, J. R., Swanson, S. J., Hammeke, T. A., & Sabsevitz, D. S. (2008). A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia, 49*(12), 1980-1997.
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125*(8), 1829-1838.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage, 32*(2), 821-841.
- Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain and Cognition, 70*(1), 31-41.
- 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.
- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. S. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain, 126*, 1193-1201.

- Crinion, J. T., & Price, C. J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*, *128*(Pt 12), 2858-2871.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914-3932.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423-3431.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Benvivuto, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799-816.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Benvivuto, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional Magnetic Resonance Imaging and lesion data. *Journal of Cognitive Neuroscience*, *19*(5), 799-816.
- Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., et al. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*, 59-90.
- Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 949-958). Cambridge, MA: MIT Press.

- Dronkers, N. F., Wilkins, D. P., Van Valin Jr., R., Redfern, B. B., & Jaeger, J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*, 145-177.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914-927.
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, *367*(1896), 2399-2421.
- Fridriksson, J., Moser, D., Ryalls, J., Bonilla, L., Rorden, C., & Baylis, G. (2009). Modulation of Frontal Lobe Speech Areas Associated With the Production and Perception of Speech Movements. *Journal of Speech, Hearing and Language Research*, *52*, 812-819.
- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, *31*(3), 448-457.
- Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech Language and Hearing Research*, *51*(5), 1183-1202.
- Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*, 1324-1338.

- Gonzalez-Castillo, J., & Talavage, T. (2011). Reproducibility of fMRI activations associated with auditory sentence comprehension. *NeuroImage*, *54*, 2138-2155.
- Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, *817*, 45-58.
- Heim, S., Friederici, A. D., Schiller, N. O., Ruschemeyer, S. A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human Brain Mapping*, *30*(3), 928-940.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393-402.
- Holt, L., & Lotto, A. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42-46.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101-144.
- Ischebeck, A. K., Friederici, A. D., & Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cerebral Cortex*, *18*(3), 541-552.
- Jardri, R., Pins, D., Bubrovsky, M., Desprez, P., Pruvo, J. P., Steinling, M., et al. (2007). Self awareness and speech processing: an fMRI study. *NeuroImage*, *35*(4), 1645-1653.
- Kell, C. A., Morillon, B., Kouneiher, F., & Giraud, A. (2011). Lateralization of Speech Production Starts in Sensory Cortices—A Possible Sensory Origin of Cerebral Left Dominance for Speech. *Cerebral Cortex*, *21*, 932-937.
- Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., et al. (2005). ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*, 155-164.

- Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2009). The cortical dynamics of intelligible speech. *Journal of Neuroscience*, 28(49), 13209-13215.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends Cogn Sci*, 4(5), 187-196.
- Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59, 203-243.
- Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the Processing Continuum of Single-Word Comprehension in Aphasia. *Journal of Speech, Hearing and Language Research*, 48, 884-896.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12), 1362-1368.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, 28(32), 8116-8123.

- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633-640.
- Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *NeuroImage*, *56*(4), 2310-2320.
- Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions *Journal of Neuroscience*, *27*, 2283-2289.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486-2495.
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage*, *52*(4), 1410-1419.
- Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, *4*, 1-3.
- Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language*, *91*, 315-325.
- Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., et al. Distinct representations of phonemes, syllables, and supra-

- syllabic sequences in the speech production network. *NeuroImage*, 50(2), 626-638.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105-110.
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the Neural Bases of Dynamic Auditory Processing and Phonological Processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, 13(5), 687-697.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724.
- Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. *NeuroImage*, 16(1), 169-176.
- Rimol, L. M., Specht, K., & Hugdahl, K. (2006). Controlling for individual differences in fMRI brain activation to tones, syllables, and words. *NeuroImage*, 30(2), 554-562.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, 15(8), 1261-1269.
- Rodd, J. M., Longe, O. L., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, 48, 1324-1335.

- Saur, D., Schelte, B., Schnell, S., Kratochvil, D., Küpper, D., Kellmeyer, P., et al. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension *NeuroImage*, *49*(3187–3197).
- Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory fMRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *21*(317-325).
- Schon, D., Gordon, R., Campagne, A., Magne, C., Astesano, C., Anton, J. L., et al. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450-461.
- Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America*, *44*, 1735-1736.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123 Pt 12*, 2400-2406.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *Journal of the Acoustical Society of America*, *120*(2), 1075-1083.
- Seifritz, E., Di Salle, F., Esposito, F., Herdener, M., Neuhoff, J. G., & Scheffler, K. (2006). Enhancing BOLD response in the auditory system by neuro-physiologically tuned fMRI sequence. *Neuroimage*, *29*, 1013-1022.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303-304.
- Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., Vigliocco, G., & Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Human Brain Mapping*, *31*(3), 365-377.

- Shuster, L. I. (2009). The effect of sublexical and lexical frequency on speech production: An fMRI investigation. *Brain Lang*, *111*(1), 66-72.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*(1), 20-31.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 250-285). Cambridge: Cambridge University Press.
- Soros, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *NeuroImage*, *32*(1), 376-387.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*, 1210-1223.
- Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., et al. (1998). Localizing components of a complex task: sentence processing and working memory. *Neuroreport*, *9*(2995-2999).
- Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*.
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*, 1-15.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta_analysis of the functional neuroanatomy of single word reading: Method and Validation. *NeuroImage*, *16*, 765-780.

- Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex*, *20*(2), 352-364.
- Utman, J., Blumstein, S. E., & Sullivan, K. (2001). Mapping from Sound to Meaning: Reduced Lexical Activation in Broca's Aphasics. *Brain and Language*, *79*, 444-472.
- Vigneau, M., Van Deuren, V., Herve, P., Jobard, G., Petit, L., Crivello, F., et al. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*, 577-593.
- Whalen, D., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America*, *119*(1), 575-581.
- Whitney, C., Weis, S., Krings, T., Huber, W., Grossman, M., & Kircher, T. (2009). Task-dependent modulations of prefrontal and hippocampal activity during intrinsic word production. *Journal of Cognitive Neuroscience*, *21*(4), 697-712.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384-1389.
- Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230-242.

- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701-702.
- Wong, P. C. M., Uppanda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical Mechanisms of Speech Perception in Noise. *Journal of Speech, Hearing and Language Research*, *51*, 1026-1041.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2009). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770-1781.

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

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).

<i>Study</i>	<i>Method</i>	<i>Participa nts</i>	<i>Languag e</i>	<i>Stimuli</i>	<i>Task</i>	<i>Contrast</i>	<i>Paradigm</i>	<i>Distortion</i>	<i>Smooth (FWHM in mm)</i>	<i>Foci</i>	<i>Source</i>	<i>Design</i>
(Adank & Devlin, 2010)	fMRI	18	British- English	sentenc es	semantic verification	time-compressed > normal speed sentences	continuous	artificial time- compression	6	12	Table 1	block
(Adank, Davis, & Hagoort, in press)	fMRI	26	Dutch	sentenc es	semantic decision	unfamiliar accent > familiar accent	sparse	unfamiliar accent	8	10	Table III ("(DA + DSDA) > (SS + DS)")	event- related
(Adank, Noordzij, & Hagoort, 2012)	fMRI	20	Dutch	sentenc es	passive listening + after task	unfamiliar accent > familiar accent	continuous	unfamiliar accent	8	1	Table III ("Accent > clear")	repetition - suppression
(Allen et al., 2005)	fMRI	11	British- English	words	voice recognition	pitch-shifted words > normal words	sparse	pitch-shifting	7.2	3	Table III ("Dist > Undist")	event- related
(Davis & Johnsruide, 2003)	fMRI	12	British- English	sentenc es	intelligibility judgment	three types of distorted speech > normal speech and SCN	sparse	background noise, noise- vocoded, segmented	12	13	Table 1 ("Form-independent activity increases for distorted speech ...")	event- related
(Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010)	fMRI	6	American English	sentenc es	target matching	Standard > Quiet EPI sequence	n/a	different scanning sequence	10	8	Table 4	block
(Peelle,	fMRI	8	American	sentenc	grammatical	time-compressed > normal	continuous	artificial time-	8	7	Table 3 ("65 >	event-

McMillan, Moore, Grossman, & Wingfield, 2004)			English	es	decision	speed sentences		compression			Baseline")	related
(Poldrack et al., 2001)	fMRI	8	American English	sentenc es	rhyme decision	compression-related increases in BOLD	continuous	artificial time- compression	6	5	Table 1 ("Compression- Related Increase")	block
(Sharp et al., 2010)	PET	12	British- English	words	semantic decision (on probe after triplets)	noise-vocoded words > non-degraded normal words	n/a	noise- vocoding	16	2	Table II ("SLPH vs. SLPL")	block
(Wong, Uppanda, Parrish, & Dhar, 2008)	fMRI	11	British- English	words	target to picture matching	speech in noise > speech in quiet	sparse	background noise	6	20	Table 1 ("5 dB vs. Quiet")	block

Table II

Table II. Meta-analysis 1 (difficult speech comprehension): activated clusters for all included studies, including number of contributing foci ([]).

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

<i>Cluster</i>	<i>Location</i>	<i>mm³</i>	<i>ALE</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>BA</i>	<i>Contributing studies</i>	
1	Left anterior STS	1824	0.022	-60	-14	-2	22	Adank & Devlin (2010) [1] Adank, et al. (2012) [1] Adank, et al. (in press) [1]	Davis & Johnsrude (2003) [2] Peelle, et al. (2010) [1] Wong, et al. (2008) [1]
2	Left posterior MTG	1136	0.021	-58	-46	4	22	Adank & Devlin (2010) [1] Adank, et al. (2012) [1]	Davis & Johnsrude (2003) [1] Peelle, et al. (2010) [1]
3	Pre-SMA	968	0.019	0	22	44	6	Adank & Devlin (2010) [1] Adank, et al. (2012) [1]	Davis & Johnsrude (2003) [1] Wong, et al. (2008) [1]
4	Left anterior Insula	880	0.022	-36	24	-4	13	Adank & Devlin (2010) [1] Adank, et al. (2012) [1]	Allen, et al. (2005) [1]
5	Right anterior Insula	720	0.018	36	26	2	13	Adank & Devlin (2010) [1] Adank, et al. (2012) [1]	Poldrack et al. (2001) [1]
6	Right anterior STS	592	0.017	64	-14	0	22	Adank & Devlin (2010) [1]	Adank, et al. (2012) [1]
7	Pre-SMA	472	0.016	0	12	60	6	Adank & Devlin (2010) [1]	Adank, et al. (2012) [1]
8	Right posterior MTG	456	0.016	56	-32	4	22	Adank & Devlin (2010) [1]	Adank, et al. (2012) [1]

Table III

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).

<i>Study</i>	<i>Method</i>	<i>Participants</i>	<i>Language</i>	<i>Stimuli</i>	<i>Task</i>	<i>Contrast</i>	<i>Paradigm</i>	<i>Smooth (FWHM in mm)</i>	<i>Foci</i>	<i>Source</i>	<i>Design</i>	<i>Baseline</i>
<i>Prelexical Speech Production</i>												
(Soros et al., 2006)	fMRI	9	Canadian English	vowels	repeat speech sounds	speech production > rest	sparse	5	29	Table 1	event-related	rest (silence)
(Bohland & Guenther, 2006)	fMRI	13	American English	syllables	produce three syllables in sequence	speech production > rest	sparse	8	40	Table 1 ("S_syl S_seq")	event-related	rest (silence)
(Fridriksson et al., 2009)	fMRI	13	American English	syllables	read nonsense syllables	speech production > observation speech video	sparse	8	5	Table 1 ("Speech production > speech viewing")	event-related	observing speech (audiovisual)
(Ghosh, Tourville, & Guenther, 2008)	fMRI	10	American English	syllables	produce isolated monosyllables	speech production > rest	sparse	12	61	Table 3 ("Monosyllables > Baseline")	event-related	rest (silence)
(Riecker, Wildgruber, Dogil, Grodd, & Ackermann, 2002)	fMRI	12	German	syllables	produce three syllables	speech production > speech perception	continuous	10	13	Table 1 ("Isochronous vs Perceptual baseline")	event-related	speech perception
(Wilson, Saygin, Sereno, & Iacoboni, 2004)	fMRI	10	American English	syllables	produce syllable	speech production > rest	continuous	4	5	Supplementary Table 1	block	rest (scanner noise)
(Zheng, Munhall,	fMRI	21	Canadian	syllable	produce syllable	speech production >	sparse	10	7	Table 5	event-	speech

& Johnsrude, 2009)			English		(‘Ted’)	speech perception					related	perception
(Golfinoopoulos et al., 2011)	fMRI	13	American English	pseudowords	read pseudo words	speech production > rest	sparse	8	40	Table 1 (“Speech > Baseline”)	event-related	rest (silence)
(Peeva et al.)	fMRI	18	American English	pseudowords	reading bisyllabic pseudowords	speech production > rest (viewing strings of ‘XXXX’)	continuous	12	34	Table 2 (“Collapsed”)	block	scanner noise
(Shuster, 2009)	fMRI	14	American English	pseudowords	produce pseudowords	speech production > listen to pink noise stimuli	continuous	6	21	Table 2	event-related	pink noise
<i>Postlexical Speech Production</i>												
(Alario, Chainay, Lehericy, & Cohen, 2006)	fMRI	10	French	words	repeat words after auditory presentation	speech production > rest	continuous	5	3	Table 1 (“Reading > Fixation”)	block	rest (scanner noise)
(Blank, Scott, Murphy, Warburton, & Wise, 2002)	PET	8	British-English	words	respond to queries about personal experience/ cite nursery rhymes/ repeat word list	speech production > rest	n/a	10	12	Caption Figure 1	block	rest (silence)
(Crescentini, Shallice, & Macaluso)	fMRI	14	Italian	words	read aloud words	speech production > reading	continuous	8	24	Table 1 (“... Generation versus Read”)	block	reading
(Dogil et al., 2002)	fMRI	9	German	words	repeat months of the year	overt > covert speaking	continuous	10	6	Table 1 (“overt Speech”)	event-related	covert speaking
(Heim, Friederici, Schiller, Ruschemeyer, &	fMRI	16	German	words	picture naming	picture naming > rest	sparse	8	39	Table I	event-related	rest (silence)

Amunts, 2009)												
(Shuster & Lemieux, 2005)	fMRI	10	American English	words	produce words	overt speaking > covert speaking	sparse	4	24	Table 1 ("Over > Covert")	event-related	covert speaking
(Tremblay & Small, 2011)	fMRI	20	American English	words	read words	speech production > observation speech video	continuous	6	22	Table 1 ("Production > Perception")	event-related	observing speech (audiovisual)
(Turkeltaub, Eden, Jones, & Zeffiro, 2002)	fMRI	32	American English	words	reading words	speech production > reading	sparse	9	28	Table 2	block	reading
(Whitney et al., 2009)	fMRI	18	German	words	phonological verbal fluency	speech production > reading	continuous	10	11	Table 3 ("PVF > Repeat")	event-related	reading
(Kell, Morillon, Kouneiher, & Giraud, 2011)	fMRI	26	French	sentences	read sentences	speech production > reading	continuous	8	16	Table 1 ("Execution of overt vs. covert reading")	event-related	reading
(Brown et al., 2009)	fMRI	16	Canadian English	narrative/poem	read aloud Beowulf	speech production > rest	continuous	8	31	Table 1 ("Speech production > speech viewing")	block	scanner noise

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.

<i>Cluster</i>	<i>Location</i>	<i>mm³</i>	<i>ALE</i>	<i>x</i>	<i>y</i>	<i>z</i>	<i>BA</i>	<i>Contributing studies</i>	
1	Left pre-SMA	3360	0.033	-2	16	46	32	Alario, et al. 2006 [1]	Golfinopoulos, et al. 2011 [2]
	Left SMA		0.027	2	0	66	6	Blank, et al. 2002 [1]	Peeva, et al. 2010 [1]
	Left pre-SMA		0.026	-2	6	62	6	Bohland & Guenther 2006 [4]	Riecker, et al. 2002 [1]
								Brown, et al. 2009 [1]	Soros, et al. 2006 [2]
								Dogil, et al. 2002 [1]	Tremblay & Small 2011 [2]
								Fridriksson, et al. 2009 [1]	Turkeltaub, et al. 2002 [1]
								Ghosh, et al. 2008 [2]	Whitney, et al. 2010 [1]
2	Left Precentral Gyrus	2880	0.023	-56	-12	26	4	Bohland & Guenther 2006 [1]	Peeva, et al. 2010 [2]
	Left Precentral Gyrus		0.022	-60	-6	36	4	Brown, et al. 2009 [2]	Kell, et al. 2010 [1]
	Left Precentral Gyrus		0.022	-54	0	20	6	Fridriksson, et al. 2009 [1]	Riecker, et al. 2002 [1]
								Ghosh, et al. 2008 [3]	Shuster, et al. 2005 [2]
								Golfinopoulos, et al. 2011 [1]	Turkeltaub, et al. 2002 [2]
3	Right Lentiform Nucleus	2712	0.029	18	-2	4	-	Bohland & Guenther 2006 [2]	Kell, et al. 2010 [1]
	Right Thalamus		0.024	12	-12	6	-	Ghosh, et al. 2008 [1]	Shuster, et al. 2009 [1]
								Golfinopoulos, et al. 2011 [2]	Soros, et al. 2006 [4]
								Heim, et al. 2008 [1]	
4	Right posterior STG	2416	0.025	60	-20	2	41	Bohland & Guenther 2006 [1]	Peeva, et al. 2010 [2]
	Right posterior STG		0.024	46	-22	6	13	Brown, et al. 2009 [1]	Tremblay & Small 2011 [1]
	Right posterior STG		0.021	54	-28	6	41	Ghosh, et al. 2008 [3]	Turkeltaub, et al. 2002 [2]
	Right posterior STG		0.020	64	-22	14	40	Golfinopoulos, et al. 2011 [2]	Wilson, et al. 2004 [1]
5	Left Lentiform Nucleus	2008	0.030	-22	-2	6	-	Bohland & Guenther 2006 [1]	Golfinopoulos, et al. 2011 [2]
	Left Lentiform Nucleus		0.022	-24	-4	-4	-	Brown, et al. 2009 [2]	Peeva, et al. 2010 [1]

								Fridriksson, et al. 2009 [1] Ghosh, et al. 2008 [4]	Shuster, et al. 2009 [1] Soros, et al. 2006 [1]
6	Left Precentral Gyrus	1888	0.024	58	-4	20	6	Bohland & Guenther 2006 [1]	Golfinopoulos, et al. 2011 [1]
	Left Precentral Gyrus		0.024	58	-8	26	4	Brown, et al. 2009 [1] Fridriksson, et al. 2009 [2] Ghosh, et al. 2008 [1]	Kell, et al. 2010 [2] Riecker, et al. 2002 [1]
	Left Precentral Gyrus		0.018	56	-4	44	4	Blank, et al. 2002 [1]	Heim, et al. 2008 [1]
7	Left Thalamus	1584	0.029	-12	-20	0	-	Bohland & Guenther 2006 [1] Ghosh, et al. 2008 [1] Golfinopoulos, et al. 2011 [2]	Kell, et al. 2010 [1] Soros, et al. 2006 [2]
8	Left Dentate Gyrus	1424	0.032	-16	-64	-22	-	Blank, et al. 2002 [1] Bohland & Guenther 2006 [1] Brown, et al. 2009 [1] Ghosh, et al. 2008 [1]	Golfinopoulos, et al. 2011 [1] Peeva, et al. 2010 [1] Kell, et al. 2010 [1] Soros, et al. 2006 [3]
9	Left anterior STG	1328	0.028	-60	-12	-2	22	Bohland & Guenther 2006 [1] Brown, et al. 2009 [1] Peeva, et al. 2010 [1]	Heim, et al. 2008 [1] Kell, et al. 2010 [1] Turkeltaub, et al. 2002 [1]
10	Left Heschl's Gyrus	1128	0.031	-40	-28	12	41	Blank, et al. 2002 [1] Bohland & Guenther 2006 [1] Brown, et al. 2009 [1] Ghosh, et al. 2008 [2]	Soros, et al. 2006 [1] Tremblay & Small 2011 [1] Turkeltaub, et al. 2002 [1]
11	Left Precentral Gyrus	504	0.021	44	-8	34	6	Brown, et al. 2009 [2] Golfinopoulos, et al. 2011 [1]	Peeva, et al. 2010 [1] Tremblay & Small 2011 [1]
12	Left anterior Insula	456	0.019	-50	12	4	13	Kell, et al. 2010 [1]	Shuster, et al. 2009 [2]
	Left IFG/PO		0.019	-52	10	12	44	Riecker, et al. 2002 [1]	

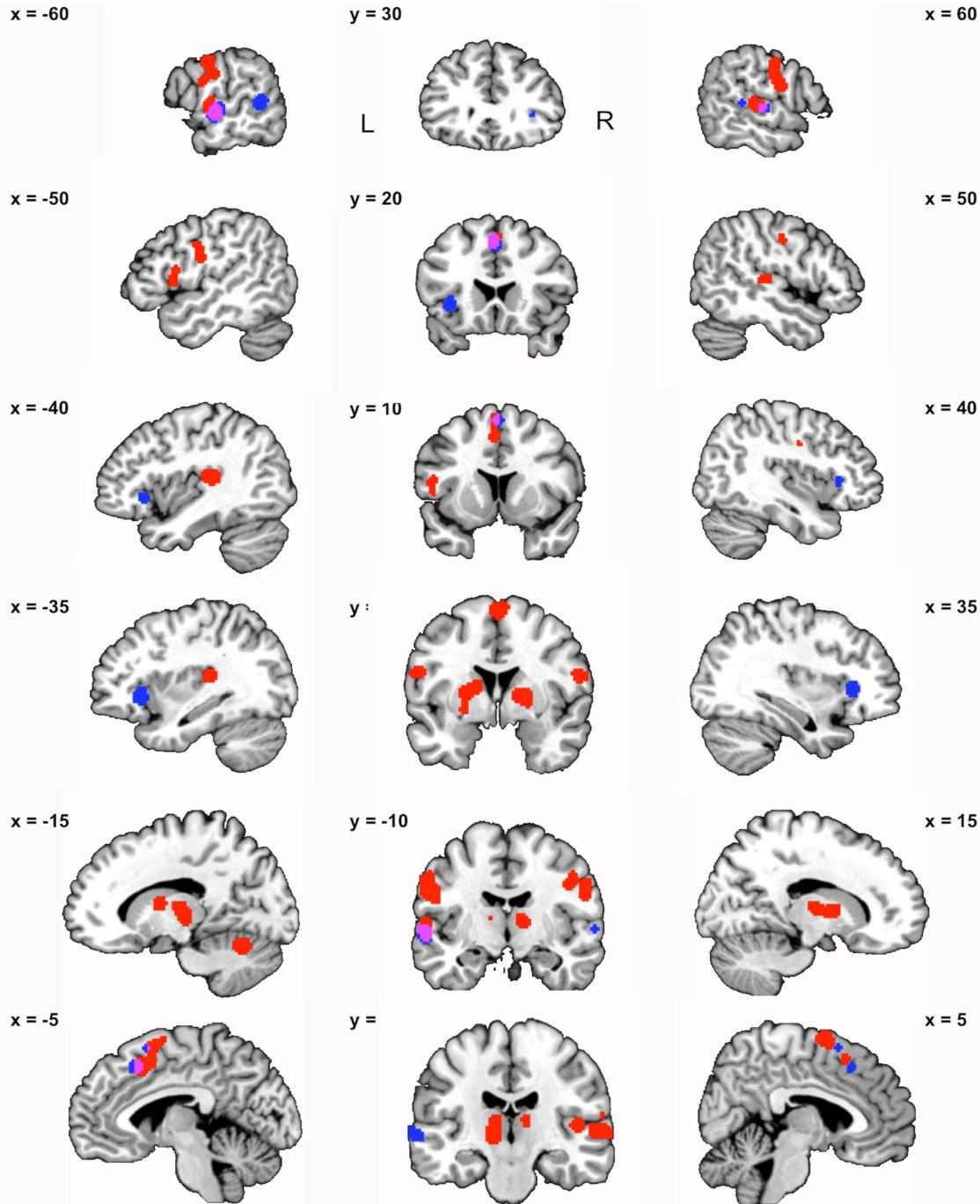
Table V

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.

Cluster	Location	mm ³	ALE	x	y	z	BA	Contributing studies
Pre-lexical								
1	Left Putamen	2064	0.024	-22	-2	6	0.024	Bohland & Guenther (2006) [1] Peeva et al 2010 [1] Ghosh et al 2008 [4] Shuster et al 2009 [1]
	Left Putamen		0.017	-24	0	-6	0.017	Fridriksson et al 2009 [1] Golfinopoulos et al 2011 [2]
2	Left Lentiform Nucleus	1464	0.024	20	0	2	0.024	Bohland & Guenther 2006) [1] Peeva et al (2010) [1] Ghosh et al (2008) [2] Shuster et al (2009) [1] Golfinopoulos et al (2011) [1]
3	Left Thalamus	1192	0.022	-10	-20	-2	0.022] Golfinopoulos et al (2011) [2] Soros et al (2006) [3]
	Left Thalamus		0.020	-12	-18	8	0.020	Bohland & Guenther (2006) [1] Ghosh et al (2008) [1]
4	Right Thalamus	1120	0.021	12	-12	8	0.021	Bohland & Guenther 2006) [3] Golfinopoulos et al (2011) [1]
	Right Thalamus		0.015	14	-24	0	0.015	Ghosh et al (2008) [1] Soros et al (2006) [2]
5	SMA	1056	0.022	0	-2	66	0.022	Bohland & Guenther (2006) [1] Golfinopoulos et al (2011) [1]

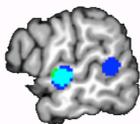
								Ghosh et al (2008) [1] Fridriksson et al (2009) [1]	Peeva et al (2010) [1]
6	Right Precentral Gyrus	896	0.019	60	-6	36	0.019	Fridriksson et al (2009) [1] Golfinopoulos et al (2011) [1] Peeva et al (2010) [1] Wilson et al (2004) [1]	
	Right Precentral Gyrus		0.018	58	-4	44	0.018		
7	Right Precentral Gyrus	736	0.014	56	-8	26	0.014	Bohland & Guenther (2006) [1] Golfinopoulos et al (2011) [1] Peeva et al (2010) [1] Soros et al (2006) [2]	
	Left Cerebellum		0.019	- 20	-64	-24	0.019		
8	Left TTG	520	0.019	- 24	-60	-22	0.019	Bohland & Guenther 2006) [1] Ghosh et al (2008) [2]	Soros et al (2006) [1]
<i>Post-lexical</i>									
1	Left pre-SMA	888	0.024	-2	18	46	6	Alario et al (2006) [1] Crescentini et al (2010) [1] Dogil et al (2002) [1]	Turkeltaub et al (2002) [2] Whitney et al (2010) [1]
	Right pre-SMA		0.014	6	18	54	6		
2	Left anterior STS	744	0,023	21	-60	-12	-4	Brown et al (2009) [1] Heim et al (2008) [1]	Kell et al (2010) [1] Turkeltaub et al (2002) [1]

3	Right Precentral Gyrus	416	0.017	6	58	-4	20	Brown et al (2009) [1] Kell et al (2010) [1]	Turkeltaub et al (2002) [1]
---	---------------------------	-----	-------	---	----	----	----	---	-----------------------------

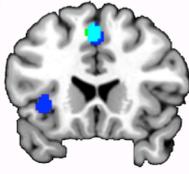


■ Distorted speech comprehension
■ Speech production
■ Overlap between speech comprehension and production

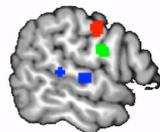
x = -60



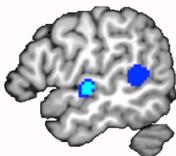
y = 20



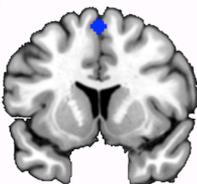
x = 60



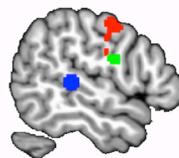
x = -55



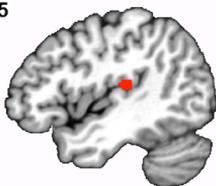
y = 10



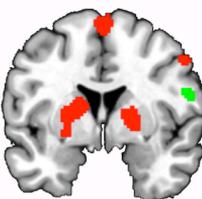
x = 55



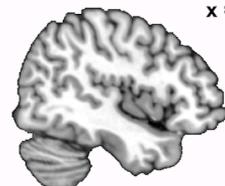
x = -45



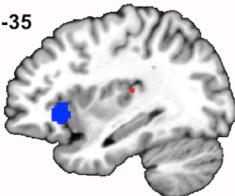
y = 0



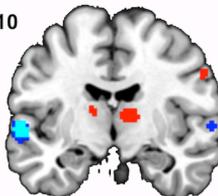
x = 45



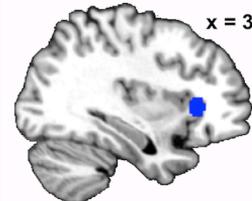
x = -35



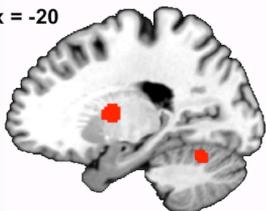
y = -10



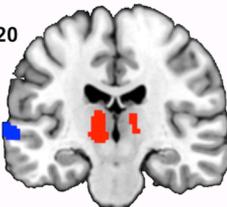
x = 35



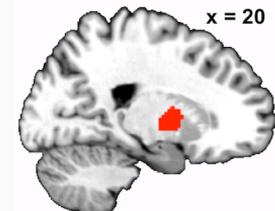
x = -20



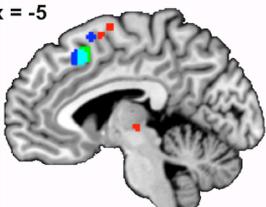
y = -20



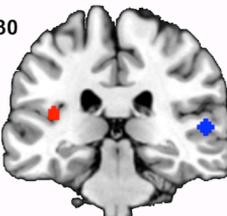
x = 20



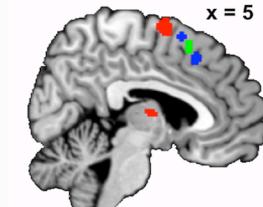
x = -5



y = -30



x = 5



Blue box: Distorted speech comprehension

Red box: Pre-lexical speech production

Green box: Post-lexical speech production

Cyan box: Overlap between speech comprehension and post-lexical speech production