Design Choices in Imaging Speech Comprehension: An Activation Likelihood
Estimation (ALE) Meta-Analysis.

Patti Adank

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

Short title: Design Issues in Imaging Speech

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

Abstract

The localisation of spoken language comprehension is debated extensively: is processing located anterior or posterior on the left temporal lobe, and is it left- or bilaterally organised? An Activation Likelihood Estimation (ALE) analysis was conducted on functional MRI and PET studies investigating speech comprehension to identify the neural network involved in comprehension processing. Furthermore, the analysis aimed to establish the effect of four design choices (scanning paradigm, non-speech baseline, the presence of a task, and the type of stimulus material) on this comprehension network. The analysis included 57 experiments contrasting intelligible with less intelligible or unintelligible stimuli. A large comprehension network was found across bilateral Superior Temporal Sulcus (STS), Middle Temporal Gyrus (MTG) and Superior Temporal (STS) bilaterally, in left Inferior Frontal Gyrus (IFG), left Precentral Gyrus, and Supplementary Motor Area (SMA) and pre-SMA. The core network for post-lexical processing was restricted to the temporal lobes bilaterally with the highest ALE values located anterior to Heschl's Gyrus. Activations in the ALE comprehension network outside the temporal lobes (left IFG, SMA/pre-SMA, and Precentral Gyrus) were driven by the use of sentences instead of words, the scanning paradigm, or the type of non-speech baseline.

Key words: Speech Comprehension, Neuroimaging, Meta-analysis, Methodology, fMRI, PET

1. Introduction

The human speech comprehension system is remarkable in its ability to quickly extract the semantic message from a transient acoustic signal. The intelligibility of a spoken utterance can be defined as the degree to which it can be comprehended by a listener of the language in question (Scott, Blank, Rosen, & Wise, 2000). The neural bases of auditory speech processing has received considerable attention in recent years (e.g., Davis & Johnsruide, 2003; Obleser, Wise, Dresner, & Scott, 2007; Rodd, Davis, & Johnsruide, 2005; Scott, et al., 2000; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006); and has been investigated at sublexical (phonemes, syllables) (Heinrich, Carlyon, Davis, & Johnsruide, 2008; Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005) and post-lexical (words, sentences, and narratives) levels (Davis & Johnsruide, 2003; Obleser, et al., 2007; Scott, et al., 2000; Zekveld, et al., 2006). This study refers to speech *perception* whenever processing occurs at sublexical levels and to speech *comprehension* whenever processing requires understanding, or access to semantic representations (cf. Hickok & Poeppel, 2007).

A recent meta-analysis (Turkeltaub & Coslett, 2010) mapped out the neural network for sublexical perception across 23 functional MRI and PET experiments. Turkeltaub & Coslett placed processing of sublexical components of speech (e.g., phonemes, or syllables) in bilateral Superior Temporal Gyrus (STG) and left posterior Middle Frontal Gyrus (MFG). More specifically, phonetic and phonological processing is located in left middle STS, and phoneme discrimination is located in left Supramarginal and left Angular Gyrus (AG).

Scott et al. (2000), in a Positron Emission Tomography (PET) study focusing on speech comprehension, concluded that left anterior STS responds in particular to

intelligible speech. Davis and Johnsrude (2003) report that the blood oxygenation level-dependent (BOLD) signal correlates with intelligibility along the left superior and middle temporal gyri (as well as in a smaller right-lateralised homologous area), left Inferior Frontal Gyrus (IFG), and left hippocampus. Obleser & Kotz (2010) report that bilateral anterior STS and left AG show increases in BOLD with increasing intelligibility, i.e., respond more to (acoustic) form-independent information.

Several theoretical models outline the neural bases of speech processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003). These models generally converge on the idea that speech is processed via parallel streams in the temporal, parietal and frontal lobes, usually referred to as the dorsal and ventral streams of speech processing. It is generally agreed that the ventral stream is involved in processing auditory perception/comprehension and that its role is to map sound onto meaning. The dorsal auditory pathway is proposed to be involved in mapping heard speech from posterior regions of the STG onto inferior parietal, prefrontal, and premotor areas that are associated with speech production. The models make different predictions on the role of the dorsal and ventral streams of processing speech in general (see Turkeltaub & Coslett, 2010) for an overview) and, more specifically, on the cortical areas involved in processing of speech at post-lexical (i.e., intelligible) levels. All three models propose that the speech comprehension network is placed in the temporal lobe, but disagree about whether it is left-lateralised or bilateral, and on its precise location on the temporal lobe.

First, some propose that comprehension is predominantly left-lateralised (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; McGettigan et al., 2012; Narain et al., 2003; Rauschecker & Scott, 2009; Scott, et al., 2000; Scott & Johnsrude, 2003). Scott and Johnsrude (2003) state that “activation specific to intelligible speech is observed

in the left anterior STS...” (p. 102). Narain et al. (2003) propose an entirely left-lateralised positioning of intelligibility. They argue that other fMRI studies find right-lateralised activations for comprehension of intelligible speech because these studies failed to use baseline conditions that “...adequately controlled for the complex physical properties of speech in contrasts aimed at defining processing of intelligibility.” (p. 1365). Similarly, Rauschecker and Scott (2009) suggest that the network for processing intelligible speech is unilaterally organised and has its centre of gravity in left anterior STS (Superior Temporal Sulcus). Others, however, hold that speech processing is bilaterally organised (Hickok, Houde, & Rong, 2011; Hickok & Poeppel, 2007; Okada et al., 2010). For instance, Hickok and Poeppel propose that processing of intelligible speech activates a bilaterally distributed network of areas, noting that evidence from lesion studies illustrates that left unilateral lesions do not always lead to speech comprehension problems, and that that bilateral damage to temporal regions is associated with severe comprehension problems.

With respect to the second issue, some place the centre of gravity of speech comprehension anterior to Heschl’s Gyrus (Davis & Johnsrude, 2003; Narain, et al., 2003; Rauschecker & Scott, 2009; Scott, et al., 2000; Scott, Rosen, Lang, & Wise, 2006), while others hold that speech processing is distributed anteriorly and posteriorly¹ to Heschl’s Gyrus (Hickok & Poeppel, 2007; Okada & Hickok, 2006; Okada, et al., 2010). Okada et al. (2010) investigated speech comprehension using a design similar to Scott et al. (2000), contrasting intelligible sentences (sentences in quiet) with less intelligible sentences (noise-vocoded sentences), and unintelligible versions of these sentences (spectrally rotated, and noise-vocoded spectrally rotated

¹ Hickok and Poeppel predict that these posterior regions are specifically associated with phonological processes in spoken language processing.

stimuli). They report activations in anterior and posterior STG/STS when contrasting processing of intelligible vs. unintelligible sentences in a whole-brain analysis, while a subsequent multivariate pattern analysis (MVPA) showed that these areas were insensitive to acoustic detail. Okada et al.'s results partially contradict assumptions by Scott et al. (2000), Scott and Johnsrude (2003), and Rauschecker and Scott (2009) assuming that processing of intelligible speech irrespective of phonetic-acoustic detail is localised predominantly in the anterior temporal lobe, as Okada et al. also found posterior activations. Yet, Peelle, Johnsrude, and Davis (Peelle, Johnsrude, & Davis, 2010) - in a response to Okada et al. - reason that Okada et al.'s posterior activations may be partially explained by the fact that they used continuous fMRI scanning (i.e., they presented their stimuli over scanner noise), so background noise was always present when speech stimuli were presented. Peelle et al. further argue that the presence of scanner noise introduces additional task components associated with increased listening effort. Increases in listening effort have in turn been associated with increased BOLD-activations in posterior temporal regions. Peelle et al. base this suggestion on results from an fMRI study by Peelle, Eason, Schmitter, Schwarzbauer, and Davis (2010), in which speech comprehension was investigated in scanning sequences varying in loudness intensity. Here, increases in listening effort were associated with greater blood oxygen level dependent (BOLD)-response in left Supramarginal gyrus and posterior STS.

Peelle et al. (2010) thus refer to a pertinent issue in neuroimaging of spoken language comprehension, i.e., the possibility that specific design choices directly lead to the activation of specific neural areas in the comprehension network. The present study aims to establish the effects of four methodological issues in neuroimaging on the location of the speech comprehension network, namely the choice of scanning

paradigm, the selected non-speech baseline, the type of stimulus material (words or sentences), and the decision to include an experimental task.

Design Choices

When running an fMRI experiment to study speech comprehension, it must be decided whether to present the speech stimulus while the scanner is acquiring functional volumes – i.e., during continuous scanning – or whether to present the stimulus in quiet and collect a single (or multiple) volume(s) after the stimulus has finished playing (a ‘sparse’ temporal sampling design, (Hall et al., 1999)). Continuous and sparse paradigms have specific (dis)advantages. A continuous scanning paradigm has the disadvantage that speech stimuli are presented over the noise associated with the echoplanar imaging (EPI) sequence commonly used in fMRI studies. The rapid gradient switching in the EPI sequence results in oscillations of the scanner coils that can create sound waves that may exceed 90 dBA (cf., Peelle et al., 2010). This noise in EPI sequences is problematic for speech research, for several reasons. First, the presence of EPI-related noise results in acoustic masking of the auditory sentences (Shah, Jäncke, Grosse-Ruyken, & Muller-Gartner, 1999). Second, in a regions-of-interest (ROI) -based study, Bandettini, Jesmanowicz, Van Kylen, Birn, and Hyde (1998) focused on auditory cortex and found evidence of EPI-related contamination of the BOLD signal response. In addition, the presence of noise negatively affects the comprehension process: speech understanding in poorer signal-to-noise (SNRs) conditions is slower and less efficient than speech in better SNRs (Kalikow, Stevens, & Elliott, 1977). Sparse designs do not have the problems associated with EPI-related noise during stimulus presentation. However, sparse designs usually have considerably longer repetition times (of up to 12 seconds) than continuous studies, as sparse designs allow a silent interval of often up to several seconds between

acquisitions of successive volumes. Consequently, fewer volumes can be collected within a given timeframe compared with continuous data acquisition. This means that fewer data points are available in a single experiment, and the experiment usually needs to be lengthened to acquire all necessary data. The choice between continuous and sparse sampling is thus based on a trade-off between the ability to reliably detect changes in the BOLD signal and experimental duration. The present study aims to establish the effects of either continuous or sparse designs on the neural network of speech comprehension.

Second, one should consider the non-speech baseline stimulus. This baseline is very often ‘rest’, which implies that, in continuous designs, speech comprehension in the presence of scanner noise is compared with the sound of scanner noise alone, while sparse designs speech compare comprehension in quiet with a condition of relative quiet. A disadvantage of using rest as a baseline is that the network associated with listening to speech versus rest also includes areas involved with processes other than speech intelligibility, such as acoustic and phonetic/phonological processing. Instead, it has been suggested (Scott, et al., 2000) to use an acoustically matched baseline, i.e., a signal that physically resembles the speech stimulus as much as possible but lacks its intelligibility. The present study defines an acoustically matched baseline as a signal that is closely matched to the speech stimulus material without being intelligible and extends this to include less intelligible signals, such as speech in noise or time-compressed speech, as well as more intelligible signals lacking meaning, such as pseudowords. A variety of matched baseline stimuli have been used: pitch-shifted speech, time-reversed speech, spectrally-rotated speech, noise-vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), low-pass filtered speech, time-reversed speech, signal-correlated noise (Schroeder, 1968),

pseudowords, speech in noise, and musical rain (Zaehle, Geiser, Alter, Jancke, & Meyer, 2008). In noise-vocoded speech, only the temporal characteristics of the signal are preserved, while spectral information is mostly absent (although noise vocoded speech with more than one channel may allow listeners to build up a profile of spectral information across channels), as the excitation source has been replaced by white noise. The resulting speech sounds like a harsh whisper. Noise-vocoded speech is created by feeding the signal through a filter bank in a channel vocoder (Flanagan, 1972). The intelligibility of the output signal is determined by the number of channels (or bands) that the original signal is filtered through. Spectrally rotated speech involves inversion of the spectral characteristics of the speech stimulus. Rotated speech represents a close preservation of the (phonetic-) acoustic form while preventing comprehension of (aspects of) the stimulus (Blessner, 1972). Signal-correlated noise (Schroeder, 1968), stimuli have the same spectral profile and amplitude envelope as the original speech signal, but all spectral detail is replaced with noise, ensuring that the signal-correlated signal is unintelligible. The long-term spectrotemporal characteristics of musical rain stimuli are matched to that of corresponding speech stimuli and the temporal envelope of each segment is modulated by the temporal envelope extracted from the corresponding speech. Musical rain does not sound like speech; yet produces a similar level of BOLD activation to that of speech stimuli in primary auditory and association cortices (Zaehle, et al., 2008).

Third, the experimenter has to decide whether to include a task or not. A drawback of using a task is that the participant's response may influence BOLD-activity in areas under research, such as (pre-)motor cortex, and the task may direct attention away from comprehension of the speech stimulus. Also, an overt (hand)

motor response could compromise the ability to uncover speech-relevant activations in motor cortex. An advantage of including a task is that it may raise participants' attention levels and that it is possible to assess to which extent the stimulus material was correctly processed.

Finally, studies on speech comprehension have either focused on identifying areas associated with processing spoken words or spoken sentences. It can be argued that sentences represent a more valid ecological stimulus than words, as they represent speech materials that listeners most often encounter in daily life. Also, recent studies show that sentences may be more effective than words in eliciting BOLD-activations compared to baseline (Awad, Warren, Scott, Turkheimer, & Wise, 2007; Xu, Kemeny, Park, Frattali, & Braun, 2005). Studies using words and sentences may have slightly different aims and target different levels in the linguistic hierarchy. Studies using word stimuli target processing at lexico-semantic levels and investigate processes such as lexical access (Kotz, Cappa, Von Cramon, & Friederici, 2002; Raettig & Kotz, 2008) or semantic priming (Gagnepain et al., 2008), while studies focusing on sentence-length stimuli may also target syntactic (Tyler et al., 2010) and pragmatic processing, for instance by investigating disambiguation processes at sentence level (Rodd, et al., 2005; Rodd, Longe, Randall, & Tyler, 2010).

Groups of researchers select specific combinations of these four design options. For instance, some investigate comprehension using a sparse temporal sampling designs, sentence stimuli, an acoustically matched baseline, and no task (Narain, et al., 2003; Obleser, et al., 2007; Scott, et al., 2000; Scott, et al., 2006), while others use the same combination, but include a task (Davis, Ford, Kherif, & Johnsrude, 2011; Davis & Johnsrude, 2003; Rodd, et al., 2005; Rodd, et al., 2010). The present study seeks to elucidate the effect of such (combinations of) design choices on the

intelligibility network using the Activation Likelihood Estimation (ALE) method (Laird et al., 2005; Turkeltaub, Eden, Jones, & Zeffiro, 2002). ALE can be used to determine the overlap between coordinates obtained from neuroimaging studies by modelling them as probability distributions centred at the reported coordinates. The meta-analysis applies ALE to coordinates collected from neuroimaging studies comparing comprehension of intelligible speech at the post-lexical level with less intelligible or unintelligible stimuli. This meta-analysis is used to establish the effects of four design choices in studying intelligibility using fMRI and PET on areas consistently activated during speech comprehension.

2. Method

2.1.1 Selection of studies

Neuroimaging studies were selected that investigated speech comprehension at post-lexical levels using Pubmed. The Pubmed online database was searched for papers using the keywords: “speech”, “intelligibility”, “intelligible”, “comprehension”, “fmri”, “post-lexical”, “narrative”, “word”, “PET”, “neuroimaging”, and appropriate combinations of these words. Additional papers were collected by searching for prominent researchers in the field. Papers were selected from January 1, 2000 onwards up to March 29 2012.

2.1.2 Inclusion and exclusion criteria

Experiment were included that fulfilled the following criteria: i) neural responses were collected using fMRI or PET, ii) only healthy, neurotypical subjects with intact hearing and no known neurological or psychiatric disorders were tested², iii) the

² Two papers were selected in which several left-handed participants had been included in the data set Christensen, Antonucci, Lockwood, Kittleson, & Plante, (2008) and Tyler, Shafto, Randall, Wright, Marslen-Wilson, & Stamatakis. (2010). Authors of both papers verified that neural responses of the left-handed participants did not deviate from responses of the right-handed participants.

experiments contained conditions in which intelligible speech was presented as well as conditions in which less intelligible or unintelligible speech was presented, iv) the intelligible speech consisted of stimuli at the post-lexical level, including words, sentences, paragraphs, 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.

The following criteria were used to exclude papers from the analysis: single-subject studies and studies that report only the results from pre-specified ROIs, or papers that presented intelligibility results for multimodal (e.g., audiovisual) stimuli.

Most papers in Table I included a single experiment, except Devauchelle, Oppenheim, Rizzi, Dehaene, and Pallier (2008), Mitchell, Elliott, Barry, Cruttenden, and Woodruff (2003), Rodd, Johnsrude, and Davis (2005), and Schön et al. (2010). Results from experiment 1 were included for Devauchelle et al., as their experiment 2 included visual stimuli only, and experiment 3 was behavioural. Mitchell et al. included six experiments, all run on the same group of participants. The results from experiment 3 were included (“regions activated by neutral prosody vs. rest”). These results were included as the other five experiments focused to a greater extent on identifying brain regions involved in processing prosody, and not intelligibility specifically. For Schön et al., the results from experiment 1 were included, as no coordinates for intelligible speech vs. unintelligible speech were listed for experiment 2. For Rodd et al., the combined results from experiments 1 and 2 were included. Finally, Tyler et al. (2010) tested a group of younger and a group of older listeners. Their two groups were included as separate studies in the meta-analysis, as Tyler et al. do not include a set of coordinates based on the results from both groups, as was the case in Rodd et al. (2005).

2.2 ALE methods

The ALE analysis was implemented using GingerALE 2.1 (www.brainmap.org). This version of the likelihood estimation algorithm was selected, as it allows for subtraction analysis, using the method described in Eickhoff, Bzdok, Laird, Kurth, and Fox (2011). Coordinates from studies that reported coordinates in Talairach space were converted to MNI space using the *tal2icbm_spm* algorithm implemented in GingerALE (www.brainmap.org/ale).

In GingerALE, first, modelled activation maps are computed for each set of foci for each included experiment. 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 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: $q < 0.05$), using a cluster extent of 400mm^3 (default: 200mm^3).

3. Results

3.1. ALE Comprehension Network

The selected studies (Table I) were based on the results from 57 experiments (58 if the foci from Rodd et al., 2005, are counted as coming from two experiments), 898 participants, and 513 foci, published in 57 papers. The distribution of the studies of all

studies over combinations of the four design options is presented in Table II. All experiments used continuous or sparse scanning, or both (Peelle, Eason, et al., 2010). Some used a low-level or acoustically matched baseline, while others used rest or an implicit baseline (i.e., scanner noise in continuous designs). Stimuli were words, sentences, stories or narratives. Some studies used a task, while others employed passive listening. The following tasks were used: gap detection, grammatical tense decision, semantic decision, gender decision, intelligibility judgment, button press after stimulus presentation, 2AFC task (two alternative forced choice), 4AFC task, anomaly judgment, similarity judgment, emotion judgment, prosody judgment, gender decision, repetition decision, emotion judgment, and word detection in the stimulus, or indicate if a post-stimulus probe was present in the stimulus or not.

Thirteen types of matched baseline stimuli were used: pitch-shifted speech, time-reversed speech, spectrally-rotated speech, noise-vocoded speech (8, 4 bands, and 1 band), low-pass filtered speech, spectrally rotated and time-reversed speech, signal-correlated noise, pseudowords, pseudoword sentences, speech in noise, musical rain, prosodic speech (i.e., unintelligible sentences containing only speech-melodic information such as the intonation contours and amplitude envelope, see Meyer, Alter, Friederici, Lohmann, and von Cramon (2002), and repetitions of different versus identical sentences. Finally, five types of non-matched baselines were used: scanner noise, tones, environmental sounds, pink noise, and white noise.

--- Tables I - II about here ---

Before evaluating the effect of the design choices, the reference comprehension network (Figure 1) was outlined. Analysis of the 57 experiments in Table I resulted in six clusters of significant ALE peaks (Table III). The highest ALE values were found in left anterior Superior Temporal Sulcus (STS), extending into Middle Temporal

Gyrus (MTG), STG, pars opercularis (POp) and par Orbitalis (POrb), precentral gyrus, ventral premotor cortex (PMv) and Angular Gyrus (AG). Second, a smaller homologous cluster was found in right anterior STS, extending into MTG, STG and right insula. A third cluster was found in the Supplementary Motor Area (SMA) extending anteriorly into pre-SMA. A fourth cluster was found in left fusiform gyrus, a fifth cluster in left PMv, and a sixth in left precentral gyrus. All peaks were based on at least four contributing studies.

--- Table III and Figure 1 about here ---

3.2 Design choices

GingerALE 2.1 was used to conduct four subtraction analyses (sparse vs. continuous paradigm, matched vs. non-matched baseline, passive listening vs. use of a task, and word-length vs. sentence-length stimuli). The same significance levels were adhered to as in the ALE analysis in 3.1.

3.2.1. Sparse vs. Continuous Scanning Paradigm

The studies were split into groups depending on whether they used a sparse temporal scanning paradigm (19 experiments, 201 foci) or a continuous scanning paradigm (34 experiments, 276 foci). Crinion et al. (2003), Scott et al. (2000), and Scott et al. (2006) were excluded from this analysis as PET was used, as was Peelle et al. (2010), as they combined the coordinates for their sparse and continuous scanning protocols. GingerAle reported four clusters for the contrast Sparse > Continuous (Table IV and Figure 2). Studies using a sparse scanning paradigm reported higher ALE scores in bilateral in anterior MTG/STG/STS, left posterior STG, and left AG. One cluster was found for the contrast Continuous > Sparse in Anterior Cingulate Gyrus (ACG) extending to pre-SMA/SMA.

--- Table IV and Figure 2 about here ---

3.2.2. Matched vs. Non-Matched Baseline

The experiments were divided into two groups depending on whether they used an acoustically matched baseline (37 experiments, 316 foci) or a non-matched baseline (20 experiments, 197 foci). Using a matched baseline resulted in more activity in left anterior STS, extending into left MTG/STG, than studies using a non-matched baseline. Studies using a non-matched baseline show more activation in two clusters in left Precentral Gyrus/MFG.

3.2.3. Task vs. Passive Listening

The studies were split into groups depending on whether they used a task (35 experiments, 347 foci) or whether no task (passive listening) was used (21 experiments, 161 foci). No significant clusters were found for the contrasts Task > Passive listening and Passive listening > Task. Lowering the cluster-level threshold to 200mm^3 did not result in any significant clusters.

3.2.4. Words vs. Sentences

The studies were split into groups depending on whether they used words (18 experiments, 162 foci) or sentences (34 experiments, 323 foci). Two significant cluster was found for the contrast Sentences > Words spanning, one in left POp, one in left Pars Triangularis (PTr) extending medially into the insula. No significant clusters were found for the contrast Words > Sentences. Lowering the cluster-level threshold to 200mm^3 did not result in any significant clusters.

3.2.5. Core comprehension network

The thresholded maps associated with all six ALE analyses (i.e., sparse studies, continuous studies, matched-baseline studies, non-matched baseline studies, studies using words, studies using sentences) were superimposed on each other using Mango's logical (binary) overlay tool. This resulted in the selection of a set of areas

present in all six ALE-maps, or the ‘core’ comprehension network, i.e., a network of areas assumed to be unaffected by the six design choices. This core network is outlined in Figures 1 and 2.

4. Discussion

4.1 Core network and selective recruitment

The ALE comprehension network in Table III includes bilateral anterior STS, extending into MTG, STG, left IFG (POp and POrb), left PMv, left AG, pre-SMA/SMA, left Fusiform Gyrus, and Left Precentral Gyrus. The ALE network is thus more widespread in the left hemisphere and extends into frontal and parietal regions, while it is restricted to the temporal lobe in the right hemisphere. The smaller core network in Figures 1-2 is restricted to the temporal lobes only.

ALE clusters in left STS/MTG anterior and posterior to the core network and areas anterior to the core network in right STS/STG/MTG were driven by studies using sparse scanning, while the midline clusters in SMA/pre-SMA were driven by studies using continuous scanning. Midline structures including SMA/pre-SMA have been associated with speech *production* processes (Alario, Chainay, Lehericy, & Cohen, 2006, Soros et al., 2006; Shuster, 2009; Shuster & Lemieux, 2005), including lexical selection, linear sequence encoding, and control of motor output (Alario, et al., 2006). It has been suggested that speech motor information may be forwarded to the speech comprehension network to resolve ambiguities in the speech signal (Pickering & Garrod, 2007; Skipper, Nusbaum, & Small, 2006). Indeed, midline activations during speech comprehension have been associated with speech comprehension under difficult listening conditions (e.g., the presence of background noise, or listening to multiple speakers) (Adank & Devlin, 2010; Davis & Johnsruide, 2003; Wong, Uppanda, Parrish, & Dhar, 2008). Speech comprehension under challenging listening

conditions may rely in part on recruitment of specific speech motor areas, including SMA (Adank, 2012). Adank performed two ALE meta-analyses: one on speech comprehension studies including a less intelligible > more intelligible contrast (e.g., time-compressed sentences versus normal-speed sentences,), and one focusing on studies evaluating speech production. Adank found that the network for effortful speech processing was located mostly outside the network for processing undistorted speech signals, and that effortful processing most likely relies on selective recruitment of cortical areas related to cognitive processes such as working memory, as well as those related speech production processes, particularly SMA. Selective recruitment during effortful speech comprehension may explain increased activation of pre-SMA/SMA when scanner noise is present (as is the case during continuous acquisition).

The results of the contrast Sentences > Words may point to another instance of such selective recruitment: experiments using sentences showed higher ALE values in left POp and PTr than experiments using words. Recruitment of areas in IFG for processing sentence-length stimuli has been reported on various occasions (Mazoyer et al., 1993; Xu, et al., 2005). Furthermore, it has been argued that POp is involved in syntax processing (Dapretto & Bookheimer, 1999; Newman, Just, Keller, Roth, & Carpenter, 2003) while PTr is associated with semantic integration processing (Newman, et al., 2003; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) in sentence comprehension. Based on this view, it could be expected that left IFG would be part of the core network of speech comprehension. Nevertheless, this expectation was not borne out in the present results. In contrast, it has been proposed that IFG is not an essential part of the comprehension network (Hickok & Poeppel, 2007; Meyer, et al., 2002; Rauschecker & Scott, 2009; Scott, et al., 2000; Scott, McGettigan, &

Eisner, 2009; Hickok, et al. 2011). The results of the current analysis support this view, as the core network did not include inferior frontal areas.

However, before it can be concluded that inferior frontal areas are not an integral part of the core comprehension network, two aspects of the current study should be examined. The first of these was the decision to include studies that used pseudowords as a matched baseline. Pseudowords deviate from other matched baselines in Table I, as they do not represent speech that is degraded or unintelligible. When processing pseudowords, the listener perceives a clear speech signal that does not allow for lexical access to familiar representations. The inclusion of pseudowords could thus have skewed the results towards lexical access processing, which may not necessarily involve left inferior frontal areas (e.g., Kotz, Cappa, von Cramon, & Friederici, 2002). This possibility was examined by repeating the ALE intelligibility network without the eight pseudowords studies. The results (displayed in Table S1) showed a lower ALE score for a sub-cluster in left POp and the distribution of ALE clusters is similar to the network described in Table III. It therefore seems unlikely that the absence of left inferior frontal areas in the ALE network is due to the inclusion of pseudoword studies. A second issue is whether the inclusion of studies using words -regardless of the type of non-speech baseline employed- could have reduced the significance of left inferior frontal clusters. Therefore, the core network in figure analysis was mapped out again (areas showing overlap regardless of the baseline, scanning paradigm, or the presence of a task), excluding the 18 studies examining word comprehension. The core network showed no overlap in inferior frontal areas when words were excluded (nor was overlap found in any area outside the core network in Figures 1-2), indicating that inferior frontal areas may not be part of the core comprehension network.

[Insert Inline Supplementary Table SI about here]

It has been postulated (Rauschecker & Scott, 2009) that inferior frontal areas, particularly IFG, may be involved in resolving what is known as “the lack of invariance problem” in speech comprehension (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), i.e., how does the comprehension system map real-life pronunciation of speech sounds (including co-articulation, or differences between and within speakers’ pronunciations) and words onto canonical and abstract forms in the mental lexicon. Results from the present study suggest that left IFG may also be recruited for general effortful linguistic processing problems, including syntactic complexities or semantic ambiguities, as well as processing acoustically distorted speech (Adank, 2012; Skipper, et al., 2006).

Studies using a matched baseline drove clusters anterior to the core network in left STS. Studies focusing on intelligibility processing, (i.e., McGettigan, et al., 2012; Narain, et al., 2003; Scott, et al., 2000; Scott, et al., 2006), report activations located anterior to the core network in Figures 1-2 in left STS/STG. This more anterior position is in agreement with the proposal that left anterior temporal lobe as is involved in more form-independent processing of speech (Davis & Johnsrude, 2003). This (more) anterior location associated with the aforementioned studies may be attributable to the use of fMRI. Only two studies used PET instead of fMRI (Scott, et al., 2000; Scott, et al., 2006). A limitation of fMRI, but not PET, is its reduced ability to reliably detect BOLD changes in (anterior) temporal regions (Devlin et al., 2000). Using controlled materials and a single type of stimulus degradation involve (more anterior) areas specific to intelligibility processing. Nevertheless, as the majority of studies included in the meta-analysis used fMRI (54) and only a minority used PET (3), contributions of anterior temporal regions may not be entirely accurately

represented in the present meta-analysis. Using distortion correction (e.g., Embleton, Haroon, Morris, Lambon-Ralph, & Parker, 2010) in future fMRI studies may uncover increased contributions of areas anterior to the core network involved in form-independent processing. Note that the matched baselines used in the studies included in this analysis were not only matched an acoustic level, but also very similar at phonetic-phonological levels, which may have cancelled out activations associated with pre-lexical levels of processing. This may have contributed to the absence of more posterior areas, which have been proposed to be specifically involved with this type of phonological processing (Hickok & Poeppel, 2007). Also, using a non-matched baseline did not result in increased activations in the right temporal lobe, as predicted, for instance, by Narain et al. (2003). Instead, studies using a non-matched baseline resulted in higher ALE scores in left Precentral/Postcentral gyrus. These activations may be due to biases towards specific combinations of design choices; e.g., of the 20 studies employing a sparse design, 18 used a matched baseline, while 18 of the 34 studies employing a continuous scanning paradigm used a non-matched baseline (thus requiring speech comprehension against a background of scanner noise). Adank (2012) reports left precentral gyrus as part of the network of areas recruited for such effortful speech comprehension. However, further research is required to untangle the effects of continuous scanning and the choice of the non-speech baseline, for instance in an fMRI experiment varying these two factors in a factorial design.

Finally, no ALE clusters were found for the contrast Task vs. Passive Listening. This result runs counter to the expectation that using a task contaminates responses in (speech) motor cortex.

4.2 Location of speech comprehension

The ALE-analysis first support those proposing that speech comprehension is bilaterally organised (Hickok, et al., 2011; Hickok & Poeppel, 2007; Okada, et al., 2010), although it should be mentioned that the network was more extensive in the left hemisphere. Second, the ALE-results support a more anterior centre of gravity of the comprehension network (Davis & Johnsrude, 2003; Narain et al., 2003; Rauschecker & Scott, 2009; Scott, et al., 2000; Scott, et al., 2006).

Hickok & Poeppel's primary argument is for a bilaterally organized network, which was supported by both the ALE analysis as well as the placement of the core network. Moreover, their model predicts that posterior MTG and ITG serve as a lexical interface. This lexical interface is proposed to form part of the ventral stream and map phonological forms in the posterior STS region onto a distributed semantic network. The present results imply that this lexical interface may be located more anterior than originally proposed in the model. No role is allocated in Hickok & Poeppel's model to other areas in the ALE network: SMA or pre-SMA, left fusiform gyrus, or left AG. Finally, the model proposes that a region on left posterior STG (area Spt) connects the speech comprehension system to the speech production system, however, the present analysis found no evidence that this area is consistently activated during comprehension. Nevertheless, it may be case that area Spt is involved primarily at pre-lexical processing levels, as also postulated in Hickok and Poeppel's model.

The ALE analysis provides partial support for Rauschecker and Scott's model as it shows the involvement of left AG, left anterior STS, left STG, and left IFG, and its centre of gravity of the network is placed anterior to Heschl's Gyrus. Yet, except for left STS and STG, none of the areas proposed in Rauschecker and Scott's model outside the temporal lobe are part of the core network or speech comprehension.

4.3 Suggestions for future study

The results suggest that spoken language comprehension, especially at lexical levels, is located in a core network in bilateral anterior-to-mid STS, extending to STG and MTG. Any areas outside this core network may be selectively recruited for addressing specific problems or levels of linguistic complexity. For instance, the results suggest that POp and PTr are recruited for processing the added complexity involved in processing sentence length stimuli, and that pre-SMA/SMA and left precentral gyrus may be recruited for effortful processing, for instance of distorted speech signals, and previous studies have shown that these areas are also recruited for adaptation processes during comprehension (Adank & Devlin, 2010; Eisner, et al., 2010). Future research may focus on elucidating on the specific role of areas in the wider comprehension network, and on their interactions (Obleser, et al., 2007). It is suggested that fMRI studies use a combination of a sparse temporal scanning paradigm and an acoustically matched baseline³, as these two combinations did not drive ALE clusters outside the temporal lobes and do not require the listeners to process speech under adverse listening conditions. Alternatively, continuous scanning could be used if the design allowed for identification of areas associated with listening effort or increased reliance on executive functioning (e.g., by including parametrically varying background noise levels or degradation levels), so that areas associated with listening effort can be identified (and accounted for).

Finally, it seems relevant to use the appropriate level of participants to ensure sufficient statistical power to detect BOLD signal changes. The average number of participants did not differ (established through three paired samples *t*-tests, $p > .2$)

³ This suggestion will have implications for studies investigating the neural bases of effortful processing, e.g., accented speech > noise-vocoded speech. Note that such comparisons may show different sites of activations for processing intelligible speech compared to studies using continuous scanning, as sites of activation vary depending on the type of adverse listening condition used (e.g., Adank, Davis & Hagoort, 2012).

across the sub-analyses: this was 15.2 (SD 5.2) for sparse scanning, 15.3 (SD 6.0) for continuous scanning, 15.2 (SD 6.5) for a matched baseline, 16.1 (SD 5.8) for a non-matched baseline, 15.7 (SD 4.7) for experiments involving a task, and 15.4 (SD 6.9) for experiments using passive listening. The present results can therefore not be explained solely by power differences between design choices. Yet, it is a concern that the average number of participants per experiment is overall relatively low. Desmond & Glover (2002) proposed a minimum of 20 participants for fMRI experiments, which suggests that the majority of experiments (i.e., 34 out of 42) included in this meta-analysis may actually have been underpowered (unless the lower number of subject was compensated for by a longer experimental duration, as suggested by Desmond & Glover).

In sum, the present results indicate that activation of areas outside the core comprehension network in bilateral anterior STS may be a consequence of the choice of scanning paradigm, type of baseline, or the type of stimulus material used. It is proposed that future research explicitly addresses the effects of design choices on the activation of areas outside the core network identified in the present research.

Acknowledgements

I wish to thank Ajay Halai, Jonathan Peelle, Matt Davis, Peter Turkeltaub and two anonymous reviewers for useful and interesting discussions and/or comments on the manuscript.

References

- Adank, P. (2012). The Neural Bases of Difficult Speech Comprehension and Speech Production and Their Overlap: Two Activation Likelihood Estimation (ALE) Meta-analyses. *Brain Lang.*, 122(1), 42-54.
- Adank, P., Davis, M., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, 50(1), 77-84.
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, 49(1), 1124-1132.
- Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Res.*, 1076(1), 129-143.
- 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. *Hum. Brain Mapp.*, 26(44-53).
- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A common system for the comprehension and production of narrative speech. *J. Neurosci.*, 27(43), 11455-11464.
- Bandettini, P. A., Jesmanowicz, A., van Kylen, J., Birn, R. M., & Hyde, J. S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magn. Reson. Med*, 39(3), 410-416.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex*, 10(), 512-528.

- 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.
- Blessner, B. (1972). Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J. Speech Hear. Res*, 15, 5-41.
- Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bihemispheric foundations for human speech comprehension. *PNAS*, 107(40), 17439-17444.
- Christensen, T. A., Antonucci, S. M., Lockwood, J. L., Kittleson, M., & Plante, E. (2008). Cortical and subcortical contributions to the attentive processing of speech. *Neuroreport*, 19(11), 1101-1105.
- Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., et al. (2004). Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage*, 22(1), 11-21.
- 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(5), 1193-1201.
- Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, 24, 427-432.
- language comprehension. *J. Neurosci.*, 23(8), 3423-3431.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.*, 23(8), 3423-3431.
- 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. *J. Cogn. Neurosci.*, 23(12), 3914-3932.

- Desmond, J. E., & Glover, G. H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J. Neurosci. Methods*, 118(2), 115-128.
- Devauchelle, A., Oppenheim, C., Rizzi, L., Dehaene, S., & Pallier, C. (2008). Sentence Syntax and Content in the Human Temporal Lobe: An fMRI Adaptation Study in Auditory and Visual Modalities. *J. Cogn. Neurosci.*, 21(5), 1000-1012.
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *Neuroimage*, 11(6 Pt 1), 589-600.
- Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, 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. *J. Cogn. Neurosci.*, 19(5), 799-816.
- Eickhoff, S. B., Bzdok, D., Laird, A., Roski, C., Caspers, S., Zilles, K., et al. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *NeuroImage*, 57, 938-949.
- Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Phil. Trans. R. Soc. A: Math. Phys. Engineer. Sci*, 367(1896), 2399-2421.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *J. Neurosci.* 30(21), 4040-4049.
- Embleton, K. V., Haroon, H. A., Morris, D. M., Lambon-Ralph, M. A., & Parker, G. J. (2010). Distortion Correction for Diffusion-Weighted MRI Tractography and fMRI in the Temporal Lobes. *Hum. Brain Mapp.*, 31, 1570-1587.

- Flanagan, J. L. (1972). *Speech analysis, synthesis and perception* (2nd ed.). Berlin: Springer-Verlag.
- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum. Brain Mapp.*, 31(3), 448-457.
- Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex*, 13(2), 170-177.
- Gagnepain, P., Chetelat, G., Landeau, B., Dayan, J., Eustache, F., & Lebreton, K. (2008). Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. *J. Neurosci.*, 28(20), 5281-5289.
- Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb. Cortex*, 14(3), 247-255.
- Gonzalez-Castillo, J., & Talavage, T. (2011). Reproducibility of fMRI activations associated with auditory sentence comprehension. *NeuroImage*, 54(3), 2138-2155.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliot, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.*, 7, 213-223.
- Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *J. Cogn. Neurosci.*, 20, 1737-1752.

- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69, 407-422.
- Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, 49(1), 875-884.
- Ischebeck, A. K., Friederici, A. D., & Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cereb. Cortex*, 18(3), 541-552.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.*, 23(29), 9541-9546.
- Jardri, R., Pins, D., Bubrovszky, M., Desprez, P., Pruvo, J. P., Steinling, M., et al. (2007). Self awareness and speech processing: an fMRI study. *NeuroImage*, 35(4), 1645-1653.
- Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.*, 61(5), 1337-1351.
- Klein, E., Moeller, K., Nuerk, H. C., & Willmes, K. (2010). On the neuro-cognitive foundations of basic auditory number processing: an fMRI study. *Behav. Brain Func.*, 6, 42.
- Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. *NeuroImage*, 17(4), 1761-1772.
- Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Bilingual and monolingual brains compared: a functional magnetic resonance imaging investigation of syntactic

- processing and a possible "neural signature" of bilingualism. *J. Cogn. Neurosci.*, 20(1), 153-169.
- 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. *Hum. Brain Mapp.*, 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. *J. Neurosci.*, 28(49), 13209-13215.
- Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.*, 74, 431-461.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb. Cortex*, 15, 1621-1631.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.*, 8, 393-402.
- LoCasto, P. C., Krebs-Noble, D., Gullapalli, R. P., & Burton, M. W. (2004). An fMRI investigation of speech and tone segmentation. *J. Cogn. Neurosci.*, 16(9), 1612-1624.
- Love, T., Haist, F., Nicol, J., & Swinney, D. (2006). A functional neuroimaging investigation of the role of structural complexity and task-demand during auditory sentence processing. *Cortex*, 42(4), 577-590.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., & Levrier, D. (1993). The cortical representation of speech. *J. Cogn. Neurosci.*, 5, 467-479.
- McGettigan, C., Evans, S., Rosen, S., Agnew, Z. K., Shah, P., & Scott, S. K. (2012). An application of univariate and multivariate approaches in FMRI to quantifying

- the hemispheric lateralization of acoustic and linguistic processes. *J. Cogn. Neurosci.*, 24, 636-652.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.*, 17(2), 73-88.
- Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, 41(10), 1410-1421.
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb. Cortex*, 13(12), 1362-1368.
- Newman, S. D., Just, M. A., Keller, T. A., Roth, J., & Carpenter, P. A. (2003). Differential effects of syntactic and semantic processing on the subregions of Broca's area. *Cog. Brain Res.*, 16, 297-307.
- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex*, 20, 633-640.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.*, 28(32), 8116-8123.
- 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. *J. Neurosci.*, 27, 2283-2289.

- Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. *NeuroReport*, 17, 1293-1296.
- 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. *Cereb. Cortex*, 20(10), 2486-2495.
- Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *J. Cogn. Neurosci.*, 18, 1237-1252.
- 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. *Front. Hum. Neurosci.*, 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 Lang.*, 91, 315-325.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.*, 11(3), 105-110.
- Raettig, T., & Kotz, S. A. (2008). Auditory processing of different types of pseudowords: An event-related fMRI study. *NeuroImage*, 39, 1420-1428.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.*, 12(6), 718-724.

- 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.
- Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An event-related FMRI investigation of implicit semantic priming. *J. Cogn. Neurosci.*, 15(8), 1160-1175.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. 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(5), 1324-1335.
- Röder, B., Stock, O., Neville, H., Bien, S., & Rosler, F. (2002). Brain Activation Modulated by the Comprehension of Normal and Pseudo-word Sentences of Different Processing Demands: A Functional Magnetic Resonance Imaging Study. *NeuroImage*, 15, 1003-1014.
- Schirmer, A., Zysset, S., Kotz, S. A., & Yves von Cramon, D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, 21(3), 1114-1123.
- Schön, 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. *J. Acoust. Soc. Am.*, 44, 1735-1736.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.*, 26, 100-107.

- Scott, S. K., Blank, S. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400-2406.
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little less action a little more conversation — candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.*, 10, 295-302.
- 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. *J. Acoust. Soc. Am.*, 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.
- Shah, N. J., Jäncke, L., Grosse-Ruyken, M. L., & Muller-Gartner, H. W. (1999). Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. *J. Magn. Res. Im*, 9(1), 19-25.
- 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. S., Vigliocco, G., & Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Hum. Brain Mapp.*, 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 Lang.*, 93(1), 20-31.

- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *NeuroImage*, 25, 76-89.
- 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, MA: 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.
- Takeichi, H., Koyama, S., Terao, A., Takeuchi, F., Toyosawa, Y., & Murohashi, H. (2010). Comprehension of degraded speech sounds with m-sequence modulation: an fMRI study. *NeuroImage*, 49(3), 2697-2706.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A re-evaluation. *PNAS*, 94, 14792-14797.
- Tracy, J., Flanders, A., Madi, S., Natale, P., Delvecchio, N., Pyrros, A., et al. (2003). The brain's response to incidental intruded words during focal text processing. *NeuroImage*, 18(1), 117-126.
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain Lang.*, 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. *Cereb. Cortex*, 20(2), 352-364.
- 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. *Cereb. Cortex*, 14(12), 1384-1389.
- Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex*, 18(1), 230-242.
- Wong, P. C. M., Uppanda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical Mechanisms of Speech Perception in Noise, *J. Speech Hear. Lang. Res.*, 51, 1026-1041.
- Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., et al. (2005). Differential Activity in Left Inferior Frontal Gyrus for Pseudowords and Real Words: An Event-Related fMRI Study on Auditory Lexical Decision. *Hum. Brain Mapp.*, 25, 212-221.
- Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *NeuroImage*, 25(3), 1002-1015.
- Zaehle, T., Geiser, E., Alter, K., Jancke, L., & Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. *Brain Res.*, 1220, 179-190.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32(4), 1826-1836.

Figure Captions

Figure 1

ALE clusters for the 57 studies in the meta-analysis. ALE scores are indicated in the legend. The solid red line outlines the areas within the speech intelligibility network activated across all eight groupings of design choices, i.e., the core comprehension network.

Figure 2

Figure 2. A: ALE clusters for the contrasts Matched > Non-matched baseline and Non-matched > Matched baseline, plus the core speech comprehension network (cf. Figure 1); B: ALE clusters for the contrasts Continuous > Sparse scanning paradigm and Sparse > Continuum scanning paradigm, plus the core speech comprehension network; C: ALE clusters for the contrast Sentences > Words, plus the core speech comprehension network.

Table I. Experiments included in the main meta-analysis on spoken language comprehension, SNR: signal to noise ratio, * = no info given in paper.

#ID	Study	Method	N	Language	Task	Stimulus	Paradigm	Design	Contrast	Source	Foci	FWHM (mm)
<i>Matched baseline</i>												
1	(Allen et al., 2005)	fMRI	11	British-English	semantic decision	sentences	sparse	block	normal sentences>pitch-shifted sentences	Table III	1	7.2
2	(Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010)	fMRI	12	British-English	gap detection	words	sparse	event-related	words>musical rain	Table S2	21	10
3	(Christensen, Antonucci, Lockwood, Kittleson, & Plante, 2008)	fMRI	14	American-English	gender decision + semantic decision (combined)	words	continuous	block	words>time-reversed words	Table I (“Dichotic listening”)	10	4
4	(Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003)	PET	17	British-English	passive listening	narratives	N/A	block	narratives>time-reversed narratives	Table I	11	10
5	(Davis & Johnsrude, 2003)	fMRI	12	British-English	intelligibility judgment	sentences	sparse	event-related	positive intelligibility correlation	Table 1	8	12
6	(Davis, Ford, Kherif, & Johnsrude, 2011)	fMRI	13	British-English	word detection	sentences	sparse	event-related	positive intelligibility correlation	Table 1	26	10
7	(Devauchelle, Oppenheim, Rizzi, Dehaene, & Pallier, 2008)	fMRI	21	French	button press	sentences	continuous	block	identical sentences>different sentences	Table 3	4	5
8	(Friederici, Kotz, Scott, & Obleser, 2010)	fMRI	17	German	passive listening	sentences	sparse	event-related	sentences>spectrally-rotated sentences	Table I	2	8
9	(Gagnepain et al., 2008)	fMRI	18	French	lexical decision	words	continuous	event-related	words>pseudowords	Table 1	8	8

10	(Giraud et al., 2004)	fMRI	8	German	semantic decision	sentences	continuous	block	sentences>“BBSN2”	Table 1	5	10
11	(Jardri et al., 2007)	fMRI	12	French	passive listening	narrative	continuous	block	narrative>time- reversed narrative	Table 1	7	4
12	(Klein, Moeller, Nuerk, & Willmes, 2010)	fMRI	17	German	passive listening	words	continuous	block	words>pseudowords	Table 3	13	7
13	(Kotz, Cappa, von Cramon, & Friederici, 2002)	fMRI	13	German	lexical decision	words	continuous	event-related	words>pseudowords	Table 3	7	4 seconds
14	(Leff et al., 2009)	fMRI	26	British-English	gender decision	words	continuous	block	words>time-reversed words	Table 1	3	8
15	(McGettigan et al., 2010)	fMRI	20	British-English	passive listening	sentences	sparse	event	sentences>unintelligible sentences	Table 2	15	8
16	(Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002)	fMRI	14	German	Tense judgment (passive vs. active)	sentences	continuous	block	sentences>“prosodic sentences”	Table III	5	4 seconds
17	(Narain et al., 2003)	fMRI	11	British-English	passive listening	sentences	sparse	block	sentences>spectrally rotated and time-reversed sentences	Table 1	3	6
18	(Obleser, Wise, Dresner, & Scott, 2007)	fMRI	16	British-English	passive listening + after task	sentences	sparse	event-related	intelligibility correlation	Table 1	6	8
19	(Obleser, Eisner, & Kotz, 2008)	fMRI	16	German	intelligibility judgment	words	sparse	event-related	positive intelligibility correlation	Table 1	4	8
20	(Obleser & Kotz, 2010)	fMRI	16	German	passive listening + after task	sentences	sparse	event-related	16 bands noise-vocoded sentences>4>1 bands	Table 1	3	8
21	(Obleser, Meyer, & Friederici, 2011)	fMRI	14	German	passive listening	sentences	sparse	event-related	32 bands noise-vocoded sentences>16>8 bands	Table 2	3	8
22	(Okada et al., 2010)	fMRI	20	American-English	intelligibility judgment	sentences	continuous	block	sentences>spectrally rotated and time-reversed sentences	Table 1	14	6
23	(Orfanidou, Marslen-	fMRI	13	British	lexical decision	words	sparse	event-	words>pseudowords	Table 2	33	12

	Wilson, & Davis, 2006)			English				related				
24	(Peelle, McMillan, Moore, Grossman, & Wingfield, 2004)	fMRI	8	American-English	gender decision	sentences	continuous	event-related	80% compressed sentences>low-pass filtered sentences	Table 3	5	8
25	(Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010)	fMRI	6	British-English	semantic decision	sentences	continuous and sparse pooled	block	sentences>signal correlated noise	Table 3	19	10
26	(Roder, Stock, Neville, Bien, & Rosler, 2002)	fMRI	19	German	passive listening	sentences	continuous	block	sentences>pseudoword sentences	Table 5b	6	5
27	(Raettig & Kotz, 2008)	fMRI	16	German	lexical decision	words	continuous	event-related	words>pseudowords	Table 3	4	8
28	(Rissman, Eliassen, & Blumstein, 2003)	fMRI	15	German	lexical decision	words	continuous	event-related	words>pseudowords	Table 2	5	6
29	(Rodd, Davis, & Johnsrude, 2005)	fMRI	15	British-English	semantic decision	sentences	sparse	event-related	sentences (low-ambiguity)>signal correlated noise	Table 3	23	12
30	(Rodd, Longe, Randall, & Tyler, 2010)	fMRI	14	British-English	button press	sentences	sparse	event-related	sentences>signal correlated noise	Table 2	13	8
31	(Scott, Blank, Rosen, & Wise, 2000)	PET	6	British-English	passive listening	sentences	n/a	block	sentences in quiet and intelligible vocoded sentences>spectrally rotated and noise-vocoded sentences	Figure 2	2	10
32	(Scott, Rosen, Lang, & Wise, 2006)	PET	7	British-English	passive listening	sentences	n/a	block	positive intelligibility correlation	Table 1	4	10
33	(Stevenson & James, 2009)	fMRI	11	American-English	2AFC task	words	continuous	event-related	high SNR sentences>low SNR sentences	Table 1	6	6
34	(Takeichi et al., 2010)	fMRI	23	Japanese	passive listening	sentences	sparse	block	sentences>time-reversed sentences	Table 1	5	8
35	(Tyler et al., 2010) (older	fMRI	14	British-	detect target words	sentences	sparse	block	sentences>musical rain	Table 2 ("NP-	4	8

	group)			English						MuR")		
36	(Tyler, et al., 2010) (younger group)	fMRI	44	British- English	detect target words	sentences	sparse	block	sentences>musical rain	Table 3 ("NP- MuR")	4	8
37	(Xiao et al., 2005)	fMRI	14	Chinese	lexical decision	words	continuous	event	words>pseudowords	Table II	4	4
<i>Non-matched baseline</i>												
38	(Adank & Devlin, 2010)	fMRI	18	British- English	semantic decision	sentences	continuous	block	sentences>scanner noise	Table I	11	8
39	(Binder et al., 2000)	fMRI	28	American- English	passive listening	words	continuous	block	words>tones	Table 2	8	2
40	(Binder, Swanson, Hammeke, & Sabsevitz, 2008)	fMRI	26	American- English	semantic decision	words	continuous	block	Sentences presented in semantic decision task>scanner noise	Table 1	3	6
41	(Constable et al., 2004)	fMRI	20	American- English	anomaly decision	sentences	continuous	block	sentences>tones	Table 1	8	- *
42	(Dick et al., 2007)	fMRI	15	Italian	semantic decision	sentences	sparse	block	sentences>environmental sounds	Table 1	19	6
43	(Gonzalez-Castillo & Talavage, 2011)	fMRI	16	American- English	passive listening	sentences	continuous	event- related	sentences>(response and instructions)	Table 3	30	6
44	(Holle, Obleser, Rueschemeyer, & Gunter, 2010)	fMRI	16	German	passive listening	sentences	continuous	event- related	sentences>scanner noise	Table 1	9	6
45	(Ischebeck, Friederici, & Alter, 2008)	fMRI	16	German	passive listening	sentences	continuous	block	sentences>scanner noise	Table 1	8	12
46	(Kovelman, Baker, & Petitto, 2008)	fMRI	10	American- English	anomaly judgment	sentences	continuous	block	sentences>scanner noise	Table 3	11	6
47	(LoCasto, Krebs-Noble, Gullapalli, & Burton, 2004)	fMRI	20	American- English	similarity judgment	words	continuous	block	words>tones	Table 3	13	4
48	(Love, Haist, Nicol, &	fMRI	10	American-	various tasks	sentences	continuous	block	sentences>scanner noise	Table II	16	8

	Swinney, 2006)			English								
49	(Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003)	fMRI	13	British-English	passive listening	sentences	continuous	block	sentences>scanner noise	Table 2	5	*
50	(Rimol, Specht, & Hugdahl, 2006)	fMRI	18	Norwegian	repetition decision	words	continuous	block	words>tones	Table 3	6	8
51	(Schirmer, Zysset, Kotz, & Yves von Cramon, 2004)	fMRI	27	German	emotion judgment	words	continuous	block	words>scanner noise	Table 1	5	4
52	(Schon et al., 2010)	fMRI	13	French	similarity judgment	words	continuous	block	words>pink noise	Table 1	5	9
53	(Skipper, Nusbaum, & Small, 2005)	fMRI	9	American-English	passive listening + after task	stories	continuous	block	stories>scanner noise	Table 1	10	4
54	(Tracy et al., 2003)	fMRI	15	American-English	passive listening	narratives	continuous	block	narratives>scanner noise	Table 2	2	8
55	(Wildgruber et al., 2004)	fMRI	10	German	prosody judgment	sentences	continuous	event-related	sentences>scanner noise	Table 1 (“Linguistic versus rest”)	12	10
56	(Wilson, Molnar-Szakacs, & Jacoboni, 2008)	fMRI	12	American-English	passive listening + after test	narratives	continuous	block	narratives>scanner noise	Table 1 (“Signal increases in standard analysis”)	7	8
57	(Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006)	fMRI	10	Dutch	decide which sentences was just presented (4aFC)	sentences	sparse	block	sentences>white noise	Table 2	9	6

Table II. Overview of distribution of the selected studies over all combinations of the four design options: words vs. Sentences vs. Narratives, presence of a task, scanning paradigm (sparse temporal sampling or continuous scanning), and the use of an acoustically matched or non-matched baseline. The three PET studies (Crinion et al. 2003; Scott et al., 2000, Scott et al., 2006) and Peelle et al. 2010 were excluded as they could not be included in any of the combinations (as PET studies do not use continuous or sparse scanning and Peelle et al. pooled their data across both paradigms).

Combination	# of studies	Studies
Words, task, continuous, matched	8	Christensen et al. 2008; Gagnepain et al., 2008; Kotz, et al. 2002; Leff et al., 2009; Raettig & Kotz, 2008; Rissman, et al. 2003; Stevenson & James, 2009; Xiao et al., 2005
Words, task, sparse, matched	3	Bozic et al. 2010; Obleser et al. 2008; Orfanidou et al. 2006
Words, task, sparse, non-matched	0	-
Words, task, continuous, non-matched	5	Binder et al., 2008; LoCasto, et al. 2004; Rimol et al, 2006; Schirmer et al, 2004; Schön et al., 2010
Words, no task, continuous, matched	1	Klein et al. 2010
Words, no task, sparse, matched	0	-
Words, no task, sparse, non-matched	0	-
Words, no task, continuous, non-matched	1	Binder et al. 2000
Sentences, task, continuous, matched	5	Devauchelle et al, 2008; Giraud et al., 2004; Meyer et al., 2002; Okada et al., 2010; Peelle et al., 2004
Sentences, task, sparse, matched	7	Allen et al. 2005; Davis et al. 2003; Davis et al. 2011; Rodd et al. 2010; Rodd et al. 2005; Tyler et al. 2010 et al. (both groups)

Sentences, task, sparse, non-matched	2	Dick et al. 2007; Zekveld et al. 2006
Sentences, task, continuous, non-matched	5	Adank & Devlin 2010; ; Constable et al. 2004; Kovelman et al. 2008; Love et al. 2006; Wildgruber et al. 2004
Sentences, no task, continuous, matched	1	Röder et al. 2002
Sentences, no task, sparse, matched	7	McGettigan et al. 2012; Narain et al. 2003; Obleser et al. 2011; Obleser et al. 2007; Obleser et al. 210; Friederici et al. 2010; Takeichi et al. 2010
Sentences, no task, sparse, non-matched	0	-
Sentences, no task, continuous, non-matched	4	Gonzalez-Castillo et al. 2011; Holle et al. 2010; Ischebeck et al. 2008; Mitchell et al. 2003
narratives, task, continuous, matched	0	-
narratives, task, sparse, matched	0	-
Sentences, task, sparse, non-matched	0	-
narratives, task, continuous, non-matched	0	-
narratives, no task, continuous, matched	1	Jardri et al. 2005
narratives, no task, sparse, matched	1	-
narratives, no task, sparse, non-matched	0	-
narratives, no task, continuous, non-matched	3	Skipper et al. 2005; Tracy et al. 2003; Wilson et al. 2008;

Table III. Activated clusters for the included experiments, coordinates in MNI space, including number of contributing foci ‘()’. AG: Angular Gyrus, MTG: Middle Temporal Gyrus; STG; PMv: ventral Premotor Cortex; POp: Pars Opercularis; POrb: Pars Orbitalis; Superior Temporal Gyrus; (Pre-)SMA: Supplementary Motor Cortex, STS: Superior Temporal Sulcus.

Cluster	mm ³	ALE	Location	x	y	z	BA	Contributing studies		
1	22152	0.077	L Anterior STS	-60	-12	-6	21	Adank & Devlin 2010 (2)	Ischebek et al. 2008 (1)	Rodd et al. 2010 (7)
		0.046	L MTG	-58	-32	2	-	Allen et al. 2005 (1)	Jardri et al. 2007 (3)	Röder et al. 2002 (2)
		0.045	L anterior STG	-52	8	-18	38	Binder et al. 2000 (5)	Kovelman et al. 2010 (3)	Schön et al. 2010 (3)
		0.026	L POp	-48	16	14	44	Bozic et al. 2010 (2)	Leff et al. 2009 (3)	Scott et al. 2000 (2)
		0.023	L Posterior MTG	-52	-48	10	21	Constable et al. 2004 (2)	LoCasto et al. 2004 (2)	Scott et al. 2006 (3)
		0.023	L AG	-54	-56	14	39	Crinion et al. 2003 (5)	Love et al. 2006 (1)	Skipper et al. 2005 (2)
		0.022	L PMv	-50	14	26	9	Davis et al. 2003 (2)	McGettigan et al. 2012 (7)	Stevenson et al. 2009 (1)
		0.018	L POrb	-44	24	-18	47	Davis et al. 2011 (7)	Meyer et al. 2002 (1)	Takeichi et al. 2010 (1)
2	13432							Devauchelle et al. 2009 (3)	Narain et al. 2003 (2)	Tracy et al. 2003 (1)
								Dick et al. 2007 (2)	Obleser et al. 2007 (1)	Tyler et al. young 2010 (3)
								Friederici et al. 2010 (1)	Obleser et al. 2008 (3)	Tyler et al. old 2010 (3)
								Gagnepain et al. 2008 (2)	Obleser et al. 2010 (1)	Wildgruber et al. 2004 (2)
								Giraud et al. 2004 (1)	Obleser et al. 2011 (1)	Wilson et al. 2008 (2)
								Gonzalez-Castillo et al. 2011 (2)	Orfanidou et al. 2006 (1)	Xiao et al. 2005 (1)
								Holle et al. 2010 (1)	Peelle et al. 2010 (1)	Zekfeld et al. 2006 (4)
								Rimol et al. 2006 (3)	Rodd et al. 2005 (8)	
		0.057	R Anterior STS	62	-8	-10	21	Adank & Devlin 2010 (1)	LoCasto et al. 2004 (1)	Rodd et al. 2005 (3)
		0.034	R Insula	54	-20	4	13	Binder et al. 2000 (3)	Love et al. 2006 (2)	Rodd et al. 2010 (4)
		0.030	R anterior STG	52	10	-22	38	Bozic et al. 2010 (3)	McGettigan et al. 2012 (4)	Schirmer et al. 2003 (1)
		0.024	L MTG	62	-32	-2	-	Constable et al. 2004 (1)	Meyer et al. 2002 (1)	Schön et al. 2010 (1)
								Crinion et al. 2003 (3)	Mitchell et al. 2006 (1)	Scott et al. 2006 (1)

								Davis et al. 2003 (1) Davis et al. 2011 (7) Dick et al. 2007 (3) Friederici et al. 2010 (1) Gagnepain et al. 2008 (4) Giraud et al. 2004 (1) Holle et al. 2010 (1)	Obleser et al. 2007 (1) Obleser et al. 2008 (1) Obleser et al. 2010 (1) Obleser et al. 2011 (1) Okada et al. 2010 (1) Peelle et al. 2010 (2) Rimol et al. 2006 (2)	Stevenson et al. 2009 (1) Takeichi et al. 2010 (1) Tracy et al. 2003 (1) Tyler et al. young 2010 (1) Wilson et al. 2008 (2) Zekfeld et al. 2006 (1)
3	1232	0.021	SMA	0	4	58	6	Binder et al. 2008 (1) Christensen et al. 2008 (1)	Constable et al. 2004 (2) LoCasto et al. 2004 (1)	Love et al. 2006 (1) Röder et al. 2002 (1)
		0.021	Pre-SMA	-2	12	50	6			
4	808	0.022	L Fusiform Gyrus	-36	-38	-20	20	Bozic et al. 2010 (1) Davis et al. 2011 (1)	LoCasto et al. 2004 (1) Orfanidou et al. 2006 (1)	Okada et al. 2010 (1) Rodd et al. 2005 (1)
5	640	0.023	L PMv	-52	30	10	46	Dick et al. 2007 (1) Meyer et al. 2002 (1)	Rodd et al. 2005 (1) Skipper et al. 2005 (1)	
6	528	0.019	L Precentral Gyrus	-44	-8	56	4	LoCasto et al. 2004 (1) McGettigan et al. 2012 (1)	Wildgruber et al. 2004 (1) Wilson et al. 2008 (1)	
		0.018	L Precentral Gyrus	-48	-2	50	6			

Table IV. Activated clusters for the subtraction analyses on studies using a sparse vs. continuous scanning paradigm, a matched vs. a non-matched acoustic baseline, and words vs. sentences. Coordinates are displayed in MNI space. MFG: Middle Frontal Gyrus; MTG: Middle Temporal Gyrus; POp: Pars Opercularis; PTr: Pars Triangularis; SMA: Supplementary Motor Area; STG: Superior Temporal Gyrus, STS: Superior Temporal Sulcus.

Sparse > Continuous							
Cluster	mm ³	Z-score	Location	x	y	z	BA
1	2128	3.72	L Anterior MTG	-53	0	-21	21
		3.54	L Anterior MTG	-53	-3	-22	21
		3.35	L Anterior MTG	-56	-10	-22	21
		2.99	L Anterior STG	-47	15	-25	38
2	1832	3.72	R Anterior STG	60	-7	-5	22
3	1000	3.72	L Posterior STG	-53	-37	0	-
		3.54	L Posterior STG	-48	-38	-1	22
4	432	2.77	L AG	-55	-59	13	39
		2.75	L AG	-58	-56	10	39
		2.69	L AG	-58	-56	16	39
Continuous > Sparse							
1	2128	3.04	ACG	-3	11	50	24
		2.91	Pre-SMA	0	16	52	6
		2.77	SMA	2	3	58	6
Matched > Non-matched baseline							
1	2512	3.35	L Anterior STS	-54	11	-24	38
		3.24	L Anterior STS	-56	5	-24	21
		3.09	L Anterior MTG	-58	-4	-20	21
		2.99	L Anterior STS	-58	-4	-16	21
		2.88	L Anterior STG	-56	14	-18	38

		2.73	L Anterior STG	-47	20	-22	38
Non-Matched > Matched baseline							
1	560	2.91	L Precentral Gyrus	-47	15	26	9
		2.82	L MFG	-52	18	28	9
2	472	3.04	L Precentral Gyrus	-43	-12	57	4
Sentences > Words							
1	592	3.35	L POp	-53	16	12	44
2	464	3.54	L PTr	-52	28	6	45
		3.35	L PTr/Insula	-52	28	10	13
Words > Sentences							
<i>No significant clusters</i>							

Table SI. Activated clusters for the included experiments (excluding experiments using pseudowords as a non-speech baseline), coordinates in MNI space, including number of contributing foci ‘()’. AG: Angular Gyrus, MTG: Middle Temporal Gyrus; STG; PMv: ventral Premotor Cortex; POp: Pars Opercularis; POrb: Pars Orbitalis; Superior Temporal Gyrus; SMA: Supplementary Motor Cortex, STS: Superior Temporal Sulcus.

Cluster	mm ³	ALE	Location	x	y	z	BA	Contributing studies		
1	21216	0.077	L Anterior STS	-60	-12	-6	21	Adank & Devlin 2010 (2)	Jardri et al. 2007 (3)	Rodd et al. 2005 (7)
		0.045	L MTG	-58	-32	0	-	Allen et al. 2005 (1)	Kovelman et al. 2010 (3)	Rodd et al. 2010 (7)
		0.041	L anterior STG	-52	8	-20	38	Binder et al. 2000 (5)	Leff et al. 2009 (3)	Schön et al. 2010 (3)
		0.022	L AG	-54	-56	14	39	Bozic et al. 2010 (2)	LoCasto et al. 2004 (2)	Scott et al. 2000 (2)
		0.021	L PMv	-50	14	26	9	Constable et al. 2004 (2)	Love et al. 2006 (1)	Scott et al. 2006 (3)
		0.021	L POp	-48	16	18	44	Crinion et al. 2003 (5)	McGettigan et al. 2012 (7)	Skipper et al. 2005 (1)
		0.018	L POrb	-44	24	-18	47	Davis et al. 2003 (1)	Meyer et al. 2002 (1)	Stevenson et al. 2009 (1)
								Davis et al. 2011 (7)	Narain et al. 2003 (2)	Takeichi et al. 2010 (1)
								Devauchelle et al. 2009 (3)	Obleser et al. 2007 (1)	Tracy et al. 2003 (1)
								Dick et al. 2007 (2)	Obleser et al. 2008 (3)	Tyler et al. young 2010 (3)
								Friederici et al. 2010 (1)	Obleser et al. 2010 (1)	Tyler et al. old 2010 (3)
								Giraud et al. 2004 (1)	Obleser et al. 2011 (1)	Wildgruber et al. 2004 (2)
								Gonzalez-Castillo et al. 2011 (2)	Peelle et al. 2010 (1)	Wilson et al. 2008 (2)
								Holle et al. 2010 (1)	Rimol et al. 2006 (3)	Zekfeld et al. 2006 (4)
								Ischebek et al. 2008 (1)		
2	12568	0.057	R Anterior STS	62	-8	-10	21	Adank & Devlin 2010 (1)	Love et al. 2006 (1)	Rodd et al. 2010 (4)
		0.033	R Insula	54	-20	4	13	Binder et al. 2000 (3)	McGettigan et al. 2012 (4)	Schirmer et al. 2003 (1)
		0.030	R anterior STG	52	10	-22	38	Bozic et al. 2010 (3)	Meyer et al. 2002 (1)	Schön et al. 2010 (1)
		0.024	L MTG	62	-32	-2	-	Constable et al. 2004 (1)	Mitchell et al. 2006 (1)	Scott et al. 2006 (1)
								Crinion et al. 2003 (3)	Obleser et al. 2007 (1)	Stevenson et al. 2009 (1)

								Davis et al. 2003 (1) Davis et al. 2011 (7) Dick et al. 2007 (3) Friederici et al. 2010 (1) Giraud et al. 2004 (1) Holle et al. 2010 (1) LoCasto et al. 2004 (1)	Obleser et al. 2008 (1) Obleser et al. 2010 (1) Obleser et al. 2011 (1) Okada et al. 2010 (1) Peelle et al. 2010 (2) Rimol et al. 2006 (2) Rodd et al. 2005 (3)	Takeichi et al. 2010 (1) Tracy et al. 2003 (1) Tyler et al. yng 2010 (1) Wilson et al. 2008 (2) Zekfeld et al. 2006 (1)
3	928	0.021	SMA	0	4	58	0	Adank & Devlin 2010 (1) Christensen et al. 2008 (1)	Constable et al. 2004 (2) Love et al. 2006 (1)	
4	752	0.023	L PMv	-52	30	10	46	Dick et al. 2007 (1) Meyer et al. 2002 (1)	Rodd et al. 2005 (1) Skipper et al. 2005 (1)	
5	664	0.019	L Precentral Gyrus	-44	-8	56	4	LoCasto et al. 2004 (1)	Peelle et al. 2010 (1)	Wilson et al. 2008 (1)
		0.018	L Precentral Gyrus	-48	-2	50	6	Love et al. 2006 (1) McGettigan et al. 2012 (1)	Takeichi et al. 2010 (1) Wildgruber et al. 2004 (1)	
6	608	0.021	L Cerebellum	-38	-42	-24	-	Bozic et al. 2010 (1) Davis et al. 2011 (1)	LoCasto et al. 2004 (1) Okada et al. 2010 (1)	Rodd et al. 2005 (1)

x = -60

y = -30

x = 62

L

x = -56

y = 0

x = 55

L

x = -50

y = 10

x = 45

L

0.02

ALE score

0.08