

Effects of Prior Information on Decoding Degraded Speech: An fMRI Study*

Mareike Clos,^{1*} Robert Langner,^{1,2} Martin Meyer,³ Mathias S. Oechlin,⁴
Karl Zilles,^{1,5,6} and Simon B. Eickhoff^{1,2}

¹*Institute of Neuroscience and Medicine (INM-1, INM-2), Research Center Jülich, Germany*

²*Institute of Clinical Neuroscience and Medical Psychology, Heinrich Heine University, Düsseldorf, Germany*

³*Neuroplasticity and Learning in the Normal Aging Brain (HAB LAB), University of Zurich, Switzerland*

⁴*Faculty of Psychology and Educational Sciences, Geneva Neuroscience Center, University of Geneva, Switzerland*

⁵*C. and O. Vogt Brain Research Institute, University of Düsseldorf, Germany*

⁶*JARA-Translational Brain Medicine, Germany*

Abstract: Expectations and prior knowledge are thought to support the perceptual analysis of incoming sensory stimuli, as proposed by the predictive-coding framework. The current fMRI study investigated the effect of prior information on brain activity during the decoding of degraded speech stimuli. When prior information enabled the comprehension of the degraded sentences, the left middle temporal gyrus and the left angular gyrus were activated, highlighting a role of these areas in meaning extraction. In contrast, the activation of the left inferior frontal gyrus (area 44/45) appeared to reflect the search for meaningful information in degraded speech material that could not be decoded because of mismatches with the prior information. Our results show that degraded sentences evoke instantaneously different percepts and activation patterns depending on the type of prior information, in line with prediction-based accounts of perception. *Hum Brain Mapp* 35:61–74, 2014. © 2012 Wiley Periodicals, Inc.

Key words: auditory expectations; Broca's region; middle temporal gyrus; predictive coding; speech perception

INTRODUCTION

The comprehension of spoken utterances is a highly challenging task due to the transient nature of auditory

speech stimuli and its vulnerability to ambiguity. The success of our sensory system to convey most stimuli with reasonable precision despite the regular disturbance of noise has been attributed to its constant anticipation of

Re-use of this article is permitted in accordance with the Terms and Conditions set out at http://wileyonlinelibrary.com/onlineopen#OnlineOpen_Terms.

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Human Brain Project; Contract grant number: R01-MH074457-01A1; Contract grant sponsor: DFG; Contract grant number: IRTG 1328; Contract grant sponsor: The Initiative and Networking Fund of the Helmholtz Association within the Helmholtz Alliance on Systems Biology; Contract grant sponsor: the Förderung des akademischen Nachwuchses (FAN) of

the Zürcher Universitätsvereins (ZUNIV); Contract grant sponsor: Swiss National Foundation; Contract grant number: 320030-120661.

*Correspondence to: Mareike Clos, Forschungszentrum Jülich, INM-2, 52425 Jülich, Germany. E-mail: m.clos@fz-juelich.de

Received for publication 18 March 2012; Revised 2 June 2012; Accepted 5 June 2012

DOI: 10.1002/hbm.22151

Published online 30 August 2012 in Wiley Online Library (wileyonlinelibrary.com).

upcoming events [Bar, 2007; Enns and Lleras, 2008; Friston, 2005]. That is, perception is not simply a passive reflection of sensory input but arises from an active integration of sensory data and prior expectations. According to the framework of predictive coding [Friston, 2005; Rao and Ballard, 1999], such expectations can resolve perceptual ambiguities because prior knowledge and context information incorporated in the predictions might help to decode noisy stimuli. Predictive coding is conceptually related to, e.g., semantic or contextual priming but refers to the use of any kind of prior information. In fact, priming may thus be regarded as a special instance of predictive coding in which a single piece of prior information (the prime) influences the processing of the subsequent stimulus.

Evidence for predictive coding has been observed in visual processing [e.g., Hosoya et al., 2005; Sharma et al., 2003; Summerfield et al., 2006], in tactilo-motor interactions [Blakemore et al., 1998], in motor preparation [Jakobs et al., 2009] and in audiovisual perception [Arnal et al., 2011; den Ouden et al., 2009, 2010]. However, rather little is known about the role of predictive coding in auditory speech perception despite the fact that comparable expectation-generating mechanisms involving interactions between bottom-up and top-down processing have often been deemed crucial for speech perception [e.g., Davis and Johnsrude, 2007; Grossberg, 2003]. Furthermore, auditory processing is thought to be hierarchically organized such that higher processing levels respond to increasingly more complex and abstract sound properties. In accordance with this hierarchical organization, research has identified at least three distinct levels of auditory processing in non-human primates. Specifically, with increasing complexity, sound information proceeds from the core region of the auditory cortex to the belt and the more lateral parabelt area [Kaas and Hackett, 2000; Rauschecker et al., 1995]. Neuroimaging studies have found a comparable hierarchical processing pattern in humans in response to complex sounds [Hall et al., 2002] and to spoken language [Scott et al., 2000]. Whereas the core and belt areas of the auditory cortex in the superior temporal gyrus are responsive to the amplitude and frequency modulations of speech, left lateralized cortical regions including the posterior inferior parietal lobe, middle temporal gyrus, fusiform and parahippocampal gyrus, dorsomedial prefrontal cortex, inferior frontal gyrus, ventromedial prefrontal cortex, and posterior cingulate gyrus seem to be involved in speech-specific semantic processing [Binder et al., 2009].

Previously, indications for effects of prior knowledge on speech perception have been demonstrated in studies showing that speech can be decoded even when extremely distorted [Remez et al., 1994; Saberi and Perrott, 1999; Shannon et al., 1995] and that perception of degraded speech can improve through training [Davis et al., 2005; Giraud et al., 2004; Hannemann et al., 2007]. Moreover, comprehension of degraded speech stimuli after training was associated with increased blood-oxygen-level-dependent (BOLD) activity in the right superior temporal sulcus

and bilateral middle and inferior temporal gyri [Giraud et al., 2004] and with increased gamma band activity in left temporal regions [Hannemann et al., 2007] when compared to exposure to the degraded stimuli prior to the training. As noted above, however, experiments in the visual domain have demonstrated analogous perceptual and/or neural effects of prior information also without requiring a training phase. That is, degraded images could be recognized instantaneously once the original (nondegraded) image had been shown [e.g., Ludmer et al., 2011; Porter, 1954]. However, the neural mechanisms of corresponding phenomena in the auditory domain (i.e., speech perception) are yet unknown.

To shed light on this question, we used functional magnetic resonance imaging (fMRI) to measure brain activity during the perception of spectrally degraded sentences following exposure to an equally degraded or, which is the critical condition, following exposure to a nondegraded sentence. Importantly, the employed spectral degradation (see Methods for details) produced sentences that were incomprehensible when heard in isolation but which became comprehensible (i.e., their meaning could be extracted) when having been preceded by their original (nondegraded) version. In light of the predictive-coding framework, the comprehension of degraded speech can be explained by the formation of a template based on the processing of the preceding nondegraded sentence. This template consists of predictive codes that, if matching with subsequent input, carry enough information for the successful decoding of degraded speech. Subsequently, we refer to this prediction-based understanding of speech as “meaning extraction,” although we do not claim that this process can be mechanistically equated with a direct decoding of the degraded sentence. Alternatively, understanding might stem from an indirect meaning reactivation that is triggered after the degraded sentence is “recognized” based on structural commonalities (e.g., prosody) with the template. Most probably, however, both processes—direct meaning decoding based on lexical-semantic predictions and meaning retrieval based on a structural match with previous language input—run in parallel. In any case, the prediction-dependent understanding of degraded language offers an excellent opportunity to investigate the neural mechanisms of integrating sensory data and prior knowledge in speech processing.

METHODS

Participants

Twenty-nine healthy participants took part in this study (14 females, mean age = 34.5 years, SD = 12.2 years). All participants were right-handed, native speakers of German, had no history of neurological or psychiatric diseases, and gave written informed consent prior to participation. The study was approved by the local ethics committee of the RWTH Aachen University.

TABLE I. Overview of conditions

Reference	Target	Condition
<i>Structural prior</i>		
Degraded sentence A	Degraded sentence A	Structural match
Degraded sentence A	Degraded sentence B	Structural mismatch
<i>Propositional prior</i>		
Nondegraded sentence A	Degraded sentence A	Propositional match
Nondegraded sentence A	Degraded sentence B	Propositional mismatch

Paradigm and Stimuli

The participants performed a delayed-matching-to-sample task in which a target sentence had to be compared with a preceding reference sentence. The stimuli comprised 25 sentences, each in a nondegraded and a degraded version. All stimuli had been developed for a previous fMRI experiment and are described in detail in Meyer et al. [2004]. In brief, the sentences were recordings of short declarative infinitival statements of similar length (mean duration = 3.8 s, SD = 0.3 s), spoken by the same female speaker. A transcription of these sentences can be found in the Supporting Information. Degraded versions of these sentences (see the Supporting Information for a sound example) were created by low-pass filtering and an additional removal of aperiodic signals. This procedure included a reduction of spectral information to frequencies containing the F_0 as well as the 2nd and 3rd harmonic (see Supporting Information Fig. S1 for spectrograms of a nondegraded sentence and its degraded version). Thus, the resulting degraded stimuli merely retained the prosodic parameters of the original version (i.e., intonation, duration, and suprasegmental acoustic modulations) but lack any segmental and lexical information. Unlike purely low-pass filtered sentences, the degraded sentences employed in the present study sound like a humming voice heard from behind a door and are virtually impossible to understand without prior presentation of the nondegraded version as a reference [Meyer et al., 2002].

The experiment consisted of five blocks, each containing two subblocks of 10 events. For each event, the stimuli were presented in pairs consisting of a reference sentence followed by a target sentence for comparison. While the target sentence always was a degraded sentence, the type of the preceding reference sentence alternated between sub-blocks: it was degraded in the first sub-block and nondegraded in the second sub-block. The reason for keeping the order of sub-blocks constant (rather than randomizing them) was twofold. First, we wanted to minimize task-switching demands. Second, we wanted to ensure that the target sentences preceded by degraded reference sentences (structural match and mismatch conditions, cf., Table I) were processed as spectrally degraded, incomprehensible sentences. Therefore, we decided to present them always in the first sub-block before the template based on the intact sentence could be formed (i.e., before encountering

the nondegraded version in the second sub-block). Each sub-block consisted of five matching and five nonmatching pairs presented in randomized order. Within each block, the pairs in either sub-block were based on combinations of the same five sentences to ensure equivalent stimulus material and hence sensory input for both sub-blocks. The order of sentence presentation within each sub-block and the order of blocks were pseudo-randomized across participants. In sum, the type of reference sentence (degraded vs. nondegraded) varied between sub-blocks to minimize trial-to-trial task-switching effects, while the type of reference-target match (match vs. mismatch) varied between trials (i.e., within sub-blocks) to allow for an event-related analysis of hemodynamic activity (see below). The current experiment thus uses an event- (epoch-) related design for modeling and analysis, which was embedded in an overarching block-structure of event-presentation to reduce confounding effects of task-switching.

Our pairing scheme yielded four conditions (Table I): (1) a structural match condition when two identical degraded sentences were presented; (2) a structural mismatch condition when two different degraded sentences were presented; (3) a propositional match condition when a nondegraded reference sentence was identical to the degraded target (i.e., the target sentence was the degraded version of the reference sentence); and (4) a propositional mismatch condition when a nondegraded reference was different from the degraded target (i.e., the target sentence was not the degraded version of the reference sentence). The first two conditions thus require sentence comparisons that are entirely based on “structural” information such as prosody, suprasegmental acoustic modulations, intonation, pitch, etc. The latter two conditions in contrast enable comparisons that are additionally based on lexical-semantic (“propositional”) information provided by the nondegraded reference sentence. Importantly, the propositional match condition evoked an understanding of the degraded target (as established by pretesting). Therefore, this condition allowed for meaning extraction from the degraded (and normally unintelligible) target.

Each of the 25 degraded sentences was presented exactly once as a target in every condition. Therefore, the stimulus material constituting the (crucial) second part of each event was identical across conditions. Thus, the only difference between conditions that could explain differential fMRI results was (1) whether only structural or also

propositional information was provided by the reference sentence, and (2) whether this information matched the target sentence.

Procedure

After receiving task instructions, participants performed a practice run with sentence pairs different from those in the main experiment. The practice run was introduced to familiarize participants with the auditory stimuli and with the sequence of events. After entering the MR scanner, a sequence of test scans was run while examples of practice sentence stimuli were presented. This was done to allow an individual adjustment of the headphone volume for each participant and ensure that the sentence stimuli were well audible with the scanner noise in the background. Subsequently, the experiment started. Following the presentation of each sentence pair, a display was shown for 2 s asking participants to indicate by left or right button press whether or not the sentence pair contained two identical sentences. Participants were instructed to respond as fast as possible. Left/right response assignment was counterbalanced across participants such that half the participants responded with the left hand and the other half responded with the right hand to specify identical sentences. After a jittered intertrial interval of 4–9 s (uniformly distributed), the next sentence pair was presented. The sentences within each pair were separated by an interstimulus interval of 1 s. The sub-blocks lasted about 3 min and were separated by a 20-s resting period from each other. A warning tone in combination with a warning on the display was presented 1–3 s prior to the end of the resting period to prepare participants for the upcoming sub-block. The total time spent in the scanner was ~35 min.

fMRI Data Acquisition and Preprocessing

Imaging was performed on a Siemens Trio 3-T whole-body scanner (Erlangen, Germany) using gradient-echo echo-planar imaging (EPI). T2*-weighted BOLD contrast volumes covering the whole brain were acquired (TR = 2.2 s, in-plane resolution = $3.1 \times 3.1 \text{ mm}^2$, 36 axial slices of 3.1 mm thickness, distance factor = 15%). To allow for magnetic-field saturation, image acquisition was preceded by four dummy images which were discarded prior to data analysis. Images were analyzed using SPM5 (www.fil.ion.ucl.ac.uk/spm). The EPI images were corrected for head movement by affine registration using a two-pass procedure. This included an initial realignment of all images to the first image and a subsequent realignment to the mean of the realigned images. After realignment, the mean EPI image of each participant was spatially normalized to the MNI (Montreal Neurological Institute) single-subject template using the unified segmentation approach (Ashburner and Friston, 2005). The resulting parameters

that define the deformation field necessary to move the participant's data into the space of the MNI tissue probability maps were then combined with the deformation field transforming between the latter and the MNI single-subject template. The ensuing deformation was subsequently applied to the individual EPI volumes that were thereby transformed into the MNI single-subject space and resampled at $2 \times 2 \times 2 \text{ mm}^3$ voxel size. Finally, these normalized images were spatially smoothed with a Gaussian kernel of 8-mm full width at half-maximum.

Data Analysis

Behavioral data were analyzed using SPSS 18.0.0 (SPSS, Chicago, IL). Reaction time and accuracy were subjected to 2×2 repeated-measures analyses of variance (ANOVAs) to test the effects of the factors Match (matching vs. non-matching sentence pairs) and Type of Prior (propositional vs. structural). Furthermore, reaction time of match and nonmatch trials was separately calculated for correct and incorrect trials (i.e., hits, misses, correct rejections, false alarms) and tested for the effects of signal-detection category and Type of Prior by a 4×2 ANOVA. Note that hits and misses were computed from correct and incorrect responses on match trials, respectively, and correct rejections and false alarms were computed from correct and incorrect responses on mismatch trials, respectively. Post-hoc analyses were Bonferroni-corrected for multiple comparisons. Finally, paired t tests were used to test for group-level differences in sensitivity (d') and decision criterion (c) between the different types of prior. The d' parameter was calculated based on the convention suggested for same-different designs [Macmillan and Creelman, 2005] using the formula

$$d' = 2z(0.5\{1 + [2p(c)\text{max} - 1]/2\})$$

with

$$p(c)\text{max} = \Phi(0.5\{z(H) - z(F)\}),$$

where Φ is the standard normal cumulative density function, H is the hit rate (i.e., proportion of match responses when pairs actually were matches), and F is the false-alarm rate (i.e., proportion of match responses when pairs actually were mismatches). The decision criterion c was calculated by the formula

$$c = -0.5[z(H) + z(F)].$$

Admittedly, typical same-different experiments within the signal-detection framework differ from the current experiment in terms of the number and complexity of stimuli. This analysis, however, was only performed to provide evidence for equivalent levels of difficulty across

both types of prior. For this purpose, applying the same-different convention should provide an acceptable approximation.

Imaging data were analyzed using the general linear model as implemented in SPM5. For each of the six events of interest (presentation of nondegraded or degraded reference sentences, presentation of target sentences from one of the four conditions: structural match, structural mismatch, propositional match, and propositional mismatch), the hemodynamic response was separately modeled by a boxcar reference vector (duration: 4 s) convolved with a canonical hemodynamic response function (HRF) and its first-order temporal derivative. The four target sentence regressors thus defined the four conditions which were identical with regard to sensory input but differed with regard to the type of input provided by the preceding reference sentence. Importantly, we limited our analysis to those target sentences that evoked correct match or mismatch responses (i.e., hits or correct rejections, respectively). Accordingly, a nuisance regressor for target sentences in trials with incorrect responses was included in the first-level model. The reason for restricting the analysis to correct trials was to ensure that participants paid attention to the task at hand during all trials included. However, one disadvantage of this approach is the potential reduction of statistical power due to the exclusion of trials. Furthermore, this exclusion may perturb the a-priori identical distribution of target stimuli across the four conditions because some of the excluded target sentences may still be present in one of the other conditions. We therefore computed a supplementary analysis based on all trials in which we reanalyzed the imaging data from all trials without the nuisance regressor.

Additional nuisance regressors were included for experimental events of no interest: left and right button presses and head movements as reflected by six motion parameters for translation and rotation. Finally, reaction time was included as parametric modulator for the structural match, structural mismatch, propositional match, and propositional mismatch regressors to assess intraindividual variation of brain activity related to performance level. Low-frequency signal drifts were removed by employing a high-pass filter with a cut-off period of 128 s. After correction of the time series for dependent observations according to an autoregressive first-order correlation structure, parameter estimates of the HRF regressors were calculated for each voxel using weighted least squares to provide maximum-likelihood estimators based on the temporal autocorrelation of the data [Kiebel et al., 2003]. The individual first-level contrasts (each condition relative to the implicit baseline) were then fed into a second-level random-effects ANOVA.

In this group analysis, mean parameter estimates over all participants were computed for all six regressors of interest (cf., above) as well as for the four parametric modulators (reflecting reaction times) and the two motor-response nuisance regressors (left/right button press).

Based on these estimates, separate t-contrasts within the ANOVA were calculated for testing differential effects. Furthermore, differential effects were combined into conjunctions based on the minimum t-statistic [Nichols et al., 2005]. Conjunction analysis was chosen because of its higher specificity and more conservative character as compared with a factorial analysis. In particular, by using a conjunction analysis, we constrained inference to those regions that were significantly present in all of the included conditions. All resulting activation maps were thresholded at $P < 0.05$ (family-wise error (FWE)-corrected for multiple comparisons at cluster level; cluster-forming threshold at voxel level: $P < 0.001$) and anatomically localized using probabilistic maps of cytoarchitecturally defined areas [Amunts et al., 2007; Zilles and Amunts, 2010] using version 1.6b of the SPM Anatomy toolbox [Eickhoff et al., 2005; www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox].

To identify regions implicated in the processing of all six sentence type events, the conjunction “nondegraded reference sentence \cap degraded reference sentence \cap structural match \cap structural mismatch \cap propositional match \cap propositional mismatch” was used. This analysis should thus reflect regions commonly activated by the sound stimuli or recruited by the general task demands (e.g., working memory, decision making). The contrast “degraded reference sentence $>$ nondegraded reference sentence” was employed to isolate regions that are more activated by the unintelligible sounds as compared to meaningful verbal information. The inverse contrast “nondegraded reference sentence $>$ degraded reference sentence” was analyzed to discern regions more tuned to intelligible speech than to (unintelligible) dynamic intonation contour. The latter contrast should thus identify regions that are selectively involved in processing the lexical-semantic aspects of speech.

Three conjunctions were employed to unravel the effects of the propositional prior compared to a purely structural prior, i.e., the effects of a nondegraded reference sentence compared to a degraded reference sentence on the subsequent processing of the degraded target sentence. To ensure that all activations associated with the propositional prior effect were specific for intelligible speech, the contrast “nondegraded reference sentence $>$ degraded reference sentence” was always included in these conjunctions. That is, we compared differential effects of the previously heard sentence on the processing of the identical (precisely the same target stimuli were presented in all four conditions) degraded sentences, but restricted this analysis to those regions that were actually involved in processing nondegraded (i.e., intelligible) speech, as opposed to degraded (i.e., normally unintelligible) speech. Hence, all these analyses should exclusively reveal effects within the brain network subserving lexical-semantic speech processing. First, the conjunction “[propositional match + propositional mismatch] $>$ (structural match + structural mismatch)] \cap (nondegraded reference sentence

> degraded reference sentence)'' aimed at identifying the general effect of exposure to a propositional prior as compared to a structural prior. This conjunction thus should specifically reveal those effects on the processing of the degraded target sentence that stem from any lexical-semantic influence provided by the reference sentence while controlling for (mis)matching of prosody (as this process should affect all target sentences alike, independently of whether the prior was propositional or structural). Second, the conjunction ''(propositional match > structural match) \cap (propositional match > structural mismatch) \cap (nondegraded reference sentence > degraded reference sentence)'' was used to reveal the effects of a propositional match on brain activity within the lexical-semantic network (as defined by the last component of the conjunction). Finally, the conjunction ''(propositional mismatch > structural mismatch) \cap (propositional mismatch > structural match) \cap (nondegraded reference sentence > degraded reference sentence)'' was employed to reveal the effects of a propositional mismatch on brain activity within the lexical-semantic network. Furthermore, we performed an additional analysis of these effects that was not restricted to those regions more responsive to intelligible compared to degraded speech.

Furthermore, we directly tested for differences between propositional matches and mismatches via (i) the contrast ''propositional match > propositional mismatch'' masked inclusively with the above analysis aiming at identifying effects of propositional matches [(propositional match > structural match) \cap (propositional match > structural mismatch) \cap (nondegraded reference sentence > degraded reference sentence)] and (ii) the contrast ''propositional mismatch > propositional match'' masked again inclusively with the above mentioned propositional mismatch effect ''(propositional mismatch > structural mismatch) \cap (propositional mismatch > structural match) \cap (nondegraded reference sentence > degraded reference sentence).''

In addition, we investigated the effect of matches versus mismatches across both types of priors by the conjunctions ''(propositional match > propositional mismatch) \cap (structural match > structural mismatch)'' and ''(propositional mismatch > propositional match) \cap (structural mismatch > structural match).''

RESULTS

Behavioral Results

Overall, the accuracy of identity judgments was comparable for targets following propositional and structural priors (83 and 79% correct responses, respectively). The 2×2 (Match: yes/no \times Type of Prior: propositional/structural) ANOVA of accuracy revealed a significant main effect of Match [$F(1, 28) = 29.93, P < 0.001$] and a significant Match \times Type of Prior interaction [$F(1, 28) = 12.02, P < 0.002$; Fig. 1A]. Type of Prior had no significant main effect

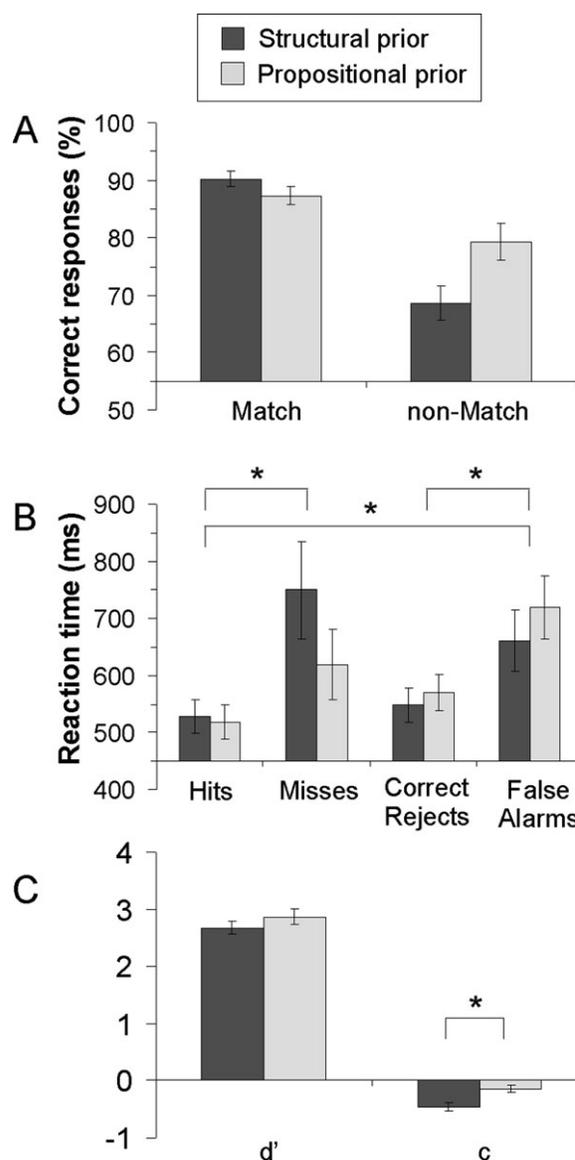


Figure 1.

Behavioral data. Accuracy (A) and reaction time (B) of target sentences following structural (dark gray) and propositional (light gray) priors. Panel C depicts measures of discriminability (d') and bias (c) as derived from same-different signal-detection analyses for the two types of prior. Positive d' values reflect discriminability above chance level. Negative c values indicate a bias toward ''same'' responses (i.e., a more lenient criterion), whereas positive c values signifies a bias toward ''different'' responses. Error bars represent the standard error of the means. * = significant at $P < 0.01$.

[$F(1, 28) = 3.86, p > .05$]. Thus, although accuracy was generally lower on nonmatch than match trials, this drop of accuracy was significantly stronger for targets following structural rather than propositional priors. An ANOVA of reaction time indicated a significant main effect of Match

[$F(1, 28) = 12.91, P = 0.001$]. Responses were faster for matching than nonmatching pairs, independent of Type of Prior (propositional/structural). When splitting match and nonmatch trials into correct and incorrect trials (i.e., hits, misses, correct rejections, false alarms), a 4(Category) \times 2(Type of Prior) ANOVA showed a main effect of Category [$F(3,18) = 10.99, P < 0.001$; Fig. 1B], indicating that in particular incorrect decision trials (misses and false alarms) were associated with slower responses. Specifically, after Bonferroni correction, hits (i.e., correct match decisions) were significantly faster than misses (i.e., incorrect nonmatch decisions), and correct rejections (i.e., correct nonmatch decisions) were faster than false alarms (i.e., incorrect match decisions). Furthermore, hits were also faster than false alarms. These results suggest that the above-reported main effect of Match on reaction time is mainly due to the fast responses on hit trials.

Signal-detection parameters for observer sensitivity (d') and decision criterion (c) were calculated to investigate differences in discriminability and response tendency, respectively (Fig. 1C). Paired t tests revealed no difference in d' between propositional and structural priors [$t(1, 28) = 1.19, P = 0.243$] but showed that c was significantly lower for structural than for propositional priors [$t(1, 28) = 3.89, P < 0.001$]. Thus, discrimination difficulty in trials with structural versus propositional priors was comparable, while responses following structural priors were relatively more biased toward “same” responses (i.e., match decisions) than responses following propositional priors. Finally, no sex differences were observed for any of the behavioral measures.

fMRI Results

General overview

Activations common to all six sentence types (nondegraded reference sentence \cap degraded reference sentence \cap structural match \cap structural mismatch \cap propositional match \cap propositional mismatch) were found bilaterally in temporal and frontal areas, in the supplementary motor area (SMA), premotor cortex, and cerebellum (Fig. 2A). Next, the contrast between nondegraded and degraded reference sentences revealed significant clusters in the left IFG [area 44, 45; Amunts et al., 1999], bilateral MTG/STG [TE1.0, TE 1.1; Morosan et al., 2001], precuneus, posterior cingulate cortex, hippocampus [SUB; Amunts et al., 2005], amygdala [CM; Amunts et al., 2005], thalamus, as well as dorsomedial and ventromedial prefrontal cortex (PFC; Fig. 2B). The inverse contrast (Fig. 2C) revealed higher activation in bilateral insula, middle frontal gyrus, and middle cingulate cortex, in right IFG (area 44), rolandic operculum [OP4; Eickhoff et al., 2006], and inferior parietal cortex, as well as left cerebellum, premotor cortex, and pre-SMA.

Furthermore, reaction times of the sentence comparisons were included as parametric modulators of the BOLD response to target sentence presentation in each of the

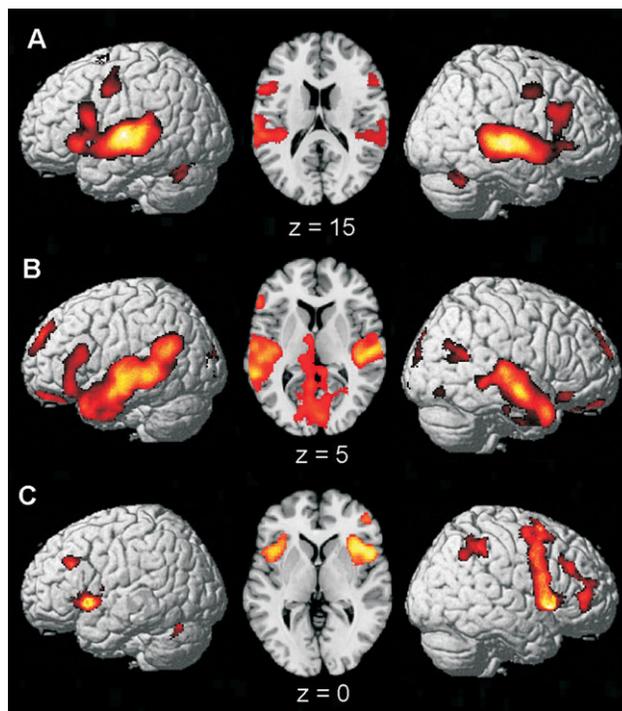


Figure 2.

Overview of the general fMRI findings. (A) The conjunction (nondegraded reference sentence \cap degraded reference sentence \cap structural match \cap structural mismatch \cap propositional match \cap propositional mismatch) revealed bilateral activations reflecting auditory processing common to all six sentence type events. (B) Regions representing the lexical-semantic rather than the prosodic aspects of speech (nondegraded reference sentence $>$ degraded reference sentence) and (C) the inverse contrast (degraded reference sentence $>$ nondegraded reference sentence). All images are thresholded at $P < 0.05$ (FWE-corrected at cluster-level; cluster forming threshold at voxel level: $P < 0.001$).

four conditions. No significant correlations between reaction time and BOLD response were observed in the group analysis.

Effect of propositional priors

To identify the general effect of exposure to propositional priors on the processing of (degraded) targets, we employed the following conjunction: [(propositional match + propositional mismatch) $>$ (structural match + structural mismatch)] \cap (nondegraded reference sentence $>$ degraded reference sentence). Three left-hemispheric clusters localized in the IFG (area 44/45), MTG and AG (PGa/PGp) resulted from this conjunction (Fig. 3A and Table II).

When focusing on activations related to propositional matches [(propositional match $>$ structural match) \cap (propositional match $>$ structural mismatch) \cap

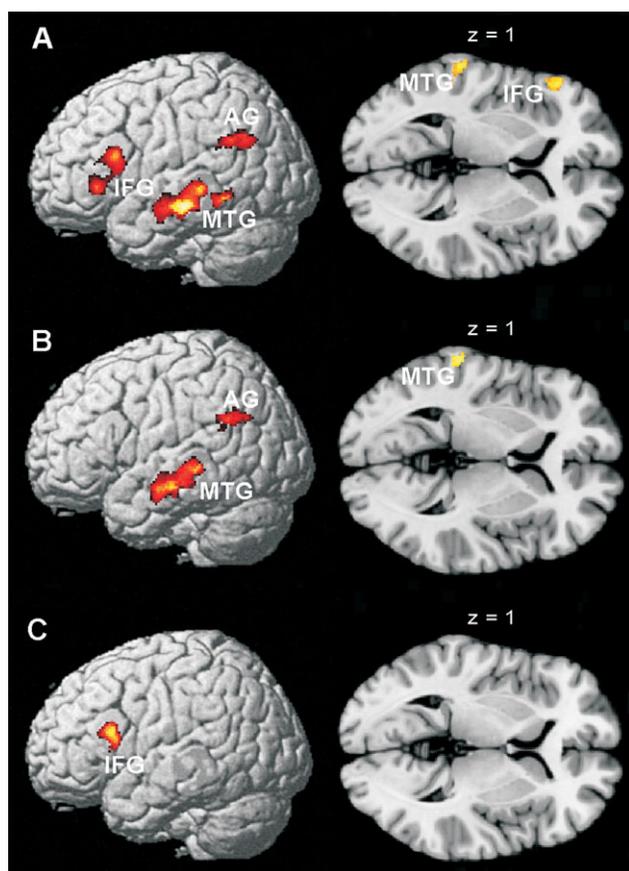


Figure 3.

Effects of propositional prior were all left-lateralized. **A)** Regions within the lexical-semantic network responding more to degraded targets that were preceded by propositional compared to structural priors “[$(\text{propositional match} + \text{propositional mismatch}) > (\text{structural match} + \text{structural mismatch}) \cap (\text{nondegraded reference sentence} > \text{degraded reference sentence})$].” Dissociation of this network into **B)** left MTG and AG for propositional matches [$(\text{propositional match} > \text{structural match}) \cap (\text{propositional match} > \text{structural mismatch}) \cap (\text{nondegraded reference sentence} > \text{degraded reference sentence})$] and **C)** Broca’s area for propositional mismatches [$(\text{propositional mismatch} > \text{structural mismatch}) \cap (\text{propositional mismatch} > \text{structural match}) \cap (\text{nondegraded reference sentence} > \text{degraded reference sentence})$]. All images are thresholded at $P < 0.05$ (FWE-corrected at cluster-level; cluster forming threshold at voxel level: $P < 0.001$). See also Supporting Information Figures S2 and S4.

(nondegraded reference sentence $>$ degraded reference sentence)], the resulting pattern no longer included Broca’s region (area 44/45) but was confined to the left AG and left MTG (Fig 3B and Table II). Broca’s region (area 44/45), however, was the only region associated with propositional mismatches [$(\text{propositional mismatch} > \text{structural$

match) \cap (propositional mismatch $>$ structural mismatch) \cap (nondegraded reference sentence $>$ degraded reference sentence)] (Fig. 3C and Table II). Direct comparison between propositional matches and mismatches revealed that Broca’s region (area 44/45) was the only cluster showing significantly increased activation in response to propositional mismatches compared to propositional matches. No such specific association could be established for the MTG; however, the AG was more strongly activated in response to propositional matches compared to propositional mismatches.

We repeated this analysis without including the contrast “nondegraded reference sentence $>$ degraded reference sentence” in the conjunctions. For the general effect of propositional prior and propositional match, we observed additional activations in some areas including the left inferior temporal gyrus, hippocampus, amygdala, basal ganglia, and cerebellum (see Supporting Information Fig. S2 for details).

Effects of reference–target match versus mismatch

Regions responding more strongly to reference–target matches than mismatches in general comprised the ventromedial PFC (vmPFC) and the bilateral putamen. Conversely, reference–target mismatches selectively activated the pre-SMA bilaterally as well as the right insula and right IFG (see Supporting Information Fig. S3).

Supplementary analysis

In addition to the main analysis based only on correct trials, a supplementary analysis was performed involving all trials. The results (see Supporting Information Table SI for details) were largely comparable to those of the main analysis; however, some interesting differences emerged. When testing for the effect of a propositional (relative to a structural) prior on the processing of a (degraded) target sentence, we again found the left MTG and IFG, but instead of activation in the left AG, we now observed significant activity in the left thalamus. According to the thalamic connectivity atlas (<http://www.fmrib.ox.ac.uk/connect>), 65% of this thalamic activation was localized in a section having connections with the prefrontal cortex, and 29% of the cluster was localized in a section connecting to the temporal cortex. These thalamic areas are thought to include the mediodorsal nucleus and nuclei of the anterior complex [Behrens et al., 2003]. Analogous to the main analysis, the comparison to targets with structural priors revealed that the left MTG and thalamus (60% of the cluster predominantly connecting to the prefrontal cortex, 36% predominantly connecting to the temporal cortex) were responsive to propositional matches, whereas Broca’s region (area 44/45) was responsive to propositional mismatches (see Supporting Information Fig. S4).

TABLE II. Overview of activations

Region	Cytoarchitectonic area (percent overlap)	x	y	z	Z-score	Cluster size
<i>Effect of propositional prior</i>						
L IFG	Area 44 ^a (48% overlap); Area 45 ^a (36% overlap)	-54	20	16	5.29	632
L MTG		-62	-26	-10	4.71	1443
L AG	PGa ^b (41% overlap); PGp ^b (18% overlap)	-46	-62	28	3.96	405
<i>Effect of matching propositional prior</i>						
L MTG		-62	-10	-15	4.68	1085
L AG	PGa ^b (51% overlap); PGp ^b (14% overlap)	-48	-54	30	3.96	297
<i>Effect of nonmatching propositional prior</i>						
L IFG	Area 44 ^a (58% overlap); Area 45 ^a (40% overlap)	-52	18	18	5.00	351

All activations $P < 0.05$ (cluster-level FWE-corrected). x , y , z coordinates refer to the peak voxel in MNI space. R, right; L, left; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; AG, angular gyrus.

^aAmunts et al., 1999.

^bCaspers et al., 2006.

DISCUSSION

This study investigated the behavioral and neural effects of propositional priors carrying lexical-semantic information on the decoding of degraded speech. As noted above, such decoding should reflect an interaction of sensory input and prior information via lexical-semantic predictions and meaning retrieval. We demonstrated that processing physically identical stimuli may result in distinct patterns of neural activation depending on the type of prior information available to the listener. In particular, prior propositional information provided by intelligible speech (compared to purely “structural” information provided by degraded speech) resulted in stronger recruitment of a left-lateralized network comprising the MTG, AG (PGa/PGp), and area 44/45 of Broca’s region. Within this network a direct comparison between propositional matches and mismatches revealed a selective association of activity in Broca’s region with propositional mismatches and a selective association of activity in the AG with propositional matches. A supplementary analysis based on all (instead of correct-only) trials indicated an involvement of the left thalamus (rather than left AG) with propositional priors.

Importantly, reaction time and observer sensitivity did not differ between trials with structural and propositional priors. Therefore, the fMRI results reported here are highly unlikely to be explained by different degrees of task difficulty. Furthermore, we would like to stress that the results are also very unlikely to arise primarily from the matching or recognition of prosody, as this process should be initiated by both types of prior information and is controlled for by the contrasts included in the conjunctions. Rather, the resulting activations most likely stem from the (attempted) lexical-semantic processing of the degraded target sentence when prior propositional information was provided, as this was the only difference between the conditions. Furthermore, in trials with matching propositional priors, this lexical-semantic processing should reflect the

subjective impression of understanding the target sentence. In our view, this perceptual phenomenon did not lead to an observable behavioral benefit compared to structural matches, because the two sentences of propositional match trials were physically not entirely identical, in contrast to structural match trials. This physical difference between reference and target in propositional match trials will have made the matching process more challenging, thereby reducing or even outweighing the (presumably) facilitatory effect of understanding the target sentence. We would moreover suggest that the behavioral benefit observed for propositional mismatches (compared to structural mismatches) might be due to the absence of the “sudden understanding” phenomenon normally associated with propositional matches: while targets in both propositional and structural mismatch conditions were physically different from the reference, the fact that the target could not be understood despite the nondegraded (propositional) prior should have provided a potent clue facilitating the (overall more difficult) mismatch decision under these circumstances, relative to trials with a nonmatching degraded (structural) prior.

According to the dual-stream model of language [Hickok and Poeppel, 2007; Hickok et al., 2011] processing of speech sounds recruits temporal lobe structures in a hierarchical dorsal-to-ventral fashion. While the core and belt auditory areas on the planum temporale process simpler aspects of sounds, the more ventrally located superior temporal gyrus and superior temporal sulcus (STG/STS) are more sensitive to complex amplitude and frequency modulations present in speech sounds. Even further ventrally, the MTG and ITG are thought to be involved in the more abstract analysis of semantic and syntactic features of speech. In accordance with this model, the results of the conjunction across all reference and target sentences indicated the involvement of the STG in response to the complex sound properties present across all sentence types. In contrast, comparing nondegraded with degraded reference sentences yielded activity in the MTG but not in the STG.

Of note, this contrast and the inverse contrast revealed an activity pattern very similar to that reported by Meyer et al. [2004] for comparing nondegraded and degraded speech. Additionally, other areas including the left AG, left IFG, precuneus, and posterior cingulate that resulted from the comparison of nondegraded vs. degraded speech are associated with the lexical-semantic analysis of meaningful speech [Binder et al., 2009; Price, 2010].

Furthermore, the left MTG together with the left AG (or, when considering all trials in the supplementary analysis, the left MTG and the left thalamus) were activated when target sentences matched the propositional prior, i.e., when meaning could potentially be decoded from a degraded sentence. This finding indicated that speech processing in the MTG, in line with Binder and Price [2001], does not depend on the physical properties of speech sounds conveyed by bottom-up signaling because it responded differentially to physically identical target sentences. Rather, the MTG was recruited when more abstract linguistic processing was enabled by a top-down application of stored lexical-semantic information stemming from the matching propositional prior. Indeed, the left MTG has been identified as a key region for semantic processing and meaning extraction [Binder et al., 2009; Price 2010]. Activation of this area has been observed in various lexical-semantic tasks ranging from comprehension of degraded sentences [e.g., Adank and Devlin, 2010; Davis and Johnsruide, 2003] to attempts to derive meaning from gestures supporting spoken speech [Dick et al., 2009; Hubbard et al., 2009]. In line with these findings, lesions of this region are associated with impairments in language comprehension [e.g., Dick et al., 2007; Dronkers et al., 2004].

In addition to the left MTG, also the AG in the left temporo-parietal junction has frequently been associated with semantic processing [Binder et al., 2009]. The AG, which corresponds to the cytoarchitectonic areas PGa and PGp [Caspers et al., 2006], is considered to be a heteromodal association area with access to higher-order concepts and long-term memory. The left AG has been suggested to provide top-down “semantic constraints” in language processing [Price, 2010] and may thus facilitate meaning extraction from ambiguous sentences [Obleser and Kotz, 2010]. Interestingly, Seghier et al. [2010] found that such a function is most likely attributable to the medial or ventral portion of the AG which corresponds well to the AG cluster observed in the current study. Thus, the selective AG activation on correct trials might be the origin of top-down signals mediating predictions that facilitate decoding of the degraded sentences and enabling correct match/non-match decisions based on lexical-semantic content. While both left MTG and AG showed stronger activation when lexical-semantic expectations were present and fulfilled, their specific contribution to the processing of the speech signal is probably not equivalent. The MTG has been proposed to be involved in mapping sound (represented in the STS) to meaning [which is thought to be distributed

throughout the cortex; Hickok and Poeppel, 2007] and thereby enabling comprehension of speech signals.

The left AG, on the other hand, is thought to be a hierarchically higher node [Binder et al., 2009] which aids language processing by top-down modulation [Price, 2010; Seghier et al., 2010]. Potentially, this top-down influence might have been more pronounced on trials with clearer evidence that consequently could be answered correctly. Alternatively however, top-down modulation originating in the AG might also have rendered the evidence clearer and might have been a precondition for correct match/nonmatch decisions.

While the pattern of activation in the AG indicated that activity in this region is mainly linked with correct trials, we observed significant thalamic activation only in the supplementary analysis. This suggests that both correctly and incorrectly answered trials contributed to the observed thalamic activation but that limiting the analysis to the correct trials might have provided insufficient statistical power to detect the thalamic activation. The effect of excluding ~20% of trials in the main analysis may have manifested itself particularly in a small structure such as the thalamus, especially in combination with the cluster threshold we used. Indeed, the thalamic activation was significant in the main analysis, too, when slightly lowering the cluster-forming threshold. Nonetheless, we think that the thalamic activity is an interesting finding that deserves closer attention. Presumably, this activation could be due to top-down modulation of sensory processing by signals from temporal or frontal areas. Indeed, cortico-thalamic feedback is known to influence thalamic responses to auditory stimulation by amplifying those sensory features that optimally represent the signal predicted by cortical areas and inhibit all other response features [Alitto and Usrey, 2003; Suga et al., 2002]. Furthermore, based on patient studies associating thalamic lesions with language deficits, Nadeau and Crosson [1997] proposed that thalamic nuclei can selectively gate and integrate the flow of lexical information between frontal and temporo-parietal cortices and regulate the access to lexical information when semantic input is provided. More recent support for a thalamic involvement in language processing beyond the relay of auditory information has been found in electrophysiological studies implicating cortico-thalamic networks in processing the semantic and syntactic features of spoken sentences [David et al., 2011; Wahl et al., 2008], in fMRI studies reporting stronger thalamic responses to normal compared to (unintelligible) prosodic speech [Kotz et al., 2003] and demonstrating a thalamic contribution to resolving ambiguity of linguistic input [Ketteler et al., 2008].

In contrast to the left MTG, AG and (when considering all trials) the thalamus, the left IFG was selectively activated in response to mismatches between the target sentence and prior propositional information, i.e., when attempts to decode the degraded sentence based on the propositional prior failed. The activation was localized in

a portion of the IFG that has been cytoarchitecturally defined as area 44/45 [Broca's region; Amunts et al., 1999] and is known to play a role in speech perception, in particular when speech is syntactically complex [Friederici, 2011; Friederici et al., 2010]. This region, however, is not restricted to language processing but appears to be generally involved in the sequencing of spatiotemporal structures of various modalities including language, music and action [Fadiga et al., 2009]. Importantly, left IFG activation has also been associated with detecting incompatibility in speech and other hierarchically organized sequences [Embick et al., 2000; Friederici et al., 2010; Myers et al., 2009] which might be interpreted as prediction errors signaling the need for reanalysis to prevent misinterpretation [Christensen and Wallentin, 2011; Novick et al., 2005; Price, 2010]. Furthermore, Giraud et al. [2004] highlighted the role of Broca's region for search of meaningful content in auditory input. Accordingly, the involvement of Broca's region in the "propositional prior" network might arise from the potential meaningful content provided by nondegraded reference sentences compared to degraded ones. In line with the above reasoning, mismatches between the target and the preceding nondegraded sentence might have evoked even stronger activation in Broca's region because the incoming signal was incompatible with the expected sequence of auditory events while the presence of meaningful content was hard to determine. Of interest, activation of Broca's region has also been linked with effects of complexity and task difficulty [Fadiga et al., 2009]. The behavioral data indicated that trials with propositional mismatches were more difficult to discriminate than trials with propositional matches. However, this held also true for structural mismatches. Furthermore, the contrast between all mismatch and match trials demonstrated that in particular the pre-SMA (in addition to the right insula and right IFG) was associated with the overall effect of the higher task difficulty and response conflict associated with mismatches. Therefore, we would suggest that the selective activation of Broca's region in response to propositional mismatches reflects its specific involvement in challenging linguistic tasks, namely when the presence of meaning is hard to determine (search for meaning in noise) and an attempt is made to decode this potential meaning (reanalysis and possibly prevention of misinterpretation).

However, it should also be noted that Broca's region is a multifunctional area [see Rogalsky and Hickok, 2011, for a recent review]. The present study cannot definitively determine the exact mechanism reflected by the activation of Broca's region, and alternative accounts such as the phonological working-memory function of Broca's region cannot completely be ruled out. Nonetheless, if Broca's region merely reflected the storage and inner rehearsal of auditory speech stimuli, the comparison of trials with propositional versus structural priors should not evoke activity in this region because these working-memory processes are also required for matching decisions purely

based on the prosodic speech stimuli. Therefore, it seems more likely that the recruitment of Broca's region is due to the influence of lexical-semantic content provided by (particularly nonmatching) propositional priors on the subsequent processing of the target sentences, as this is the only aspect that distinguishes these conditions.

While Broca's region thus showed higher sensitivity for mismatches with propositional priors, the left MTG and AG (or, when considering all trials, the left MTG and the left thalamus) were implicated in lexical-semantic processing and meaning extraction by means of using prior information. When an informative propositional template is available, originally incomprehensible speech stimuli can be subjected to a more profound analysis. That is, prior exposure to the intelligible original sentence results in a dramatic change in the appraisal of a hitherto meaningless speech-like auditory stimulus that can suddenly be perceived as a salient and meaningful sentence. Such an expectancy-guided reappraisal of formerly noisy and meaningless sensory stimuli corresponds well to the notion of predictive coding which proposes that the brain actively participates in the perceptual process by anticipating upcoming events [Friston, 2005; Rao and Ballard, 1999]. This inferential process may result in striking effects of prior information on perception, even if the perceived stimuli are physically identical [Hunter et al., 2010].

Conceptually, accounts of predictive coding in perception [Friston, 2005; Rao and Ballard, 1999] assert that sensory predictions are generated at each level of the cortical hierarchy based on integration of prior knowledge with neural activity from lower levels. These predictions are thought to be fed back to lower levels where they are compared to the actual neural activity representing the sensory data. Differences between predicted and observed information are fed forward to the hierarchically higher node as the prediction error. This prediction error, in turn, is used to optimize subsequent predictions, as it indicates the fit of the current priors. Therefore, when a prediction fits well with the incoming sensory data, potential ambiguities among the stimuli can be resolved because the perceptual alternatives are weighted by the predicted template. With regard to the current study, propositional templates could only be successfully employed for decoding a degraded target sentence when (1) the reference sentence was a nondegraded sentence (propositional prior) and (2) the target sentence matched that reference sentence. Presumably, interactions within a left-hemispheric network including the AG and the MTG are an important generator of these lexical-semantic predictions. Possibly, these predictions were sent to lower levels of the auditory processing hierarchy and potentially modified the response profile in the left thalamus. Alternatively, it is also possible that top-down feedback from the AG affected the processing in the MTG such that sound could be successfully mapped to meaning, resulting in the percept of an intelligible sentence. The importance of these regions is supported by previous studies reporting involvement of

left temporal areas in successful decoding of originally unintelligibly degraded speech stimuli [Eulitz and Hanne- mann, 2010; Giraud et al., 2004; Hannemann et al., 2007]. In contrast, unsuccessful decoding attempts of target sentences following a mismatching propositional prior were selectively associated with activation of Broca's region. This could indicate that Broca's region contributes to speech perception by searching for meaningful content and comparing expected auditory sequences with the actual input. Consequently, activation of Broca's region might represent the prediction error when prior knowledge cannot be used to decode meaningful speech due to mismatches between propositional priors and degraded targets. Alternatively, the comparison between the predicted and the actual signal might also happen elsewhere in the brain. In this case, involvement of Broca's region could reflect the updating of the expectations to be generated. Finally, a response to these mismatches might also signal the need for reanalysis of the auditory sequence to prevent misinterpretation. This signal might then lead to a reduced involvement of the left MTG and AG in processing propositional mismatches. Thus, top-down influences of Broca's region on the MTG in particular but also on the AG might be relevant for the perceptual phenomenon of sudden understanding of a heavily degraded sentence and, at the same time, the lack of comprehension when prior knowledge cannot be applied to such degraded input. However, the exact mechanisms cannot be determined with the current analysis but would ultimately require dynamic causal modeling or related approaches.

Although we interpret our results within the framework of predictive coding, we do not claim that this framework is the only one that can account for the findings. Alternatively, it might also be warranted to refer to priming mechanisms to explain the recognition of degraded sentences by an exposure effect of the original stimulus. However, predictive coding is the more generic framework, encompassing all kinds of contextual effects on perception, ranging from subliminal priming to instructed expectations. Accordingly, we prefer to interpret our findings in the context of this more general model of brain function, although the current experiment did not aim to test the predictive-coding account itself.

CONCLUSION

The current study demonstrated that, in line with the notion of predictive coding, prior information has a decisive effect on speech perception. In particular, the processing of degraded sentences, which were incomprehensible when heard in isolation, was shown to be sensitive to the availability of memory templates ("priors") carrying propositional (i.e., lexical-semantic) information. It may be assumed that this effect results from a combination of direct meaning decoding, based on lexical-semantic predictions, and meaning retrieval, based on a structural match

with previous speech input. Prior propositional information also influenced the neural response to degraded sentences and hence revealed the neural correlates of these processes. Specifically, successful meaning extraction from degraded sentences based on prior information was associated with increased activity in the left MTG and AG. These areas are known to play a role in high-level lexical-semantic processing that presumably modulates more basic speech processing at lower levels including the thalamus. In contrast, unsuccessful decoding due to a misleading propositional prior was selectively associated with stronger activation of Broca's region which may reflect the search for relevant acoustic cues in auditory sequences as well as processing prediction errors, thereby potentially preventing misinterpretation.

REFERENCES

- Adank P, Devlin JT (2010): On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage* 49:1124–1132.
- Alitto HJ, Usrey WM (2003): Corticothalamic feedback and sensory processing. *Curr Opin Neurobiol* 13:440–445.
- Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, Habel U, Schneider F, Zilles K (2005): Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat Embryol (Berl)* 210:343–352.
- Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings HB, Zilles K (1999): Broca's region revisited: Cytoarchitectonic and intersubject variability. *J Comp Neurol* 412:319–341.
- Amunts K, Schleicher A, Zilles K (2007): Cytoarchitecture of the cerebral cortex—More than localization. *Neuroimage* 37: 1061–1065.
- Arnal LH, Wyart V, Giraud AL (2011): Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14:797–803.
- Ashburner J, Friston KJ (2005): Unified segmentation. *Neuroimage* 26:839–851.
- Bar M (2007): The proactive brain: Using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289.
- Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM (2003): Noninvasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* 6:750–757.
- Binder JR, Price CJ (2001): Functional imaging of language. In: Cabeza R, Kingstone A, editors. *Handbook on Functional Neuroimaging of Cognition*. Cambridge (MA): MIT Press. pp 187–251.
- Binder JR, Desai RH, Graves WW, Conant LL (2009): Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Blakemore SJ, Wolpert DM, Frith CD (1998): Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1:635–640.
- Caspers S, Geyer S, Schleicher A, Mohlberg H, Amunts K, Zilles K (2006): The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *Neuroimage* 33:430–448.

- Christensen KR, Wallentin M (2011): The locative alternation: Distinguishing linguistic processing cost from error signals in Broca's region. *Neuroimage* 56:1622–1631.
- David O, Maess B, Eckstein K, Friederici AD (2011): Dynamic causal modelling of subcortical connectivity of language. *J Neurosci* 31:2712–2717.
- Davis MH, Johnsrude IS (2003): Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Davis MH, Johnsrude IS (2007): Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hear Res* 229:132–147.
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005): Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen* 134: 222–241.
- den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009): A dual role for prediction error in associative learning. *Cereb Cortex* 19:1175–1185.
- den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010): Striatal prediction error modulates cortical coupling. *J Neurosci* 30:3291–3209.
- Dick AS, Goldin-Meadow S, Hasson U, Skipper JL, Small SL (2009): Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum Brain Mapp* 30:3509–3526.
- Dick F, Saygin AP, Galati G, Pitzalis S, Bentrovato S, D'Amico S, Wilson S, Bates E, Pizzamiglio L (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* 19:799–816.
- Dronkers NF, Wilkins DP, van Valin RD Jr, Redfern BB, Jaeger JJ (2004): Lesion analysis of the brain areas involved in language comprehension. *Cognition* 92:145–177.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005): A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Eickhoff SB, Schleicher A, Zilles K, Amunts K (2006): The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. *Cereb Cortex* 16:254–267.
- Embick D, Marantz A, Miyashita Y, O'Neil W, Sakai KL (2000): A syntactic specialization for Broca's area. *Proc Natl Acad Sci USA* 97:6150–6154.
- Enns JT, Lleras A (2008): What's next? New evidence for prediction in human vision. *Trends Cogn Sci* 12:327–333.
- Eulitz C, Hannemann R (2010): On the matching of top-down knowledge with sensory input in the perception of ambiguous speech. *BMC Neurosci* 11:67.
- Fadiga L, Craighero L, D'Ausilio A (2009): Broca's area in language, action, and music. *Ann N Y Acad Sci* 1169:448–458.
- Friederici AD (2011): The brain basis of language: From structure to function. *Physiol Rev* 91:1357–1392.
- Friederici AD, Kotz SA, Scott SK, Obleser J (2010): Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp* 31:448–457.
- Friston K (2005): A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004): Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex* 14:247–255.
- Grossberg S (2003): Resonant neural dynamics of speech perception. *J Phon* 31:423–445.
- Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140–149.
- Hannemann R, Obleser J, Eulitz C (2007): Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Res* 1153:134–143.
- Hickok G, Houde J, Rong F (2011): Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69:407–422.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hosoya T, Baccus SA, Meister M (2005): Dynamic predictive coding by the retina. *Nature* 436:71–77.
- Hubbard AL, Wilson SM, Callan DE, Dapretto M (2009): Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Hum Brain Mapp* 30:1028–1037.
- Hunter MD, Eickhoff SB, Pheasant RJ, Douglas MJ, Watts GR, Farrow TF, Hyland D, Kang J, Wilkinson ID, Horoshenkov KV, Woodruff PW (2010): The state of tranquillity: Subjective perception is shaped by contextual modulation of auditory connectivity. *Neuroimage* 53:611–618.
- Jakobs O, Wang EW, Dafotakis M, Grefkes C, Zilles K, Eickhoff SB (2009): Effects of timing and movement uncertainty implicate the temporo-parietal junction in the prediction of forthcoming motor actions. *Neuroimage* 47:667–677.
- Kaas JH, Hackett TA (2000): Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97:11793–11799.
- Ketteler D, Kastrau F, Vohn R, Huber W (2008): The subcortical role of language processing. High level linguistic features such as ambiguity-resolution and the human brain: An fMRI study. *Neuroimage* 39:2002–2009.
- Kiebel SJ, Glaser DE, Friston KJ (2003): A heuristic for the degrees of freedom of statistics based on multiple variance parameters. *Neuroimage* 20:591–600.
- Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD (2003): On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain Lang* 86: 366–376.
- Ludmer R, Dudai Y, Rubin N (2011): Uncovering camouflage: Amygdala activation predicts long-term memory of induced perceptual insight. *Neuron* 69:1002–1014.
- Macmillan NA, Creelman CD (2005). *Detection Theory: A User's Guide*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002): fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17:73–88.
- Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004): Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang* 89:277–289.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001): Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13:684–701.
- Myers EB, Blumstein SE, Walsh E, Eliassen J (2009): Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol Sci* 20:895–903.
- Nadeau SE, Crosson B (1997): Subcortical aphasia. *Brain Lang* 58:355–402.

- Nichols T, Brett M, Anderson J, Wager T, Poline JB (2005): Valid conjunction inference with the minimum statistic. *Neuroimage* 15:653–660.
- Novick JM, Trueswell JC, Thompson-Schill SL (2005): Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5:263–281.
- Obleser J, Kotz SA (2010): Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex* 20:633–640.
- Porter P (1954): Another picture-puzzle. *Am J Psychol* 67:550–551.
- Price CJ (2010): The anatomy of language: A review of 100 fMRI studies published in 2009. *Ann NY Acad Sci* 1191:62–88.
- Rao RP, Ballard DH (1999): Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Rauschecker JP, Tian B, Hauser M (1995): Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114.
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM (1994): On the perceptual organization of speech. *Psychol Rev* 101:129–156.
- Rogalsky C, Hickok G (2011): The role of Broca's area in sentence comprehension. *J Cogn Neurosci* 23:1664–1680.
- Saberi K, Perrott DR (1999): Cognitive restoration of reversed speech. *Nature* 398:760–761.
- Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123:2400–2406.
- Seghier ML, Fagan E, Price CJ (2010): Functional subdivisions in the left angular gyrus where the semantic system meets and diverges from the default network. *J Neurosci* 30:16809–16817.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995): Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Sharma J, Dragoi V, Tenenbaum JB, Miller EK, Sur M (2003): V1 neurons signal acquisition of an internal representation of stimulus location. *Science* 300:1758–1763.
- Suga N, Xiao Z, Ma X, Ji W (2002): Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36:9–18.
- Wahl M, Marzinzik F, Friederici AD, Hahne A, Kupsch A, Schneider GH, Saddy D, Curio G, Klostermann F (2008): The human thalamus processes syntactic and semantic language violations. *Neuron* 59:695–707.
- Zilles K, Amunts K (2010): Centenary of Brodmann's map—Conception and fate. *Nat Rev Neurosci* 11:139–145.