NeuroImage 70 (2013) 89–100

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Predictions in speech comprehension: fMRI evidence on the meter–semantic interface

Kathrin Rothermich [a,b], Sonja A. Kotz [b,*]

[a] International Laboratory for Brain, Music and Sound Research (BRAMS), Pavillon 1420 Boul. Mont Royal, Université de Montréal, Case Postale 6128, Station Centre-Ville, Montréal, Quebec, Canada H3C 3J7
[b] Research Group "Subcortical Contributions to Comprehension", Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

ARTICLE INFO

Article history:
Accepted 8 December 2012
Available online 3 January 2013

Keywords:
fMRI
Speech comprehension
Prediction
Facilitation

ABSTRACT

When listening to speech we not only form predictions about *what* is coming next, but also *when* something is coming. For example, metric stress may be utilized to predict the next salient speech event (i.e. the next stressed syllable) and in turn facilitate speech comprehension. However, speech comprehension can also be facilitated by semantic context, that is, which content word is likely to appear next. In the current fMRI experiment we investigated (1) the brain networks that underlie metric and semantic predictions by means of prediction errors, (2) how semantic processing is influenced by a metrically regular or irregular sentence context, and (3) whether task demands influence both processes. The results are three-fold: First, while metrically incongruent sentences activated a bilateral fronto-striatal network, semantically incongruent trials led to activation of fronto-temporal areas. Second, metrically regular context facilitated speech comprehension in the left-fronto-temporal language network. Third, attention directed to metric or semantic aspects in speech engaged different subcomponents of the left inferior frontal gyrus (IFG). The current results suggest that speech comprehension relies on different forms of prediction, and extends known speech comprehension networks to subcortical sensorimotor areas.

Introduction

One of the most fascinating features of the human brain is its ability to recognize regular patterns in a dynamically changing environment. For example, speech and music rhythms form recurring temporal patterns of energy fluctuations that optimize their perception. Likewise, perceptual regularity established by a regular rhythm leads to the prediction of upcoming salient auditory events (i.e. sounds, tones, speech units) and to the direction of attention to predicted points in time (Large and Jones, 1999). Ultimately, it has been proposed that prediction, realized in speech by repetitive prosodic information such as a regular rhythm, allows for an economy of resources, leads to faster recognition and facilitates the integration of information (Bubic et al., 2010; Cason and Schön 2012; Martin, 1972).

In speech comprehension research, the role of rhythm as a source of predictions has only recently benefitted from renewed scientific interest. Traditionally, rhythmic and metric patterns are considered to be phonological properties of speech that shape speech segmentation (Endress and Hauser, 2010; Lee and Todd, 2004), word recognition (e.g. Cutler and Norris, 1988), and language acquisition (e.g. Jusczyk et al., 1999). In the present fMRI study, we pursued the goal to determine whether speech rhythm can also (i) influence the integration of words into a sentence context, and (ii) facilitate integration by utilizing metric and semantic predictions.

Previous research suggests that speech contains regularities encoded in several acoustic features such as pitch, duration, and amplitude that allow to group speech into smaller segments (Dilley and Pitt, 2010; Yoshida et al., 2010). Languages differ in how acoustic regularities are composed, but a 'phonological grammar' may be built upon a dynamic temporal system that adapts to acoustic regularities by means of periodic oscillations in speech (Port, 2003) or in communicative interactions (Hasson et al., 2012). Importantly, in stress-time languages such as German, regularities comprise the alternation of strong and weak syllables (i.e. the metric foot) that form the metric structure in speech. Syllable alternations are very prominent and enhance the perception of regularity in speech, thereby forming temporal predictions about when the next prominent syllable will occur (Domahs et al., 2008; Magne et al., 2010; Rothermich et al., 2010). These temporal predictions may arise even if the distribution of stressed syllables does not include *precise* temporal information (Lidji et al., 2011; Rothermich et al., 2012; Schmidt-Kassow and Kotz, 2009).

In addition to temporal predictions, semantic predictions, derived from knowledge and context, also impact word integration. Some psycholinguistic research suggests that all available linguistic information (e.g. semantic, syntactic, and prosodic) is used to anticipate upcoming events during speech comprehension (e.g., McClelland and Elman, 1986). Consequently, processing ease, even prior to lexical access or selection, will be influenced by information that has been pre-activated in

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany. Fax: +49 341 9940113.
E-mail address: kotz@cbs.mpg.de (S.A. Kotz).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.12.013

a given context (Federmeier, 2007; Kutas and Federmeier, 2011). Similarly, Delong et al. (2005) proposed that words and phrases differentially trigger future expectancies via rich semantic representations. Taken together, both metric ("when") and semantic ("what") predictions influence and facilitate dynamic speech comprehension.

The goal of the current experiment was to investigate how manipulations of metric and semantic predictions impact speech comprehension at the neuronal level. We therefore manipulated the level of metric and semantic expectancy in an auditory sentence context by presenting words with an unexpected stress pattern (metrically incongruent condition) and semantically unexpected words that do not fit in the semantic context (semantically incongruent condition). Additionally, to test the effect of regular speech rhythm on semantic and metric integration we created two different kinds of sentence contexts (metrically regular and metrically irregular) by varying the alternation of stressed and unstressed syllables in words preceding the critical items. If a regular metric structure allows to temporally predict future items, this should lead to reduced activation to semantically or metrically unexpected items particularly in frontal and temporal brain regions (commonly interpreted as facilitation; Obleser and Kotz, 2010; Rossell et al., 2003; Wheatley et al., 2005).

For metrically incongruent compared to congruent trials we expected to see increased activation in an extended fronto-striatal and an extended cortico-cerebello-cortical network (e.g. for speech: Aleman et al., 2005; Geiser et al., 2008, 2012; Klein et al., 2011; Kotz and Schwartze, 2010; for music: Chen et al., 2008; Grahn and Brett, 2007). This network consists of classical language areas such as the bilateral posterior inferior frontal gyrus (IFG), as well as the bilateral superior temporal gyrus. Next to fronto-temporal regions, regions associated with motor behavior and timing such as the supplementary motor area (SMA), the basal ganglia, the cerebellum, the anterior insula, and the thalamus are implied in the perception and evaluation of rhythm. Second, semantically incongruent trials should increase activation in a left fronto-temporal network (e.g. Bookheimer, 2002; Friederici et al., 2003; Kuperberg et al., 2000; Lau et al., 2008; Newman et al., 2001; Ni et al., 2000; Rissman et al., 2003), including the left IFG, (notably in BA 47/45) as well as the middle temporal gyrus (MTG). More specifically, we hypothesized that the IFG together with the middle temporal cortex subserves top-down processes drawing predictions about incoming information, thereby easing its integration (Friederici, 2011). As already mentioned, by varying the rhythmic properties of sentence context (metrically regular vs. irregular context) we further aimed to test whether regular speech rhythm leads to facilitation effects in classical speech comprehension areas and beyond (see Pickering and Garrod, 2007 for a similar argument on context manipulation) and whether such facilitation varies as a function of attentional task demands.

In summary, we aimed to elucidate how metric and semantic predictions modulate neuronal activation within the speech comprehension network, and how this network may extend to sensorimotor circuitries when focusing on the analysis of the metric/temporal structure of speech. Thus, we hypothesized that a cortico-subcortico-cortical network is involved in the processing of metric and rhythmic characteristics of the auditory speech signal. Manipulation of semantic and metric predictions could also induce common changes in left frontal and temporal cortices revealing areas of integration for both information types (i.e. Geiser et al., 2008; Lau et al., 2008).

Methods

Participants

Sixteen right-handed participants (native speakers of German, 8 female, mean age of 26 yrs, S.D. 3,8), with no neurological history took part in the study. They received eight Euros per hour for compensation. After being informed about potential risks and screening by a physician of the Max Planck Institute for Cognitive and Brain Sciences (Leipzig, Germany) participants gave informed and written consent. The experimental procedures were approved by the local ethics committee of the University of Leipzig.

Stimulus material

Speech stimuli included 152 metrically regular sentences (38 per condition) consisting of bi-syllabic trochaic words (stress on the first syllable) as well as 152 metrically irregular sentences with words varying in syllable number and stress pattern, resulting in a total of 304 stimuli (see also Rothermich et al., 2012). In the metrically regular sentences, the metrical context leading up to a critical word consisted of five bi-syllabic words with stress on the first syllable. The stress pattern in the irregular sentences consisted of five words with one to three syllables each and with stress on the first, second or third syllable. Aside from the preceding metric sentence context, all material was the same. Stimuli were tested in four conditions: congruent, semantically incongruent, metrically incongruent and doubly incongruent. In the semantic condition, the critical noun was exchanged with a grammatically acceptable, but semantically unexpected noun. In the metric condition, the expected trochaic stress pattern was replaced by an iambic pattern. In the double condition, a semantically incongruent noun with an iambic pattern was presented (see Table 1 for examples).

All verbs as well as the preceding adverbs were matched for word frequency (according to the Leipziger Wortschatz Lexikon[1]). Sentences were spoken by a professional female German speaker at a normal speech rate and digitally recorded with a 16-bit resolution at a sampling rate of 44,100 Hz. The speaker was instructed to avoid overemphasizing the regular stress pattern. An extensive splicing procedure was applied to avoid co-articulation artifacts with respect to unexpected stressed words and to familiarize the speaker with the incongruent stress pattern in the metric conditions (see example in Table 2).

We constructed phrases consisting of three words, with the first word adhering to an iambic pattern to facilitate the speaker's expression of an incongruent metric word pattern. The last two words remained the same as in the original sentence. Then the two nouns were removed from the signal and inserted into the original sentence. Note that the same procedure was performed for the correctly pronounced sentences, in which the first word of the three-word phrases was trochaic. This ensured equal treatment of congruent and incongruent sentences.

Experimental procedure

All experimental trials were presented in two sessions, a metric task session and a semantic task session, and the sessions were acquired on two days separated by at least 30 days. Participants judged either the metric congruency (explicit task with respect to meter, implicit task with respect to semantics) or the semantic congruency of the presented sentences. The order of the tasks was counterbalanced across participants. The experimental design was a 2×4 design with the factor *condition* (congruent, semantically incongruent, metrically incongruent, and doubly incongruent) and the factor *context* (metrically regular, metrically irregular) with 38 trials each. In total, 304 trials were presented per session in 38 blocks consisting of 8 Trials. Each block contained either metrically regular or metrically irregular trials and were presented in a pseudo-randomized order. Pauses between the blocks (8–24 s), in which no stimulus was presented, served as null-events. The total trial length was 8 s; one experimental session lasted approximately 48 min. To allow for measurements at numerous time points along the BOLD signal curve to

[1] The Leipzig Wortschatz Lexikon is a corpus that is permanently updated using publicly available texts. It was developed and is maintained by the Department of Computer Science of the University of Leipzig and can be accessed at http://www.wortschatz.uni-leipzig.de.