

Expectations about word stress modulate neural activity in speech-sensitive cortical areas

Ferenc Honbolygó^{a,b,*}, Andrea Kóbor^{a,1}, Petra Hermann^a, Ádám Ottó Kettinger^{a,c}, Zoltán Vidnyánszky^a, Gyula Kovács^{a,d,1}, Valéria Csépe^{a,e}

^a Brain Imaging Centre, Research Centre for Natural Sciences, Budapest, Hungary

^b Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

^c Department of Nuclear Techniques, Budapest University of Technology and Economics, Budapest, Hungary

^d Department of Biological Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Jena, Germany

^e Faculty of Modern Philology and Social Sciences, University of Pannonia, Veszprém, Hungary

ARTICLE INFO

Keywords:

Speech
fMRI
Word stress
Predictive processing
Repetition suppression

ABSTRACT

A recent dual-stream model of language processing proposed that the postero-dorsal stream performs predictive sequential processing of linguistic information via hierarchically organized internal models. However, it remains unexplored whether the prosodic segmentation of linguistic information involves predictive processes. Here, we addressed this question by investigating the processing of word stress, a major component of speech segmentation, using probabilistic repetition suppression (RS) modulation as a marker of predictive processing. In an event-related acoustic fMRI RS paradigm, we presented pairs of pseudowords having the same (Rep) or different (Alt) stress patterns, in blocks with varying Rep and Alt trial probabilities. We found that the BOLD signal was significantly lower for Rep than for Alt trials, indicating RS in the posterior and middle superior temporal gyrus (STG) bilaterally, and in the anterior STG in the left hemisphere. Importantly, the magnitude of RS was modulated by repetition probability in the posterior and middle STG. These results reveal the predictive processing of word stress in the STG areas and raise the possibility that words stress processing is related to the dorsal “where” auditory stream.

1. Introduction

The human brain is best viewed as an inference machine, actively predicting and explaining its sensations through internal representations modeling the dynamic sensory context (Friston, 2010). One human-specific cognitive faculty where predictive processing may be especially important is linguistic communication (Donhauser and Baillet, 2019; Kuperberg and Jaeger, 2016; Lau et al., 2016; Willems et al., 2016). Predictive inference have been specifically integrated into a recent neurobiological model of language processing (Bornkessel-Schlesewsky et al., 2015; Bornkessel-Schlesewsky and Schlewsky, 2013). This model proposes a dual auditory stream network involving ventral and dorsal streams similarly to previous models (Friederici, 2011; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Saur et al., 2008; Scott et al., 2000), but suggests slightly different functions related to these streams. The antero-ventral or “what” stream of the

linguistic network (including the primary auditory cortex, the anterior superior temporal cortex, and anterior and ventral parts of the inferior frontal cortex) is thought to be responsible for the recognition of linguistic elements in an order-insensitive way, while the postero-dorsal or “where” stream (including the primary auditory cortex, the posterior superior temporal cortex, the inferior parietal lobule, the premotor cortex, and posterior and dorsal parts of the inferior frontal cortex) performs predictive sequential processing of linguistic information in successively larger temporal windows related to different linguistic levels (sounds, words, sentences, discourse). This predictive sequential processing is suggested to be based on hierarchically organized internal models, corresponding to temporal receptive windows that allow the processing of linguistic information at different time scales.

The model suggests that one of the dorsal stream’s functions is the prosodic segmentation of input. Prosody, the melodic and rhythmic aspect of speech (Cutler et al., 1997), contributes to speech

* Corresponding author. Magyar tudósok krt. 2., 1117, Budapest, Hungary.

E-mail address: honbolygo.ferenc@tk.hu (F. Honbolygó).

¹ These authors contributed equally to this work.

understanding at different levels: at the sentence level, intonation modifies the interpretation of the sentence (Friederici et al., 2007; Männel and Friederici, 2011; Sammler et al., 2015; Steinhauer et al., 1999; van der Burght et al., 2019), while at the word level, word stress plays a major role in the segmentation of continuous speech input into words (Cutler and Norris, 1988; Mattys et al., 2005; Norris et al., 1995; van Donselaar et al., 2005).

In accordance with the dual auditory stream model, previous research provided evidence that the postero-dorsal stream contributes to the prosodic segmentation of linguistic input. Particularly, intonation and discourse processing elicited increased BOLD responses in the posterior superior temporal gyrus (STG) and inferior frontal gyrus (IFG) (Geiser et al., 2008; Inspector et al., 2013; Ischebeck et al., 2008; Kandylaki et al., 2016; Meyer et al., 2004; Sammler et al., 2018, 2015). Furthermore, word stress processing has been associated with activations in the STG/superior temporal sulcus (STS), together with other areas like the IFG, SMA (supplementary motor area), and areas in the parietal (angular gyrus, superior parietal gyrus, parietal lobule) and frontal lobes (precentral, postcentral, and middle frontal gyrus); most of which could be assumed to be part of the dorsal stream (Aleman et al., 2005; Domahs et al., 2013; Heisterueber et al., 2014; Kandylaki et al., 2017; Klein et al., 2011).

Meanwhile, it remains an open question whether predictive processes are involved in the prosodic segmentation of linguistic input and specifically in words stress processing. Previous studies indicated that the processing of the prominence of words at the sentence level was guided by the acoustic and lexical predictability of words (Kakouros and Räsänen, 2016; Magne et al., 2005). In the short term, the perception of prominence was modified by the preceding prosodic exposure (Kakouros et al., 2018). Moreover, ERP evidence suggested the role of long-term expectations in the processing of stress at the word level (Honbolygó and Csépe, 2013). However, direct investigation of predictive processing of word stress related to cortical regions is missing.

To address this question, we used a possible neural marker of prediction, the probabilistic modulation of the fMRI repetition suppression (RS) effect (Summerfield et al., 2008). fMRI RS refers to reduced BOLD responses to repeated sensory stimuli (Henson and Rugg, 2003; Grill-Spector et al., 2006). The neural background of RS is still debated (Kovács and Schweinberger, 2016): the most widely accepted explanation is provided by predictive theories, according to which RS reflects the reduced prediction error in a Bayesian multi-stage model of cortical functions (Auksztulewicz and Friston, 2016; Friston, 2010, 2005; Rao and Ballard, 1999; Summerfield et al., 2008). Indeed, there is increasing neuroimaging evidence that higher-order contextual expectations modulate the magnitude of RS for visual (Grotheer et al., 2014; Grotheer and Kovács, 2016, 2014; Kovács et al., 2013; Larsson and Smith, 2012; Mayrhauser et al., 2014) as well as for acoustic stimuli (Andics et al., 2013a, 2013b; Todorovic et al., 2011; Todorovic and de Lange, 2012).

Based on these results, we hypothesized that cortical regions involved in the processing of speech stimuli might be the primary locus of RS effects related to the processing of word stress information (H1). Furthermore, we also assumed that word stress is encoded by predictive mechanisms. Therefore, we expected the RS effects to be modulated by the predictability of stress violations in areas related to the dorsal auditory stream, evoked by the repetition of the legal stress pattern (H2).

To investigate the above hypothesis, we followed the paradigm of Summerfield et al. (2008), previously widely used for visual and acoustic stimuli (for a review, see Grotheer and Kovács, 2016). Briefly, we embedded pseudoword pairs with repeated and alternating word stress in longer blocks where the repetitions were either likely, thereby predicted or rare, thereby surprising. We measured RS by comparing the repeated and alternating pseudoword pairs. Stimuli in the repeated pairs had the same stress pattern, i.e., stress on the first syllable, which is the only existing word stress pattern in Hungarian. In the alternating pairs, the difference between the two stimuli was the position of stress: stress

was on the first syllable for the first stimulus (legal stress pattern) while it was on the second syllable for second stimulus (illegal stress pattern). The reason for this choice was that in our previous ERP study (Honbolygó and Csépe, 2013), we found that only the illegal stress pattern elicited the Mismatch Negativity (MMN) component when it was in the deviant position. The legal stress pattern did not elicit MMN in the deviant position, arguably because it did not violate the predictions about the native stress pattern. Based on this, to address the primary question of the present study, i.e., the predictive processing of word stress, we focused on the conditions in which the illegal stress pattern in the alternating pairs violated the prediction formed on the legal stress pattern by the repeated pairs.

2. Materials and methods

2.1. Participants

Twenty-three healthy adults took part in the experiment. Three of them were excluded from further analysis: one because the overall hit rate was lower than 80% and the overall false alarm rate was higher than 10% in the behavioral task, the other two participants due to excessive head movements during MRI scanning. Therefore, 20 participants remained in the final sample (13 females; all right-handed, $M_{\text{Age}} = 28.6$ years, $SD = 6.1$ years, $M_{\text{Years of education}} = 18.4$ years, $SD = 2.1$ years). Note that the final sample size varies between 15 and 20 in the fMRI data analyses as a function of the number of the successfully identified region of interest (ROI) during the functional localizer runs (see sections fMRI data analysis and fMRI results). All participants were native speakers of Hungarian, had normal or corrected-to-normal vision and reported normal hearing levels. None of them reported a history of any neurological and/or psychiatric condition. All participants provided written informed consent before enrolment and received no compensation for taking part in the experiment. The study was approved by the Ethical Board of the Medical Research Council, Hungary and was conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

2.2.1. Experimental task

For the main experimental task, we used disyllabic pseudowords as auditory stimuli, uttered with legal (stress on the first syllable) and illegal (stress on the second syllable) stress patterns, according to the stress assignment rules of Hungarian language (Siptár and Törkenczy, 2007). Hungarian language is ideal to study the predictive mechanisms of word stress processing because it is a fixed-stress language (Siptár and Törkenczy, 2007). This means that in contrast to e.g., English, a variable stress language, the stress pattern of every disyllabic word is the same without exception, i.e., stress always falls on the first syllable. Therefore, it can be expected that Hungarian speakers are especially sensitive to any violation of this highly regular stress pattern. As it has been shown previously, Hungarian speakers detect stress pattern violations pre-attentively in both meaningful words (Garami et al., 2017; Honbolygó et al., 2004) and meaningless pseudowords (Honbolygó and Csépe, 2013).

All pseudowords had a consonant-vowel-consonant-vowel structure (e.g., /bidi/, /divi/, /sipi/, /tiki/, etc.) and we used the same vowel /i/ (pronounced “e” as in the word “me”) to ease clear pronunciation. Of all the possible permutations of Hungarian consonants and the /i/vowel, altogether 47 pseudowords were selected, excluding meaningful words and pseudowords that sounded odd. The average length of the pseudowords was 594 ms ($SD = 85$ ms) and ranged between 410 ms and 863 ms.

To avoid potential confounds of the acoustic features on the experimental effects, we created trial-unique auditory stimuli randomized across participants. To obtain trial-uniqueness, we recorded each of the 47 pseudowords with 4 different female speakers (who were native

Hungarian speech therapists and/or linguists and were trained to produce the required stress patterns). Speakers produced naturally both the legal and illegal stress patterns, and no post-processing was applied to artificially enhance the difference between the stress patterns. After checking all recorded tokens, we selected 40 legal-stressed and 40 illegal-stressed pseudowords from each speaker to be used during the experiment. These were the 40 best stimuli in terms of intelligibility and clearness of stress patterns, as judged by two of the authors (F.H., A.K.) and another colleague (B.Cs.). Next, we manipulated the acoustical parameters of the stimuli using the Praat software (Boersma and Weenink, 2007). We modified the fundamental frequency of the stimuli by shifting the overall f0 to 90%, 100%, and 110% of the original. This technique has been applied by Dupoux et al. (2001) and also by our group in a previous study (Honbolygó et al., 2019) in order to increase the acoustical variability of the stimuli. Consequently, we obtained 480 (40 stimuli * 4 speakers * 3 shifts of the f0) legal-stressed pseudowords and 480 illegal-stressed pseudowords.

We also created target stimuli by modifying the fundamental frequency of the original stimuli to 110%, 120%, and 130% (i.e., the frequency of the target stimuli was 20% higher than that of the respective original stimuli after shifting their overall f0). Targets were needed to maintain the attention of participants and were not included in the analysis. To create the target stimuli, a well-detectable perceptual difference between the stimulus pairs was needed that was different from the perceptual difference investigated (i.e., stress difference). One of the most prominent and easily detectable acoustical feature of speech stimuli is fundamental frequency, this is why we decided to manipulate this feature.

Finally, we equalized the loudness level of all stimuli using RMS (root mean square) normalization and added a rise/fall amplitude envelope to the beginning and the ending of the sound to avoid the “clicking” sound at the stimulus onset. The acoustical characteristics of the recorded stimuli are summarized in Table 1.

2.2.2. Functional localizer

For the independent functional localizer scans, we created four 15 s long speech segments, consisting of a sequence of disyllabic pseudowords. Pseudowords conformed to the phonotactical rules of Hungarian and were uttered by two male speakers. Each segment consisted of 15–16 pseudowords, randomly selected from both speakers, and there was a 200–300 ms long pause between successive pseudowords. Using the four original speech segments, we created two distorted, unintelligible segments which served as baseline conditions: signal correlated noise (SCN) and spectrally rotated speech (SRSP). Both manipulations were performed using scripts in the Praat software (the SCN script was written by Matt Davis, MRC Cognition and Brain Sciences Unit; the SRSP script was written by Holger Mitterer, University of Malta). The SCN was created by extracting the amplitude envelope of the original recordings and applying it to the randomized phase spectrum of the spectrogram of the original recording, i.e., to a pink noise having the same spectral profile as the recordings. This resulted in amplitude modulated noise-

like stimuli, which retained the temporal characteristics of speech but removed all spectral information, effectively making the stimuli unintelligible and completely dissimilar to speech. The SRSP was created by inverting the spectral content of the original recordings at 3600 Hz, i.e., spectro-temporal information of lower frequencies became high frequency information and vice-versa. This resulted in an alien-like speech: it had very similar temporal and spectral complexity to the original speech but it was unintelligible (see also Scott et al., 2000).

2.3. Design and procedure

2.3.1. Experimental task

The design and procedure of the present experiment were based on previous studies testing RS for human voices (Andics et al., 2013a) and for various visual stimuli such as faces and letters (Grotheer et al., 2014; Grotheer and Kovács, 2014; Kovács et al., 2013; Summerfield et al., 2008). The trial and block structure of the experiment are shown in Fig. 1.

Stimuli were presented pairwise with a stimulus onset asynchrony (SOA) varying randomly between 800 ms and 1000 ms. The stress pattern of the first stimulus (S1) was either identical to (Repetition Trial = RepT) or different from that of the second stimulus (S2; Alternation Trial = AltT). In the RepT, both S1 and S2 had the same legal stress pattern (stress on the first syllable). In the AltT, the only difference between S1 and S2 was the position of stress: stress was on the first syllable for S1 (legal stress pattern) while it was on the second syllable for S2 (illegal stress pattern), but the two stimuli were otherwise identical. Stimulus pairs were separated with a randomized inter-trial interval (ITI) of 4 or 6 s.

Besides the different trial types, two different types of blocks were presented to test the modulation of repetition probability: Repetition Blocks (RepBs) and Alternation Blocks (AltBs). In each block, 20% of the trials were target trials, which were either AltTs or RepTs with the same probability (i.e., 10% of the targets were AltTs and 10% were RepTs in each block, respectively). In target trials, the frequency of the S2 in the stimulus pair was 20% higher than that of the S1 (see Stimuli section). In the RepBs, including the target trials, 70% of the trials were RepTs while 30% were AltTs. In the AltBs, including the target trials, 70% of the trials were AltTs and 30% were RepTs. The first four trials of each block were always non-target trials and consisted of the more frequent trial type of that block (RepT in RepB, AltT in AltB). The order and identity of AltTs and RepTs in each block was random and unique for each participant with the constraint that target trials were separated by at least two non-target trials. Particularly, twelve randomly mixed and unique stimulus files per participant determined stimulus presentation during the experimental task: four of them coded the S1 in each block, another four files coded the trial type (RepT, AltT, Repetition target, or Alternation target), and the remaining four files coded the S2 that corresponded to the matching of the S1 and the trial type (e.g., if the given trial was an AltT, the exact same pseudoword as the S1 was presented as the S2 but with illegal stress pattern). There were 120 trials (stimulus pairs) in each

Table 1

Acoustical characteristics of the stimuli. The maximum (highest value measured within the syllable) and slope (increase and direction of the change measured within the syllable) values of the f0 and intensity averaged across all tokens and speakers are summarized.

	Stress position							
	Initial				Final			
	Syllable		Syllable		Syllable		Syllable	
	1st	2nd		1st	2nd		1st	2nd
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
f0 _{max}	243.81	21.94	149.19	40.03	202.34	28.31	210.97	36.04
f0 _{slope}	-2.16	2.78	-0.01	1.25	-0.17	1.38	-2.50	2.33
intensity _{max}	87.78	2.29	78.86	4.82	83.92	3.60	87.17	1.94
intensity _{slope}	-0.03	1.19	-0.30	0.42	0.04	0.99	-0.22	0.52

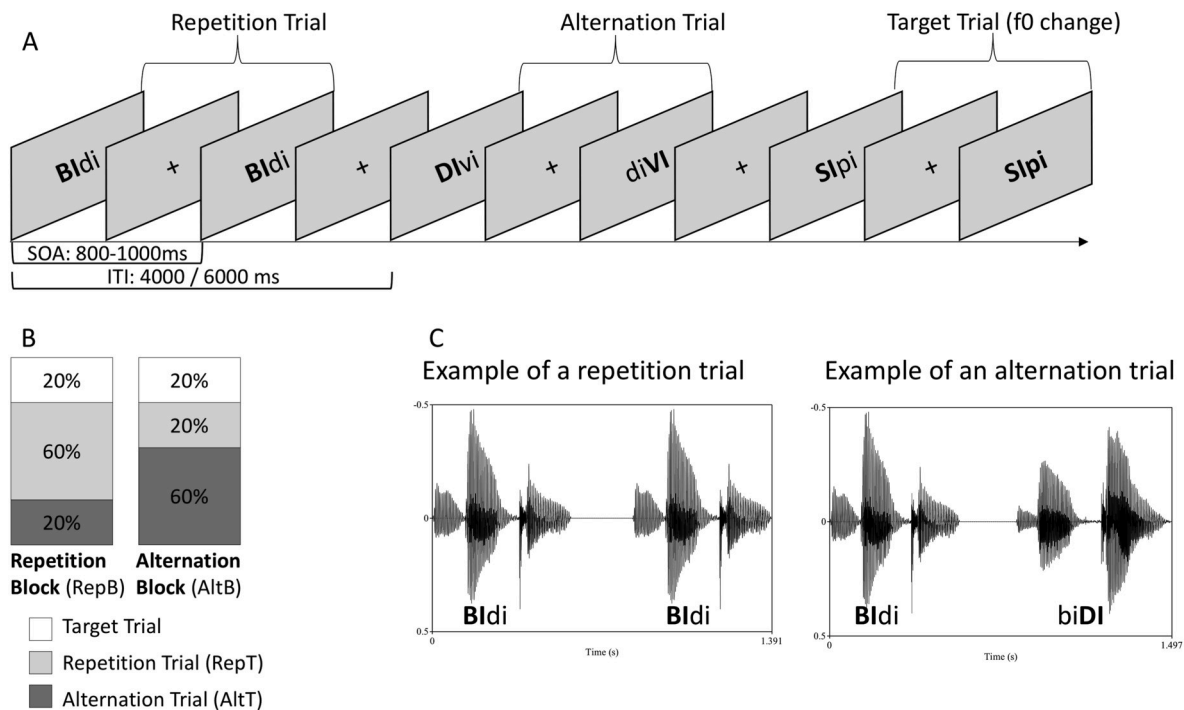


Fig. 1. Stimuli and paradigm. **A.** Schematic illustration of the trial structure. The acoustic stimulus signal that participants heard is represented by written words in the figure. Note that different pseudowords were used; and, during trial presentation, participants saw only a fixation cross on the screen. Capital letters in bold indicate stress on the syllable. Consequently, a repetition (RepT), an alternation (AltT), and a target trial are illustrated. In the target trial, bold letters signal the higher frequency of the target stimulus. **B.** The structure of the repetition (RepB) and alternation (AltB) blocks. Note that in each block, half of the target trials (10%) were AltTs and the other half (10%) were RepTs. Thus, altogether, in the RepBs, 70% of the trials were RepTs (60% non-target RepTs plus 10% target RepTs) while 30% were AltTs (20% non-target AltTs plus 10% target AltTs). Similarly, in the AltBs, 70% of the trials were AltTs and 30% were RepTs. **C.** Acoustic waveform of a typical repetition and an alternation trial.

block, and altogether four blocks (i.e., 480 stimuli) were presented during a scanning session. In order to obtain a stronger repetition probability effect, the different blocks within a particular functional run were *not* mixed (cf. Andics et al., 2013a; Grotheer and Kovács, 2014). Instead, RepBs and AltBs were presented in separate functional runs with block (run) order counterbalanced across participants following a Latin square design. The total time of one experimental run was 10 m 44 s. Between each run, breaks lasted until the next run was initiated (approx. 1 min). There was one-way communication between the experimenter and the participant during these breaks; only basic information was provided (number of the remaining runs, etc.). The task of the participants throughout these runs was irrelevant to the manipulation of the stress pattern regarding S1 and S2 stimuli, but it involved decision on the phonetic characteristics of the stimulus pair. Namely, participants had to signal the detection of the target stimulus – i.e., when the second stimulus in the pair had a *higher overall pitch* than the first one – with a button press as fast and accurate as possible using their right index finger. They were also instructed to maintain their gaze on the central fixation cross appearing on a screen throughout the experiment.

2.3.2. Functional localizer

Speech-sensitive cortical regions were defined in separate functional localizer runs, the structure of which was based on the paradigm described in Stoppelman et al. (2013). The authors used a paradigm in which blocks of continuous speech, reversed speech, and SCN were presented, and found that while SCN served as an effective baseline to contrast speech stimuli, reversed speech removed much of the speech-related responses in speech specific areas. Taking these results into account, we applied two baseline conditions: SCN and SRSP. The latter condition was selected because previous studies (Obleser et al., 2006; Scott et al., 2000) effectively used it as a baseline condition, and, in contrast to SCN, it retains much of the complex spectro-temporal

properties of speech while leaving it unintelligible.

Consequently, three conditions – Speech, SCN, SRSP – were presented in the functional localizer scan, using a block design. Each block was 15 s long followed by 12, 14, or 16 s long silent intervals. Two localizer runs were presented with 12 blocks in each (altogether 24 blocks were presented). The order of blocks was pseudorandomized such that two subsequent blocks belonged to different conditions (cf. Stoppelman et al., 2013). During the localizer runs, participants were instructed to pay attention to all auditory stimuli without any tasks. The total time of one functional localizer run was 6 m 34 s.

2.4. Stimulus presentation

Stimulus presentation was controlled via MATLAB 2013b (The MathWorks Inc., Natick, MA, USA.) using the Psychophysics Toolbox Version 3 (PTB-3) extensions (Brainard, 1997; Pelli, 1997). Auditory stimuli were delivered binaurally via MRI-compatible headphones (MR Confon, Magdeburg, Germany) at a comfortable volume (previously set based on the pilot scans and used throughout the experiment). Written instructions, feedback on performance after each block (number of hits in a given block), and a central fixation cross were displayed on an MRI-compatible LCD screen (32" NNL LCD Monitor, NordicNeuroLab, Bergen, Norway; refresh rate: 60 Hz) placed at 142 cm from the observer, and were viewed via a mirror attached to the top of the head coil.

The four experimental runs (two runs with RepBs, two runs with AltBs), the structural run, and the two runs of the functional localizer were administered in one scanning session in that order. The length of the full scanning session was around 1 h 5 m. During scanning, the presentation of S1s was synchronized to the trigger pulses of the MRI scanner. Before scanning, participants practiced the target detection task with eight trials outside the scanner. They were not informed about

the different presentation probabilities of RepTs and AltTs in the two block types. The entire experimental procedure lasted about 2 h.

2.5. Imaging parameters

A 3-T MRI scanner (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) was used with a 20-channel head-neck receiver coil. During the functional runs, we continuously acquired images (36 slices, 10° tilted relative to axial, T2*-weighted EPI sequence with twofold GRAPPA acceleration, Griswold et al., 2002), TR = 2000 ms; TE = 30 ms; flip angle = 83°; 70 × 70 matrices; in-plane resolution: 3 × 3 mm, slice thickness: 3 mm, inter-slice gap: 0.75 mm). To obtain 3D structural scans, sagittal T1-weighted images were acquired using a magnetization-prepared fast gradient echo sequence (MP-RAGE, Mugler and Brookeman, 1990; TR = 2300 ms; TE = 3.03 ms; 1 mm isotropic voxel size).

2.6. fMRI data analysis

2.6.1. Preprocessing

fMRI data preprocessing and analysis was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, UK) running under MATLAB 2013b. The functional images were realigned to spatially match the mean of the images created after a realignment to the first volume. The structural images were coregistered to the mean functional images. To spatially normalize the realigned functional images to MNI space, we applied the deformation field parameters that were obtained during the normalization of the anatomical T1-weighted image. After the normalization procedure, functional images were spatially smoothed with an 8-mm full-width at half-maximum isotropic Gaussian kernel. The same preprocessing steps were performed on the functional images of the localizer runs, as well.

2.6.2. Single-subject analysis

The separate functional localizer runs were used to determine ROIs (ROI; see Fig. 2), which were analyzed using MARSBAR 0.44 toolbox for SPM (Brett et al., 2002). Previous studies investigating RS and repetition

probability modulation effects for different visual stimuli found that only the analysis of specific ROIs was sufficiently sensitive to reliably detect these effects; and from whole-brain analyses testing the same effects, no significant activations emerged in additional brain regions (Grotheer et al., 2014; Grotheer and Kovács, 2014; Kovács et al., 2013; Summerfield et al., 2008). Therefore, in this study, we primarily focus on the analysis of specific ROIs.

We used two types of localizer contrasts (Speech > SRSP and Speech > SCN) to determine the location of the left posterior (p) STG, left middle (m) STG, and left anterior (a) STG as well as the right pSTG, right mSTG, and right aSTG. By default, the Speech > SRSP was used, with the help of which all ROIs were identified in 13 participants. Beyond this contrast, we used the Speech > SCN contrast to identify the left mSTG in one, the right mSTG in two, and the right aSTG in one participant as these ROIs could not be determined using the default Speech > SRSP contrast in their cases. Meanwhile, none of the two contrasts was used successfully in identifying the location of the left aSTG, the right mSTG, and the right aSTG in the case of two, one, and three participants, respectively. Accordingly, the left pSTG, left mSTG, and right pSTG could be defined in all participants, while the left aSTG, the right mSTG, and the right aSTG were defined in 18, 19, and 17 participants, respectively. Therefore, the number of participants in whom all ROIs were successfully identified was 15 (see Table 2).

The location of the three areas (posterior, middle, and anterior STG) in both hemispheres was determined for each participant individually as responding more strongly to speech than to SRSP or SCN stimuli in the localizer runs ($p_{\text{uncorrected}} \leq .001$). Areas closest to the corresponding reference clusters (according to the whole-brain random-effects analysis for the Speech > SRSP or SCN; $p_{\text{FWE}} < .05$, cluster extent of >100), and where the activations reached local maxima were considered as appropriate on the individual level. The individual and the average MNI coordinates for these areas are presented in Table 2.

To analyze the fMRI data of the experimental task, we extracted the mean percent signal change and a time series of the voxel values within an 8-mm radius sphere around the ROIs' centers using MARSBAR as follows. Regressors were created by modeling the four experimental conditions (AltB_AltT, AltB_RepT, RepB_AltT, and RepB_RepT) and the target trials were modeled at the onset of the S1 stimuli, using delta

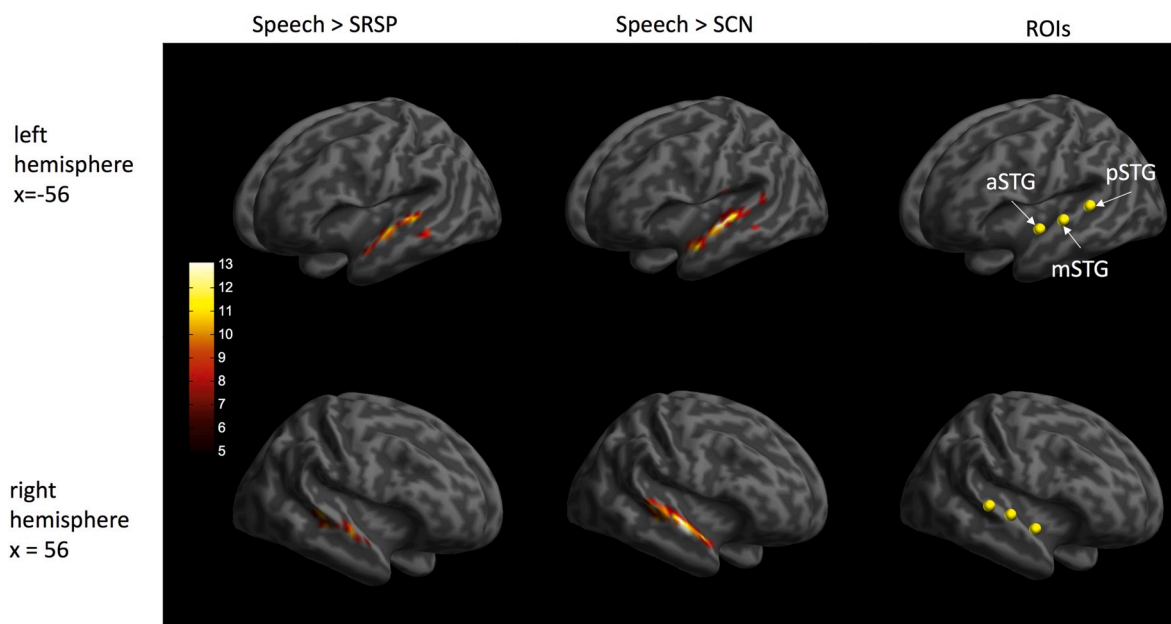


Fig. 2. Results of the whole-brain random-effects analysis of different localizer contrasts, showing the locations of the ROIs (yellow dots; see specific coordinates in Table 2). Colored areas show t-values ($p_{\text{FWE}} < .05$, with an additional cluster extent of >100 voxels). SRSP: spectrally rotated speech; SCN: signal correlated noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2

The individual ($N = 20$) and the mean (in bold) MNI coordinates of the ROIs determined based on the functional localizer.

Participant	left pSTG			left mSTG			left aSTG			right pSTG			right mSTG			right aSTG		
	x	y	z	x	y	z	x	y	z	y	x	z	x	y	z	x	y	z
1	-58	-48	8	-60	-20	2	-56	6	-14	52	-32	2	58	-10	-8	56	14	-26
2	-62	-26	4	-46	-12	4	-58	-10	-6	62	-20	2	60	-14	-6	68	-10	-18
3	-48	-40	4	-62	-26	4	-60	-16	-2	48	-38	4	54	-10	-8			
4	-66	-28	6	-56	-18	0	-62	-8	-8	44	-28	0	58	-12	-6	64	0	-16
5	-60	-44	18	-62	-20	2	-56	-4	-6	68	-34	18	60	-2	-6	50	8	-4
6	-66	-40	8	-68	-30	0	-62	-8	-6	46	-30	0	68	-22	2	56	6	-10
7	-64	-52	8	-62	-20	-4	-60	0	-12	50	-36	4	54	-28	6	58	2	-12
8	-66	-28	0	-62	-12	-2	-64	-4	-6	62	-24	2	64	-4	-8			
9	-58	-36	6	-66	-14	2	-60	-10	-2	52	-40	14	58	-18	-2			
10	-68	-42	6	-64	-10	-6	-56	0	-10	52	-28	-2	64	-12	-10	56	8	-14
11	-64	-20	6	-60	-12	0	-50	0	-8	64	-30	0	64	-10	-4	58	2	-10
12	-68	-34	8	-64	-20	-2	-64	-2	-4	46	-32	4	64	-18	2	66	-4	-4
13	-60	-46	6	-62	-24	0	-60	4	-12	66	-48	2	66	-24	-4	60	-2	-4
14	-62	-54	20	-62	-26	6	-60	-6	-2	68	-38	12	60	-18	0	58	2	-6
15	-62	-26	0	-64	-18	-6				56	-28	2	56	-20	-6	60	-8	-8
16	-64	-34	8	-60	-24	-2				64	-34	4				60	-4	-6
17	-60	-58	6	-56	-40	8	-62	-18	-2	56	-32	-2	62	-22	-2	60	0	-10
18	-68	-20	2	-56	-10	6	-58	-6	-4	64	-20	2	58	0	-8	54	10	-8
19	-52	-32	2	-60	-30	0	-64	-16	0	44	-34	2	50	-26	2	58	2	-4
20	-62	-50	4	-56	-26	-2	-54	8	-14	60	-20	-4	62	8	-10	56	14	-16
Mean	-62	-38	7	-60	-21	1	-59	-5	-7	56	-31	3	60	-14	-4	59	2	-10
SD	5	11	5	5	8	4	4	7	4	8	7	5	5	10	5	4	7	6

Note. Grey shading denotes that the given ROI could not be determined; framing denotes that the given ROI was

functions convolved with the canonical hemodynamic response function of SPM12, for the general linear model analysis of the data. Low-frequency components were excluded from the data using a high-pass filter with 128 s cut-off. Correction for temporal auto-correlations was done using an autoregressive AR(1) model and movement-related variance was accounted for by the spatial parameters resulting from the realignment procedure.

2.6.3. Multi-subject analysis

The mean percent signal change values obtained in the experimental task for all ROIs were analyzed in two steps. First, a five-way repeated measures analysis of variance (ANOVA) was conducted with Hemisphere (right, left), Region (posterior, middle, anterior), Run (1, 2), Block (Alternation, Repetition), and Trial type (Alternation, Repetition) as within-subject factors. Second, to separately analyze each ROI, three-way repeated measures ANOVAs with Run (1, 2), Block (AltB, RepB), and Trial type (AltT, RepT) as within-subject factors were conducted. In all ANOVAs, partial eta-squared (η_p^2) is reported as the measure of effect size. To control for Type I error, we used Tukey HSD tests for pair-wise comparisons.

3. Results

3.1. Behavioral performance

Participants detected the target stimuli with an average accuracy of 90.1% ($SE = 1.4\%$) and with an average RT of 843 ms ($SE = 34$ ms). Their false alarm rate was on average 4.9% ($SE = 1.1\%$). Behavioral measures (accuracy and RT) calculated for the *target stimuli* were entered into two-way repeated measures ANOVAs with Block and Trial type as within-subject factors. The Block * Trial type ANOVA performed on *accuracy data* revealed a significant main effect of Trial type, $F(1, 19) = 14.53, p = .001, \eta_p^2 = .433$, showing that AltTs were judged less

accurately than RepTs (correct %: 87.2% vs. 92.9%). The same ANOVA performed on *RT data* showed that responses were slower for AltTs than for RepTs (main effect of Trial type, $F(1, 19) = 23.59, p < .001, \eta_p^2 = .554$; 893 ms vs. 795 ms). No other significant effects were observed in these analyses.

3.2. fMRI results

In the analyses below, RS is indicated by the main effect of Trial type (overall difference between AltTs and RepTs), while the repetition probability modulation of RS is indicated by the Block * Trial type interaction (Summerfield et al., 2008). Since we report only the significant main effects and interactions in the following sections, results from ANOVAs performed on the mean percent signal change in each ROI are detailed in Table 3.

3.2.1. Overall ANOVA

In order to examine whether RS and repetition probability effects differ between ROIs and hemispheres, an overall five-way repeated measures ANOVA was performed. The main effect of Trial type was significant, $F(1, 14) = 39.00, p < .001, \eta_p^2 = .736$, which was qualified by the significant Region * Trial type interaction, $F(2, 28) = 3.85, p = .033, \eta_p^2 = .215$, indicating that although the RS effect (RepT < AltT) was significant in all ROIs (all $ps \leq .001$), it decreased from posterior to the anterior regions (pSTG difference: 0.15% signal change; mSTG difference: 0.10%; aSTG difference: 0.09%; pSTG difference > aSTG difference, $p = .032$). We found an overall repetition probability effect, irrespective of hemisphere, specific brain region, or functional run (significant Block * Trial type interaction, $F(1, 14) = 5.93, p = .029, \eta_p^2 = .298$), as well as the RepT < AltT difference was significantly larger in the RepBs than in the AltBs (0.15%, $p < .001$ vs. 0.08%, $p = .012$). The Region * Run * Trial type interaction was also significant, $F(2, 28) = 3.88, p = .033, \eta_p^2 = .217$, because of the variation of the RS effect

Table 3
Summary of results from ANOVAs performed on the mean percent signal change in each ROI.

Region	Run		Block		Trial type		Run * Block		Run * Trial type		Block * Trial type		Run * Block * Trial type	
	F	p	F	p	F	p	F	p	F	p	F	p	F	p
left pSTG	2.91	.104	0.79	.386	34.68	< .001	1.99	.175	0.18	.672	6.80	.017	2.14	.160
right pSTG	<0.01	.957	2.49	.131	45.35	< .001	0.50	.487	1.08	.311	4.07	<i>.058</i>	0.12	.738
left mSTG	0.17	.683	4.28	<i>.053</i>	18.38	< .001	0.16	.698	1.14	.300	3.50	<i>.077</i>	0.01	.927
right mSTG	1.53	.232	1.11	.306	37.44	< .001	0.88	.361	0.60	.450	7.92	.011	0.30	.589
left aSTG	0.25	.623	0.69	.418	24.02	< .001	0.21	.652	4.56	.048	1.66	.215	1.76	.202
right aSTG	1.09	.312	1.97	.179	4.27	<i>.055</i>	1.02	.328	0.35	.563	0.52	<i>.479</i>	0.13	.719

Note. *p*-values below .05 are **boldfaced**, and below 0.10 are *italicized*.

between Run 1 and Run 2 only in the anterior regions (see below the control effects in relation to the aSTG), but no other significant main effects or interactions were found involving Hemisphere, Region, or Run as a factor.

3.2.2. Repetition suppression (H1) in each ROI

We observed a significant RS in the left (main effect of Trial type, $F(1, 19) = 34.68, p < .001, \eta_p^2 = .646$) and right pSTG (main effect of Trial type, $F(1, 19) = 45.35, p < .001, \eta_p^2 = .704$), such that the BOLD signal was significantly lower for RepTs than for AltTs (see Fig. 3a). Similarly, significant RS was observed in the left (main effect of Trial type, $F(1, 19) = 18.38, p < .001, \eta_p^2 = .492$) as well as in the right mSTG (main effect of Trial type, $F(1, 18) = 37.44, p < .001, \eta_p^2 = .675$; see Fig. 3b). Although we found a significant RS in the left aSTG (main effect of Trial type, $F(1, 17) = 24.02, p < .001, \eta_p^2 = .586$), this was only a tendency in the right aSTG (main effect of Trial type, $F(1, 16) = 4.27, p = .055, \eta_p^2 = .211$; see Fig. 3c). In sum, all ROIs showed RS effect, although this was weaker in the right aSTG.

3.2.3. Repetition probability modulations of RS (H2) in each ROI

In the left pSTG, a significant Block * Trial type interaction was observed, $F(1, 19) = 6.80, p = .017, \eta_p^2 = .264$. The difference in the BOLD signal between RepTs and AltTs was significantly larger in the RepBs (Fig. 3a, difference: 0.18%, $p < .001$) than in the AltBs (Fig. 3a, difference: 0.11%, $p < .001$). In the right pSTG, the Block * Trial type interaction showed a strong tendency, $F(1, 19) = 4.07, p = .058, \eta_p^2 = .176$, and pair-wise comparisons showed larger differences between the trial types in the RepBs (Fig. 3a, difference: 0.16%, $p < .001$) than in the AltBs (Fig. 3a, difference: 0.10%, $p = .001$).

In the left mSTG, the Block * Trial type interaction, again, showed a tendency, $F(1, 19) = 3.50, p = .077, \eta_p^2 = .156$, as the BOLD signal was significantly lower for RepTs than for AltTs in the RepBs (Fig. 3b, difference: 0.12%, $p < .001$), but only tended to differ in the AltBs (Fig. 3b, difference: 0.06%, $p = .055$). In the right mSTG, the Block * Trial type significant interaction indicated a clear repetition probability effect, $F(1, 18) = 7.92, p = .011, \eta_p^2 = .305$. The difference between RepTs and AltTs was, again, significantly larger in the RepBs (Fig. 3b, difference: 0.17%, $p < .001$) than in the AltBs (Fig. 3b, difference: 0.08%, $p = .010$).

In contrast to the former ROIs, in the aSTG, the Block * Trial type interaction was not significant either in the left, $F(1, 17) = 1.66, p = .215, \eta_p^2 = .089$, or in the right, $F(1, 16) = 0.52, p = .479, \eta_p^2 = .032$, hemisphere (see Fig. 3c). Particularly, the magnitude of the BOLD response for RepTs did not differ between the AltBs and RepBs in the left ($p = .076$) and right aSTG ($p = .061$).

In sum, we found significant repetition probability modulations in the pSTG and mSTG bilaterally, and a lack of this effect in the bilateral aSTG. Although the interaction effect was only a tendency in the left mSTG and right pSTG, pair-wise tests suggested repetition probability modulations in these ROIs. These tendencies were also confirmed by the overall ANOVA.

3.2.4. Control effects

Other main effects and interactions including Run, Block, Run *

Block, Run * Trial type, and Run * Block * Trial type were not significant in any of the ROIs, except for a significant Run * Trial type interaction in the left aSTG, $F(1, 17) = 4.56, p = .048, \eta_p^2 = .211$ (see Table 3). Here, the RepTs < AltTs difference was larger in Run 1 (0.13%, $p < .001$) than in Run 2 (0.08%, $p < .001$).

3.2.5. Whole-brain analyses

To test whether other areas beyond those identified in the ROI analysis reflected the repetition probability modulation effect, we performed whole-brain random-effects analysis. We tested the main effect of Block (AltB > RepB) and the main effect of Trial type (AltT > RepT). These analyses, however, did not yield any significant activations in any brain regions at the threshold of $p_{FWE} < .05$. We also checked the potential activations at a more liberal threshold ($p_{uncorrected} < .0001$, cluster extent of >20) and no activations were found either. No significant activations were found for the opposite contrasts either (RepB > AltB, RepT > AltT).

The contrast testing the Block * Trial type interaction [(AltT_AltB vs. RepT_AltB) vs. (AltT_RepB vs. RepT_RepB)] did not yield significant activations in any brain regions even at the liberal threshold of $p_{uncorrected} < .0001$ (cluster extent of >20) either.

4. Discussion

In the present study, we used RS to investigate the processing of word stress in speech-sensitive regions of the superior temporal cortex identified with independent functional localizer. The results revealed RS effects related to word stress processing in several superior temporal cortical areas. In particular, RS was found for word stress in the bilateral pSTG and mSTG, as well as in the aSTG of the left hemisphere. The comparison of the magnitude of RS in the different ROIs revealed that its size decreased along the posterior-anterior axis, suggesting that the pSTG and mSTG regions were more actively involved in word stress encoding than more anterior regions. In addition, it was shown that fatigue effects and possible changes in the magnetic field did not influence remarkably the experimental results. Crucially, the results also revealed that the RS effect was modulated by the repetition probability in the speech specific pSTG and mSTG regions and thus provide evidence for the predictive processing of word stress in the human cortex.

In order to select the specific regions sensitive to word stress processing, we developed a functional speech localizer. According to our results, several regions of the STG were activated bilaterally by meaningless speech stimuli compared with both SCN and SRSP. Previous studies using the SRSP stimuli as contrast in localizing speech specific brain areas found activity in the STG/STS regions (Bautista and Wilson, 2016; Golden et al., 2015; Halai et al., 2015; Sabri et al., 2008) and also in IFG regions (Halai et al., 2015; Sabri et al., 2008). The lack of IFG activity in our case could be due to using pseudowords: it has been previously found that IFG regions are involved in the processing of semantically and syntactically more complex stimuli (Halai et al., 2015; Newman et al., 2003; Thompson-Schill et al., 1997). These results confirm the effectiveness of the SRSP stimuli as baseline in identifying speech specific regions.

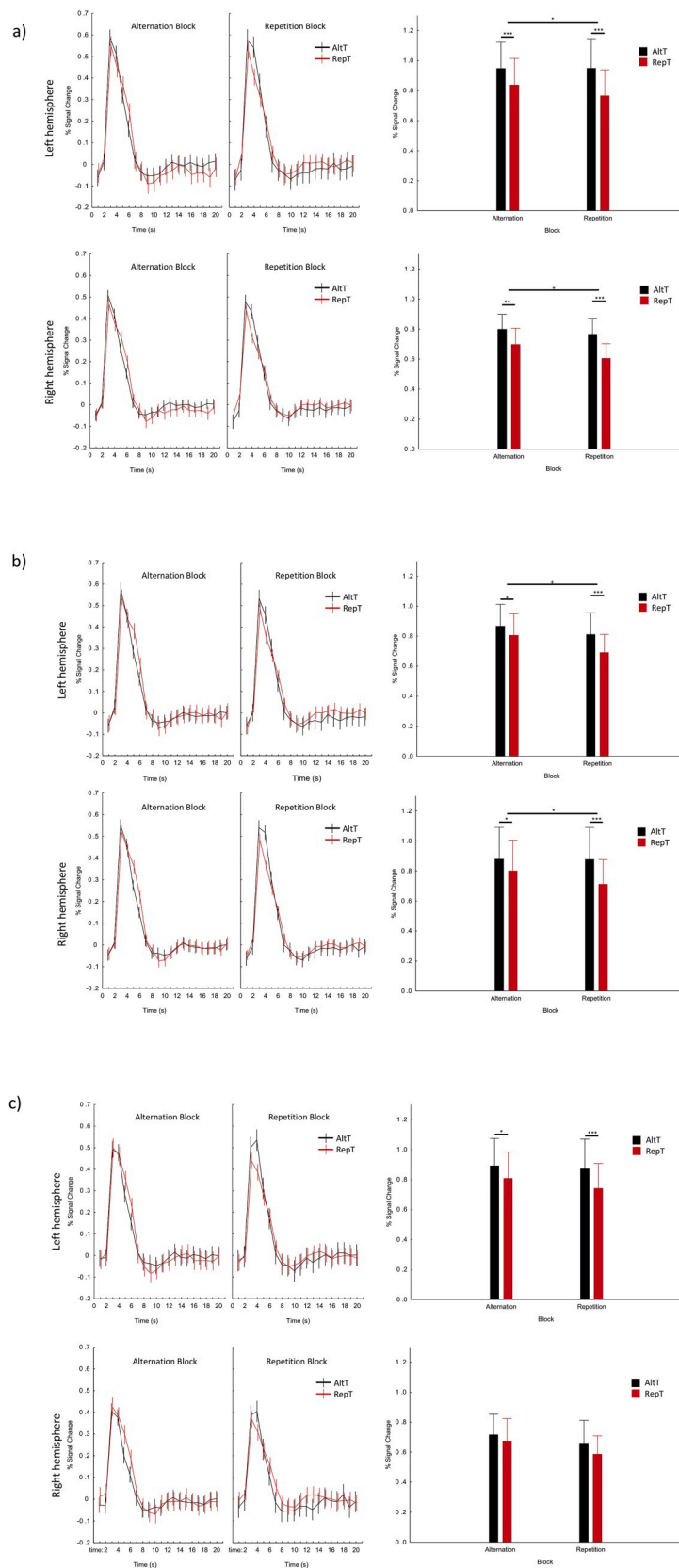


Fig. 3. Time course (mean \pm SE) and average peak activation profiles (\pm SE) in the pSTG (a), mSTG (b) and aSTG (c), separately for the different trial and block types. Short horizontal lines denote the significance of RS effects, long horizontal lines denote the significance of repetition probability effects. Note: $+p < .10$; $*p < .05$; $**p < .01$; $***p < .001$.

In previous studies investigating the nature of language comprehension, evidence for predictive mechanisms has been found at several levels of the linguistic hierarchy: prediction of upcoming words has been shown for the phonological (DeLong et al., 2005), morphosyntactic (Van Berkum et al., 2005; Wicha et al., 2004, 2003), lexical-semantic/discourse (Federmeier and Kutas, 1999; Hasson et al., 2006; Lau et al., 2016; Orfanidou et al., 2006; Otten and Van Berkum, 2008; Poppenk et al., 2016; Van Petten et al., 1999), and syntactic contexts (Arai and Keller, 2013; Bornkessel-Schlesewsky and Schlesewsky, 2013; Kuperberg and Jaeger, 2016; Kutas et al., 2011; Matchin et al., 2016; Rohde et al., 2011; Weber et al., 2016). The importance of predictive mechanisms in word stress processing is that the representation of word stress is suggested to involve hierarchical rules, i.e., rules about the assignment of stress to certain syllables at the word level or words at the sentence level (Domahs et al., 2014; Hayes, 1995; Liberman and Prince, 1977). Even in languages where stress is specified in the mental lexicon, some rule-based mechanisms exist to compute the stress pattern of unknown words (Colombo, 1992; Cutler and Isard, 1980).

One important aspect of the present study was to investigate word stress processing in Hungarian, a fixed-stress language having stress on the first syllable of words. We argued that Hungarian speakers might be especially sensitive to changes of the highly regular stress pattern, because any other stress patterns could be considered as illegal (Garami et al., 2017; Honbolygó and Csépe, 2013). This raises the question if the results obtained are language specific and valid only for fixed-stress languages. As we argue above, predictions about word stress are based on phonological rules, particularly in the case of unknown words that do not have their stress pattern specified in the mental lexicon. Therefore, it can be expected that listeners of variable stress languages would show similar results to Hungarian listeners, involving similar brain regions when processing pseudowords. We could expect a crucial difference between languages, however, when processing known words, as the lexical specification of stress is probably different for fixed-stress and variable stress languages. This is an especially interesting question for further brain imaging studies, because it is unclear if the stress pattern of words is specified or not in the mental lexicon of listeners of fixed-stress languages.

Concerning the neurobiological background of words stress processing, the superior temporal lobe (STG/STS regions) has been previously found to be active (Aleman et al., 2005; Domahs et al., 2013; Heisterueber et al., 2014; Kandylaki et al., 2017; Klein et al., 2011). In these previous studies, several other brain regions have been found active, including the IFG, SMA, areas in the parietal lobe (angular gyrus, superior parietal gyrus, parietal lobule), and frontal lobe (precentral, postcentral, and middle frontal gyrus). The possible reason of the activation of diverse brain areas is that these studies used various paradigms (discrimination, imagery, recall tasks, well-formedness judgement), which might have tapped on different cognitive functions and consequently involved differing brain areas. In our study, the RS paradigm as an implicit and passive task allowed us to investigate areas directly and specifically related to word stress processing.

Moreover, the paradigm also allowed us to demonstrate the importance of the STG region in the processing of word stress based on long-term memory traces (reflected in the observed repetition probability effect). In our previous ERP study (Honbolygó and Csépe, 2013), we suggested that word stress processing is based on so-called stress templates, which were assumed to be pre-lexical, speech specific long-term traces of word stress patterns. In that study, we found that the MMN component appeared only when the illegal stress pattern (stress on the second syllable) was the deviant and the legal stress pattern (stress on the first syllable) was the standard. However, there was no MMN in the reversed condition when the legal stress pattern was the deviant and the illegal stress pattern was the standard, indicating that the legal stress pattern in the deviant position did not violate the predictions based on the long-term traces of the native stress pattern. Here, we show that the possible neural background of the word stress processing based on these

suggested stress templates involves the bilateral pSTG and mSTG regions. Although there have been some claims about a possible right hemispheric dominance in prosodic processing (Gandour et al., 2004; Meyer et al., 2004), studies about word stress showed either left hemispheric dominance (Aleman et al., 2005) or bilateral activations (Domahs et al., 2013; Heisterueber et al., 2014; Kandylaki et al., 2017; Klein et al., 2011). Our study further supports the bilateral nature of word stress processing.

The results also fit with the assumptions of the neurobiological language model proposed by Bornkessel-Schlesewsky and Schlesewsky (2013). The model suggests that speech is processed along a dual auditory stream network, consisting of an antero-ventral stream responsible for the recognition of linguistic elements and a postero-dorsal stream responsible for the predictive sequential processing of linguistic information. The model assumes that the postero-dorsal stream, among other tasks, engages in prosodic segmentation, which includes the detection of word stress. Word stress processing is a time-sensitive mechanism, and as discussed above, its representation is based on hierarchical rules, which allow the formation of expectations about the upcoming stress information (whether a syllable will be stressed or unstressed). Given that repetition suppression is closely connected to predictive processes, the RS effects found confirm that word stress information is processed based on predictive processes. Furthermore, we found some indication that the RS repetition probability effect was present to a different extent in the various regions of the STG: although the overall ANOVA did not show a significant Block * Trial type * Region interaction, the individual analysis of ROIs revealed that the RS effect was modulated by the repetition probability in the pSTG and mSTG but not in the aSTG. This might indicate that the RS probability effects are related to the posterior part of STG, assumed to belong to the dorsal stream. Nevertheless, further studies are required to uncover if word stress processing is indeed specifically associated with the dorsal stream.

Contrary to previous studies investigating the neural background of word stress processing, we did not find any IFG activation in our whole-brain analysis. As mentioned above, the lack of IFG activity in our case could be due to using pseudowords. Bornkessel-Schlesewsky and Schlesewsky (2013) suggest that the IFG is involved mostly in cognitive control and conflict resolution, and it brings together the representations generated by the two auditory streams. Since pairs of meaningless pseudowords were presented in our case, it can be assumed that their processing did not require the involvement of the IFG. Furthermore, the activation of IFG was found in studies investigating the role of prosody in sentence processing and prosodic structure building (Sammler et al., 2018; van der Burgh et al., 2019), or when participants had to make same-different judgements about stress pairs (Klein et al., 2011); that is, when prosodic information had to be explicitly used. This was not the case in our study, which might have also contributed to the lack of IFG activation.

In summary, the present study provides further evidence that, among other linguistic features, the processing of word stress is also based on predictive mechanisms, as shown by the RS and repetition probability effects found in the posterior and middle parts of the STG bilaterally. Further studies are needed to clarify if these predictive representations are similar in meaningful and meaningless words, and if they differ between languages having different stress systems. Moreover, we need more data on the role of the dorsal auditory stream and predictive processes in prosody perception at both the word and sentence level.

Declaration of competing interest

The authors claim no conflict of interest.

CRediT authorship contribution statement

Ferenc Honbolygó: Conceptualization, Methodology, Software,

Investigation, Writing - original draft. **Andrea Kóbor**: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft. **Petra Hermann**: Software, Investigation. **Ádám Ottó Kettinger**: Software, Investigation. **Zoltán Vidnyánszky**: Conceptualization, Writing - review & editing. **Gyula Kovács**: Conceptualization, Methodology, Formal analysis, Writing - review & editing, Supervision. **Valéria Csépe**: Conceptualization, Writing - review & editing, Supervision, Funding acquisition.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors claim no conflict of interest.

Acknowledgements

This study was supported by the Hungarian Scientific Research Fund, Hungary (project number: OTKA NK 101087, OTKA K 119365, PI: V.Cs.; OTKA FK 124412, PI: A.K.), the Postdoctoral Fellowship of the Hungarian Academy of Sciences, Hungary (to A.K.), the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, Hungary (BO/00273/17 - F.H., BO/00075/19 - A.K.) and by the Deutsche Forschungsgemeinschaft, Germany (KO-3918/2-2; 5-1; PI: Gy.K.). P.H., Á.O.K., and Z.V. was supported by a grant from the Hungarian Brain Research Program, Hungary (KTIA 13_NAP-A-I/18). We are grateful to Barbara Cseri for her help in the preparation of stimuli and in the collection of data. We thank Dorottya Gyarmathy and Mária Gósy for their help in creating and recording the experimental stimuli, András Beke for the acoustical measurements of the stimuli, István Winkler and Botond Hajdú for providing the pseudoword lists for the functional localizer, and Attila Andics for his help in creating the experimental scripts.

References

- Aleman, A.A., Formisano, E., Koppenhagen, H., Hagoort, P., De Haan, E.H.F., Kahn, R.S., 2005. The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebr. Cortex* 15, 221–228. <https://doi.org/10.1093/cercor/bbh124>.
- Andics, A., Gál, V., Vicsi, K., Rudas, G., Vidnyánszky, Z., 2013a. fMRI repetition suppression for voices is modulated by stimulus expectations. *Neuroimage* 69, 277–283. <https://doi.org/10.1016/j.neuroimage.2012.12.033>.
- Andics, A., McQueen, J.M., Petersson, K.M., 2013b. Mean-based neural coding of voices. *Neuroimage* 79, 351–360. <https://doi.org/10.1016/j.neuroimage.2013.05.002>.
- Arai, M., Keller, F., 2013. The use of verb-specific information for prediction in sentence processing. *Lang. Cognit. Process.* 28, 1–36. <https://doi.org/10.1080/01690965.2012.658072>.
- Auksztulewicz, R., Friston, K., 2016. Repetition suppression and its contextual determinants in predictive coding. *Cortex* 80, 125–140. <https://doi.org/10.1016/j.cortex.2015.11.024>.
- Bautista, A., Wilson, S.M., 2016. Neural responses to grammatically and lexically degraded speech. *Lang. Cogn. Neurosci.* 31, 567–574. <https://doi.org/10.1080/23273798.2015.1123281>.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2013. Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang.* 125, 60–76. <https://doi.org/10.1016/j.bandl.2013.01.010>.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S.L., Rauschecker, J.P., 2015. Neurobiological roots of language in primate audition: common computational properties. *Trends Cognit. Sci.* 19, 142–150. <https://doi.org/10.1016/j.tics.2014.12.008>.
- Brainard, D.H., 1997. The Psychophysics toolbox. *Spatial Vis.* 10, 433–436. <https://doi.org/10.1163/156856897X00357>.
- Brett, M., Anton, J.L., Valabregue, R., Poline, J.B., 2002. Region of interest analysis using an SPM toolbox. *Neuroimage* 16, 497. [https://doi.org/10.1016/S1053-8119\(02\)90010-8](https://doi.org/10.1016/S1053-8119(02)90010-8).
- Boersma, P., and Weenink, D. (2007). Praat: Doing Phonetics by Computer (Version 4.5.) [Computer program]. Retrieved from <http://www.praat.org/>.
- Colombo, L., 1992. Lexical stress effect and its interaction with frequency in word pronunciation. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 987–1003.
- Cutler, A., Dahan, D., Van Donselaar, W., 1997. Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.
- Cutler, A., Isard, S.D., 1980. The production of prosody. In: Butterworth, B. (Ed.), *Language Production*. Academic Press, London, pp. 245–269.
- Cutler, A., Norris, D., 1988. The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 113–121. <https://doi.org/10.1037/0096-1523.14.1.113>.
- DeLong, K. a, Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. <https://doi.org/10.1038/nn1504>.
- Domahs, U., Klein, E., Huber, W., Domahs, F., 2013. Good, bad and ugly word stress - fMRI evidence for foot structure driven processing of prosodic violations. *Brain Lang.* 125, 272–282. <https://doi.org/10.1016/j.bandl.2013.02.012>.
- Domahs, U., Plag, I., Carroll, R., 2014. Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *J. Comp. German Ling.* 17, 59–96. <https://doi.org/10.1007/s10828-014-9063-9>.
- Donhauser, P.W., Baillet, S., 2019. Two distinct neural timescales. *Neuron* 1–9. <https://doi.org/10.1016/j.neuron.2019.10.019>.
- Dupoux, E., Peperkamp, S., Sebastián-Gallés, N., 2001. A robust method to study stress “deafness”. *J. Acoust. Soc. Am.* 110, 1606–1618. <https://doi.org/10.1121/1.1380437>.
- Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. <https://doi.org/10.1006/jmla.1999.2660>.
- Friederici, A.D., 2011. The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>.
- Friederici, A.D., von Cramon, D.Y., Kotz, S.A., 2007. Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron* 53, 135–145.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. <https://doi.org/10.1038/nrn2787>.
- Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. <https://doi.org/10.1098/rstb.2005.1622>.
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357.
- Garami, L., Ragó, A., Honbolygó, F., Csépe, V., 2017. Lexical influence on stress processing in a fixed-stress language. *Int. J. Psychophysiol.* 117, 10–16. <https://doi.org/10.1016/j.ijpsycho.2017.03.006>.
- Geiser, E., Zaehele, T., Jancke, L., Meyer, M., 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. *J. Cognit. Neurosci.* 20, 541–552. <https://doi.org/10.1162/jocn.2008.20029>.
- Golden, H.L., Agustus, J.L., Goll, J.C., Downey, L.E., Mummery, C.J., Schott, J.M., Crutch, S.J., Warren, J.D., 2015. Functional neuroanatomy of auditory scene analysis in Alzheimer's disease. *NeuroImage Clin* 7, 699–708. <https://doi.org/10.1016/j.nicl.2015.02.019>.
- Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cognit. Sci.* 10, 14–23. <https://doi.org/10.1016/j.tics.2005.11.006>.
- Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* 47, 1202–1210. <https://doi.org/10.1002/mrm.10171>.
- Grotheer, M., Hermann, P., Vidnyánszky, Z., Kovács, G., 2014. Repetition probability effects for inverted faces. *Neuroimage* 102, 416–423. <https://doi.org/10.1016/j.neuroimage.2014.08.006>.
- Grotheer, M., Kovács, G., 2016. Can predictive coding explain repetition suppression? *Cortex* 80, 113–124. <https://doi.org/10.1016/j.cortex.2015.11.027>.
- Grotheer, M., Kovács, G., 2014. Repetition probability effects depend on prior experiences. *J. Neurosci.* 34, 6640–6646. <https://doi.org/10.1523/JNEUROSCI.5326-13.2014>.
- Halai, A.D., Parkes, L.M., Welbourne, S.R., 2015. Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *Neuroimage* 122, 214–221. <https://doi.org/10.1016/j.neuroimage.2015.05.067>.
- Hasson, U., Nusbaum, H.C., Small, S.L., 2006. Repetition suppression for spoken sentences and the effect of task demands. *J. Cognit. Neurosci.* 18, 2013–2029. <https://doi.org/10.1162/jocn.2006.18.12.2013>.
- Hayes, B., 1995. *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press, Chicago.
- Heisterueber, M., Klein, E., Willmes, K., Heim, S., Domahs, F., 2014. Processing word prosody-behavioral and neuroimaging evidence for heterogeneous performance in a language with variable stress. *Front. Psychol.* 5, 1–14. <https://doi.org/10.3389/fpsyg.2014.00365>.
- Henson, R.N.A., Rugg, M.D., 2003. Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* 41, 263–270. [https://doi.org/10.1016/S0028-3932\(02\)00159-8](https://doi.org/10.1016/S0028-3932(02)00159-8).
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Honbolygó, F., Csépe, V., 2013. Saliency or template? ERP evidence for long-term representation of word stress. *Int. J. Psychophysiol.* 87, 165–172. <https://doi.org/10.1016/j.ijpsycho.2012.12.005>.
- Honbolygó, F., Csépe, V., Ragó, A., 2004. Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neurosci. Lett.* 363, 84–88. <https://doi.org/10.1016/j.neulet.2004.03.057>.
- Honbolygó, F., Kóbor, A., Csépe, V., 2019. Cognitive components of foreign word stress processing difficulty in speakers of a native language with non-contrastive stress. *Int. J. Biling.* 23, 366–380. <https://doi.org/10.1177/1367006917728393>.

- Inspector, M., Manor, D., Amir, N., Kushnir, T., Karni, A., 2013. A word by any other intonation: fMRI evidence for implicit memory traces for pitch contours of spoken words in adult brains. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0082042>.
- Ischebeck, A.K., Friederici, A.D., Alter, K., 2008. Processing prosodic boundaries in natural and hummed speech: an fMRI study. *Cerebr. Cortex* 18, 541–552. <https://doi.org/10.1093/cercor/bhm083>.
- Kakouros, S., Räsänen, O., 2016. Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. *Cognit. Sci.* 40, 1739–1774. <https://doi.org/10.1111/cogs.12306>.
- Kakouros, S., Salminen, N., Räsänen, O., 2018. Making predictable unpredictable with style – behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia* 109, 181–199. <https://doi.org/10.1016/j.neuropsychologia.2017.12.011>.
- Kandylaki, K.D., Henrich, K., Nagels, A., Kircher, T., Domahs, U., Schlesewsky, M., Bornkessel-Schlesewsky, I., Wiese, R., 2017. Where is the beat? The neural correlates of lexical stress and rhythmic well-formedness in auditory story comprehension. *J. Cognit. Neurosci.* 29, 1119–1131. <https://doi.org/10.1162/jocn.a.01122>.
- Kandylaki, K.D., Nagels, A., Tune, S., Kircher, X.T., Wiese, X.R., Schlesewsky, M., Bornkessel-Schlesewsky, I., 2016. Predicting "when" in discourse engages the human dorsal auditory stream: an fMRI study using naturalistic stories. *J. Neurosci.* 36, 12180–12191. <https://doi.org/10.1523/JNEUROSCI.4100-15.2016>.
- Klein, E., Domahs, U., Grande, M., Domahs, F., 2011. Neuro-cognitive foundations of word stress processing - evidence from fMRI. *Behav. Brain Funct.* 7, 15. <https://doi.org/10.1186/1744-9081-7-15>.
- Kovács, G., Kaiser, D., Kaliukhovich, D. a., Vidnyánszky, Z., Vogels, R., 2013. Repetition probability does not affect fMRI repetition suppression for objects. *J. Neurosci.* 33, 9805–9812. <https://doi.org/10.1523/JNEUROSCI.3423-12.2013>.
- Kovács, G., Schweinberger, S.R., 2016. Repetition suppression – an integrative view. *Cortex* 80, 1–4. <https://doi.org/10.1016/j.cortex.2016.04.022>.
- Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? *Lang. Cognit. Neurosci.* 31, 32–59. <https://doi.org/10.1080/23273798.2015.1102299>.
- Kutas, M., DeLong, K.A., Smith, N.J., 2011. A look around at what lies ahead: prediction and predictability in language processing. In: Bar, M. (Ed.), *Predictions in the Brain: Using Our Past to Generate a Future*, pp. 190–207. <https://doi.org/10.1093/acprof:oso/9780195395518.003.0065>.
- Larsson, J., Smith, A.T., 2012. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebr. Cortex* 22, 567–576. <https://doi.org/10.1093/cercor/bhr119>.
- Lau, E.F., Weber, K., Gramfort, A., Hämäläinen, M.S., Kuperberg, G.R., 2016. Spatiotemporal signatures of lexical-semantic prediction. *Cerebr. Cortex* 26, 1377–1387. <https://doi.org/10.1093/cercor/bhu219>.
- Liberman, M., Prince, A., 1977. On stress and linguistic rhythm. *Ling. Inq.* 8, 249–336.
- Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., Besson, M., 2005. On-line processing of "pop-out" words in spoken French dialogues. *J. Cognit. Neurosci.* 17, 740–756. <https://doi.org/10.1162/0898929053747667>.
- Männel, C., Friederici, A.D., 2011. Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Dev. Sci.* 14, 786–798. <https://doi.org/10.1111/j.1467-7687.2010.01025.x>.
- Matchin, W., Hammerly, C., Lau, E., 2016. The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. *Cortex* 88, 106–123. <https://doi.org/10.1016/j.cortex.2016.12.010>.
- Mattys, S.L., White, L., Melhorn, J.F., 2005. Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134, 477–500.
- Mayrhauser, L., Bergmann, J., Crone, J., Kronbichler, M., 2014. Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Front. Hum. Neurosci.* 8, 225. <https://doi.org/10.3389/fnhum.2014.00225>.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., Von Cramon, D.Y., 2004. Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang.* 89, 277–289. [https://doi.org/10.1016/S0093-934X\(03\)00350-X](https://doi.org/10.1016/S0093-934X(03)00350-X).
- Mugler, J.P., Brookeman, J.R., 1990. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn. Reson. Med.* 15, 152–157. <https://doi.org/10.1002/mrm.1910150117>.
- 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. *Cognit. Brain Res.* 16, 297–307. [https://doi.org/10.1016/S0926-6410\(02\)00285-9](https://doi.org/10.1016/S0926-6410(02)00285-9).
- Norris, D., McQueen, J.M., Cutler, A., 1995. Competition and segmentation in spoken-word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1209–1228.
- Obleser, J., Scott, S.K., Eulitz, C., 2006. Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cerebr. Cortex* 16, 1069–1076. <https://doi.org/10.1093/cercor/bhj047>.
- Orfanidou, E., Marslen-Wilson, W.D., Davis, M.H., 2006. Neural response suppression predicts repetition priming of spoken words and pseudowords. *J. Cognit. Neurosci.* 18, 1237–1252. <https://doi.org/10.1162/jocn.2006.18.8.1237>.
- Otten, M., Van Berkum, J.J.a., 2008. Discourse-based word anticipation during language processing: prediction or priming? *Discourse Process* 45, 464–496. <https://doi.org/10.1080/01638530802356463>.
- Pelli, D.G., 1997. The VideoToolbox Software for Visual Psychophysics: Transforming Numbers into Movies. *Spatial Vis.* <https://doi.org/10.1163/156856897X00366>.
- Poppenk, J., Moscovitch, M., McIntosh, A.R., 2016. fMRI evidence of equivalent neural suppression by repetition and prior knowledge. *Neuropsychologia* 90, 159–169. <https://doi.org/10.1016/j.neuropsychologia.2016.06.034>.
- Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. <https://doi.org/10.1038/4580>.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* <https://doi.org/10.1038/nn.2331>.
- Rohde, H., Levy, R., Kehler, A., 2011. Anticipating explanations in relative clause processing. *Cognition* 118, 339–358. <https://doi.org/10.1016/j.cognition.2010.10.016>.
- Sabri, M., Binder, J.R., Desai, R., Medler, D.A., Leitl, M.D., Liebenthal, E., 2008. Attentional and linguistic interactions in speech perception. *Neuroimage* 39, 1444–1456. <https://doi.org/10.1038/jid.2014.371>.
- Sammler, D., Cunitz, K., Gierhan, S.M.E., Anwender, A., Adermann, J., Meixensberger, J., Friederici, A.D., 2018. White matter pathways for prosodic structure building: a case study. *Brain Lang.* 183, 1–10. <https://doi.org/10.1016/j.bandl.2018.05.001>.
- Sammler, D., Grosbras, M.H., Anwender, A., Bestelmeyer, P.E.G., Belin, P., 2015. Dorsal and ventral pathways for prosody. *Curr. Biol.* 25, 3079–3085. <https://doi.org/10.1016/j.cub.2015.10.009>.
- Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vrya, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U. S. A.* <https://doi.org/10.1073/pnas.0805234105>.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. <https://doi.org/10.1093/brain/123.12.2400>.
- Siptár, P., Törkenczy, M., 2007. *The Phonology of Hungarian*. Oxford University Press.
- Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196. <https://doi.org/10.1038/5757>.
- Stoppelman, N., Harpaz, T., Ben-Shachar, M., 2013. Do not throw out the baby with the bath water: choosing an effective baseline for a functional localizer of speech processing. *Brain Behav* 3, 211–222. <https://doi.org/10.1002/brb3.129>.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006. <https://doi.org/10.1038/nn.2163>.
- 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 reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797. <https://doi.org/10.1073/pnas.94.26.14792>.
- Todorovic, A., de Lange, F.P., 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J. Neurosci.* 32, 13389–13395. <https://doi.org/10.1523/JNEUROSCI.2227-12.2012>.
- Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* 31, 9118–9123. <https://doi.org/10.1523/JNEUROSCI.1425-11.2011>.
- Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467. <https://doi.org/10.1037/0278-7393.31.3.443>.
- van der Burght, C.L., Goucha, T., Friederici, A.D., Kreitewolf, J., Hartwigsen, G., 2019. Intonation guides sentence processing in the left inferior frontal gyrus. *Cortex* 117, 122–134. <https://doi.org/10.1016/j.cortex.2019.02.011>.
- van Donselaar, W., Koster, M., Cutler, A., 2005. Exploring the role of lexical stress in lexical recognition. *Q. J. Exp. Psychol. A, Hum. Exp. Psychol.* 58, 251–273.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., Parks, M., 1999. Time course of word identification and semantic integration in spoken language. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 394–417. <https://doi.org/10.1037/0278-7393.25.2.394>.
- Weber, K., Christiansen, M.H., Petersson, K.M., Indefrey, P., Hagoort, P., 2016. fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *J. Neurosci.* 36, 6872–6880. <https://doi.org/10.1523/JNEUROSCI.3180-15.2016>.
- Wicha, N.Y.Y., Bates, E.A., Moreno, E.M., Kutas, M., 2003. Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neurosci. Lett.* 346, 165–168. <https://doi.org/10.1038/jid.2014.371>.
- Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cognit. Neurosci.* 16, 1272–1288. <https://doi.org/10.1162/0898929041920487.Anticipating>.
- Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., Van Den Bosch, A., 2016. Prediction during natural language comprehension. *Cerebr. Cortex* 26, 2506–2516. <https://doi.org/10.1093/cercor/bhw075>.