

# Where Is the Beat? The Neural Correlates of Lexical Stress and Rhythmical Well-formedness in Auditory Story Comprehension

Katerina D. Kandylaki<sup>1\*</sup>, Karen Henrich<sup>2\*</sup>, Arne Nagels<sup>3</sup>, Tilo Kircher<sup>3</sup>, Ulrike Domahs<sup>4</sup>, Matthias Schlesewsky<sup>5</sup>, Ina Bornkessel-Schlesewsky<sup>5</sup>, and Richard Wiese<sup>3</sup>

## Abstract

While listening to continuous speech, humans process beat information to correctly identify word boundaries. The beats of language are stress patterns that are created by combining lexical (word-specific) stress patterns and the rhythm of a specific language. Sometimes, the lexical stress pattern needs to be altered to obey the rhythm of the language. This study investigated the interplay of lexical stress patterns and rhythmical well-formedness in natural speech with fMRI. Previous electrophysiological studies on cases in which a regular lexical stress pattern may be altered to obtain rhythmical well-formedness showed that even subtle rhythmic deviations are detected by the brain if attention is directed toward prosody. Here, we present a new approach to this

phenomenon by having participants listen to contextually rich stories in the absence of a task targeting the manipulation. For the interaction of lexical stress and rhythmical well-formedness, we found one suprathreshold cluster localized between the cerebellum and the brain stem. For the main effect of lexical stress, we found higher BOLD responses to the retained lexical stress pattern in the bilateral SMA, bilateral postcentral gyrus, bilateral middle frontal gyrus, bilateral inferior and right superior parietal lobule, and right precuneus. These results support the view that lexical stress is processed as part of a sensorimotor network of speech comprehension. Moreover, our results connect beat processing in language to domain-independent timing perception. ■

## INTRODUCTION

In everyday language use, the brain processes speech by transforming sound signals to meaning. Identifying the rhythmical pattern of speech (i.e., where the beat is) helps humans to segment the continuous speech signal into meaningful parts, such as words. The local stress pattern of a word, for languages with idiosyncratic stress placement denoted as lexical stress, is realized by means of the phonetic properties fundamental frequency (f<sub>0</sub>), syllable duration, and intensity. For some languages, like English and German, lexical stress information is important because it is used to distinguish lexical items such as *Augúst* (German noun for a month) from *Áugust* (German male first name) and to identify word boundaries within a continuous speech stream (cf. Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). On top of lexical stress lies metrical stress information: When stressed syllables are kept at an equal distance (and thereby creating an isochronous, rhythmically well-formed structure), the rhythm qualifies the language as belonging to the

stress-timed languages (e.g., German, English). Stress-timed languages tend to obey the principle of rhythmic alternation (PRA; Selkirk, 1984; Abercrombie, 1967; Meyer & Cooper, 1960), which defines a well-formed rhythmic structure as an alternation of strong (stressed) and weak (unstressed) units. In the current study, we investigate the interplay of lexical stress and rhythmical well-formedness.

Although natural language is not and cannot be strictly isochronous (Beckman, Edwards, & Fletcher, 1992; Dauer, 1983; Bolinger, 1965), various studies have shown that rhythmic alternation is important in stress-timed languages (Couper-Kuhlen, 1986; Selkirk, 1984; Liberman & Prince, 1977). Also, a predominant sequence of stressed and unstressed syllables supports language acquisition (Nazzi & Ramus, 2003; Jusczyk, Houston, & Newsome, 1999), speech perception and comprehension, word recognition, and segmentation processes (Rothermich & Kotz, 2013; Mattys, 2000; Grosjean & Gee, 1987; Cutler & Foss, 1977). Last, regular sequences are easier to memorize (Auer & Uhmman, 1988; Bolinger, 1981), whereas deviations from regular patterns slow down speech perception (Bohn, Knaus, Wiese, & Domahs, 2013; Pitt & Samuel, 1990) as well as production (Tilsen, 2011).

In recent years, several electrophysiological and neuroimaging studies investigated how violations of both lexical stress and rhythmical well-formedness are processed.

<sup>1</sup>Imperial College London, <sup>2</sup>Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany, <sup>3</sup>Philipps-Universität Marburg, <sup>4</sup>Free University of Bozen-Bolzano, Brixen-Bressanone, Italy, <sup>5</sup>University of South Australia, Magill Campus  
\*These authors contributed equally to this paper.

Their results confirm the importance of identifying stress patterns for facilitated speech comprehension and general language processing, as the brain reacts clearly to lexical as well as to metrical stress violations (Henrich, Wiese, & Domahs, 2015; Henrich, Alter, Wiese, & Domahs, 2014; Bohn et al., 2013; Rothermich & Kotz, 2013; Klein, Domahs, Grande, & Domahs, 2011; Marie, Magne, & Besson, 2011; Rothermich, Schmidt-Kassow, Schwartze, & Kotz, 2010; Schmidt-Kassow & Kotz, 2009a, 2009b; Knaus, Wiese, & Janßen, 2007; Magne et al., 2007; Aleman et al., 2005).

To date, there are only few neuroimaging studies on the processing of lexical stress (Domahs, Klein, Huber, & Domahs, 2013; Klein et al., 2011; Aleman et al., 2005). Aleman et al. (2005) showed that primary auditory and auditory association areas play a prominent role in the processing of metrical (rhythmical) structure in the presence of a metrical decision task. Domahs et al. (2013) showed that primary auditory areas in particular are sensitive not only to the violation of lexical stress patterns but also to different degrees of deviation from these patterns.

Moving on to the systematic investigation of rhythmical well-formedness, few fMRI studies investigated the processing of rhythmically regular and irregular structures in German speech perception (Rothermich & Kotz, 2013; Geiser, Zaehle, Jancke, & Meyer, 2008). Geiser et al. (2008) used pseudosentences with either a strict isochronous or nonisochronous rhythmical structure. When contrasting the explicit rhythmical task with the implicit task, which required focus on the intonation rather than the rhythm, they observed higher activation in the SMA and insula (INS) bilaterally as well as in the left inferior parietal lobule (IPL) and right supramarginal gyrus.

A similar influence of task on speech rhythm was also found in Rothermich and Kotz (2013), a study that investigated metrical deviations in sentences of strict rhythmical regularity compared with irregular sentences. The study also used two different task settings with either explicit (metrical task) or implicit (semantic task) attention on the metrical structure of the sentences. When embedded in an irregular, that is, unpredictable, context, deviations from metrical stress increased activation in the INS bilaterally and in the inferior frontal gyrus (IFG), the head of the caudate in both hemispheres, the right thalamus, the left pre-SMA, and the right superior temporal gyrus (STG). The authors connected this effect to a network that detects irregular patterns comprising the BG, the pre-SMA, and the cerebellum. Moreover, the increased left IFG activation for metrically irregular structures was interpreted as a response to metrical and semantic prediction errors related to the evaluation of the stimuli. An alternative explanation for these activations (especially the pre-SMA and left IFG) could connect it to sensorimotor integration during speech comprehension, as explained in Hickok, Houde, and Rong (2011).

Current neurobiological models of language processing highlight forward modeling as a mechanism for incrementally updating linguistic predictions during comprehension (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015; Hickok et al., 2011; Rauschecker & Scott, 2009). In Bornkessel-Schlesewsky et al. (2015), these predictions are updated in a hierarchical manner according to the length of the predictive cue, in different temporal receptive windows (as demonstrated by Lerner, Honey, Silbert, & Hasson, 2011), and are processed in the human dorsal auditory stream. Given this architecture, deviations from rhythmical and lexical stress patterns should create prediction errors, which would be manifested as signal changes in the dorsal auditory stream (but see Skipper, Devlin, & Lametti, 2017, for a recent perspective that emphasises the importance of predictive processing for speech comprehension, albeit without subscribing to a dual-streams architecture). The “dual stream model of speech processing” (Hickok et al., 2011) also adopts the ventral-dorsal stream architecture and additionally introduces the integration of sensory and motor components of speech representation, partially motivated by Wernicke’s early work (Wernicke, 1969). Last, Hickok and colleagues (2011) assume that forward model predictions of the integrated system can modulate perception of somebody else’s speech.

The previously mentioned neuroimaging studies show the importance of preserving correct lexical stress patterns on the one hand and the compliance with an alternating metrical structure (rhythmical well-formedness) on the other. Moreover, the experimental task strongly influences the activation patterns and partly also the lateralization of activations. To date, however, no study has examined manipulations of linguistic rhythm in a setup that did not require participants to complete a rhythm-related task.

In the current study, we present a new approach to linguistic rhythm processing while listening to contextually rich stories for comprehension. We build upon previous electrophysiological studies on language rhythm by investigating lexical stress patterns and their (non-)compliance with an alternating metrical structure. To this end, we study cases in which a regular lexical stress pattern may be altered to obtain rhythmical well-formedness. The preservation of the lexical stress pattern would otherwise lead to a contravention of the above-mentioned PRA insofar as two adjacent stressed syllables would clash (so-called stress clash; Selkirk, 1984). Potential stress clashes appear most often in noun compound structures (e.g., *Hauptbahnhof* “main train station”) as well as in phonological phrases (e.g., *Termin absagen* “cancel appointment”) when particular words are combined with each other (cf. Wiese, 1996; Kiparsky, 1966). Stress clashes are resolved by shifting the lower-level stress, that is, secondary stress, away from the primary, clashing stress position onto another stressable syllable within the larger linguistic unit (e.g., *Hauptbahnhof* → *Hauptbahn**h**of* “main

train station”). This form of stress shift is also known as the result of the rhythm rule (RR; Liberman & Prince, 1977), a linguistic repair strategy to fulfill the demands of the PRA.

Although this form of stress shift is described in the literature as being optional, it is applied systematically in German (Bohn, Wiese, & Domahs, 2011; Wagner & Fischenbeck, 2002) and English (Grabe & Warren, 1995; Vogel, Bunnell, & Hoskins, 1995). Applying the RR, however, causes the correct lexical stress pattern of one of the included words to be violated (e.g., *Bahnbof*→*Bahnbof* “train station”). It thus seems as if the well-formedness of the metrical structure overrides the lexical stress pattern in these special cases (Selkirk, 1995).

Several ERP studies have investigated this specific link between lexically correct and rhythmically well-formed structures in German (Henrich et al., 2015; Bohn et al., 2013) and English (Henrich et al., 2014). These studies examined lexical and rhythmical irregularities as well as the combination of both. A combined deviation from both lexical and rhythmical stress results through so-called stress lapses. Stress lapses are a further deviation from the PRA, as at least two consecutive unstressed syllables are directly adjacent to each other. Consider cases such as *Féier àbsagen*→*Féier absàgen* “cancel the party.” The rhythmically unlicensed stress shift of a secondary stress onto the verb stem leads to two deviations: a deviation from the lexical stress within the verb as well as a metrical deviation in form of a stress lapse of two adjacent unstressed syllables. The ERP studies on stress clashes and lapses (Henrich et al., 2014, 2015; Bohn et al., 2013) demonstrated that both forms of rhythmic irregularities are perceived and processed differently from well-formed structures. Specifically, ill-formed structures in the form of stress clashes elicited an early frontocentral negativity whereas stress lapses elicited a centroparietal N400 effect (Praamstra & Stegeman, 1993, for such an effect in phonology) in comparison with well-formed structures (Bohn et al., 2013).

Whereas Bohn et al. (2013) employed a task (prosodic naturalness judgment) that directed participants’ attention toward the overall prosodic structure of the sentences, Henrich et al. (2015) showed that an attentional shift toward a word preceding the rhythmically critical structure improved the acceptability of both rhythmic deviations (clash and lapse). This attentional shift made the rhythmic deviations less salient and thus less perceptible, but dissociable brain responses depending on condition were still observable. This finding is in line with the results of the previously mentioned neuroimaging studies and shows that attention and thus the task settings (implicit vs. explicit processing) are critical for the investigation of lexical stress and rhythmical well-formedness.

One open question that remains following the ERP studies by Bohn et al. (2013) and Henrich et al. (2014, 2015) concerns the functional neuroanatomy underlying the mechanisms in question. The observed negativity effects showed different topographies for the lexical and

rhythmical deviations, thus suggesting a possible generation by different cortical networks. To this end, the current study used fMRI to gain further insights into the neurobiological correlates of lexical stress and rhythmical well-formedness in the absence of an explicit prosody-focused task. According to recent neurobiological models of language comprehension (Bornkessel-Schlesewsky et al., 2015; Rauschecker & Scott, 2009), we expect to find different BOLD responses to predicted (regular) and unpredicted (irregular) stress patterns along the dorsal auditory pathway.

## The Present Study

The novel aspect of the current study is the embedding of rhythmically well-formed and ill-formed structures not only in a metrically uncontrolled context, as in Bohn et al. (2013) and Henrich et al. (2014, 2015), but within the rich context of story comprehension. Moreover, in contrast to the previous neuroimaging studies, in which attention was mostly explicitly directed toward the critical lexical or metrical stress structure, this study investigates speech processing in the absence of such a task. Thus, no attention was directed toward speech rhythm or lexical stress distribution.

The following hypotheses were examined:

1. Lexically deviating stress patterns are processed differently depending on rhythmical licensing, thus leading us to expect an interaction of lexical stress and rhythmical well-formedness.
2. Rhythmically well-formed structures produce different activation patterns than rhythmical deviations.
3. These effects appear in the absence of a manipulation-related task.
4. The effects are localized along the dorsal auditory stream.

## METHODS

### Participants

Twenty-two monolingual native speakers of German participated in the study, were all right-handed (Edinburgh Inventory of Handedness; age mean = 24.3 years, *SD* = 2.1 years, male *n* = 6), and were recruited at the University of Marburg. Data from two participants were excluded because of movement artifacts, resulting in a total of 20 data sets for the current study. The study was approved by the ethics committee of the Faculty of Medicine of the University of Marburg. All participants gave informed written consent before participating in the study and were paid €30 for participation.

### Stimuli

For this study, we selected 40 compound noun pairs and embedded them in 22-min-long stories (available online;

Kandylaki, 2016); the same stories included additional manipulations that investigated theory of mind (Kandylaki et al., 2015) and discourse prominence (Kandylaki et al., 2016), respectively. The use of compound structures allows to test long syllable sequences with alternating and nonalternating stress patterns embedded in a larger story context in a rather natural way without creating sentences with artificial semantic structure. Compounds are complex words consisting of two or more constituents, which exist independently as words in the respective language, as in *football coach* or *physics textbook*. Formation of new compounds is common in German. All compounds used here followed the morphological structure (A(BC)), in which one letter signifies a monosyllabic morpheme. The (BC) part consists of two monosyllabic constituents with default stress on the first syllable, as in *Bábn-hof* “station.” The A-constituents are either monosyllabic or disyllabic (the disyllabic bear stress on the first syllable); for example, the word *Bábn-hof* (“station”) was combined with the monosyllabic *Haupt-* (“main”) and the bisyllabic *Gúe-ter-* (“goods”), thereby creating the pair of A(BC) compound words *Haupt-bábn-hof* (“main station,” “central station”) and *Gúe-ter-bábn-hof* (“goods station,” “freight depot”). There was no significant difference in the frequencies of compounds depending on their syllable length, as extracted from the Leipzig corpus (Goldhahn, Eckart, & Quasthoff, 2012) and compared statistically with linear models in R (R Core Team, 2014). The model comparison of a null model (including only the intercept) versus a model with main effect of Syllable number (trisyllabic vs. quattrosyllabic) showed no significant improvement of model fit for the main effect model compared with the null model,  $p > .1$ .

When *Bábn-hof*, which is stressed on the first syllable, is combined with *Háupt-* in a compound, the two stressed syllables *Háupt-* and *Bábn-* are next to each other. According to the RR and the PRA (see Introduction), two adjacent stressed syllables are rhythmically dispreferred; therefore, the stress of *Bábn-* needs to be shifted to the next syllable *-hof*. In this way, the compound word ends up having the following stress pattern: *Háupt-bábn-hòf*, in which *Háupt-* bears main stress (denoted by the acute accent) and *-hòf* bears secondary stress (denoted by the grave accent). This condition is referred to as SHIFT.

On the other hand, when *Bábn-hof* is combined with *Gúe-ter-* in the compound *Gúe-ter-bábn-hof* (“freight depot”), there is no need to shift the lexical stress of *Bábn-hof*, because the syllable preceding *Bábn-* is not stressed. Thus, in this case the phonological structure *Gúe-ter-bàbn-hof* (main stress on *Gúe-* and secondary stress on *-bàbn-*) already obeys the RR and the PRA; this condition is called NOSHIFT, because no shift of the initial stress pattern of *Bábn-hof* is needed. SHIFT and NOSHIFT are the conditions that follow the RR and the PRA, irrespective of the lexical stress of the second compound; they are therefore rhythmically well-formed.

We created two rhythmically ill-formed conditions to investigate rhythmical irregularities more systematically. The first rhythmically irregular condition was CLASH, in which the stress from *Bábn-* is not shifted to *-hof* in the compound *Haupt-bábn-hof*, thereby creating a clash of two adjacent stressed syllables and following the irregular stress pattern: *Háupt-bàbn-hof* (main stress on *Háupt-* and secondary stress on *-bàbn-*). In this condition, the initial lexical stress of *Bábn-hof* is kept at the expense of the overall rhythmic regularity. In addition to the irregularity of two stressed syllables directly following each other, there is another possible irregular structure namely two unstressed syllables to follow each other. In the compound *Gúe-ter-bábn-hof*, if we shift the initial stress of *Bábn-hof* from *Bábn-* to *-hòf*, even though that shift is not needed according to the RR, we end up with the phonological structure *Gúe-ter-bábn-hòf* (main stress on *Gúe-* and secondary stress on *hòf*), in which two unstressed syllables *-ter* and *bábn-* are adjacent. These two adjacent unstressed syllables create a rhythmical lapse; therefore, we named this condition LAPSE. These conditions result in a  $2 \times 2$  design of lexical stress of the second part of the compound (BC) and rhythmical well-formedness, as shown in Table 1.

Each compound was realized in either its well-formed or ill-formed structure as shown in Table 1. The compound *Hauptbábnhof* “central station” appeared in the SHIFT or in the CLASH condition, and the compound *Güterbábnhof* “freight depot” followed either the NOSHIFT or the LAPSE stress pattern. The 80 compound words were used in 20 stories as follows: one story included four different manipulated compound words (one in each condition) at various time points within the story, surrounded by a semantically natural context. We created such a context around the compound, so that its semantic integration to the story was as natural as possible, as in “There the manager explained where the individual parts are produced and how they are transported from the freight depot to the airport.” None of the critical words was at the end of a sentence.

An additional set of 20 “twin” stories was created, in which each compound word was realized in the opposite stress pattern; if Story 1 in version 1A included

**Table 1.** The  $2 \times 2$  Design of Lexical Stress and Rhythmical Well-formedness

		Rhythm	
		Well-formed	Ill-formed
Lexical stress	Correct	NOSHIFT <i>Gúe-ter-bàbn-hof</i>	CLASH <i>Háupt-bàbn-hof</i>
	Incorrect	SHIFT <i>Háupt-bábn-hòf</i>	LAPSE <i>Gúe-ter-bábn-hòf</i>

*Haupt-bahn-hof* in SHIFT, version 1B would include it in CLASH condition. The same would apply for *Güter-bahn-hof*; for Story 2, in version 2A, it would be presented in NOSHIFT and in version 2B, it would be presented in LAPSE condition. This stimulus design resulted in a total of 40 stories with all compounds in all conditions. We could not present the whole set of 40 stories to each participant, because they would be hearing the “twin” of each of the 20 stories and the manipulation would be transparent. Therefore, we created two lists of 20 stories each, and each participant encountered only one of the two lists (hence, only one version of Story 1, either 1A or 1B); in this way each participant heard each compound word in one of its two conditions (e.g., *Haupt-bahn-hof* in either CLASH or SHIFT). Lists were counterbalanced across participants.

All stories were spoken by a professionally trained female speaker of German at a normal speech rate and recorded in a sound-protected laboratory cabin with a sampling rate of 44.1 kHz and a 16-bit (mono) sample size. For sampling, we used the sound recording and analysis software Amadeus Pro (version 1.5.3, HairerSoft, Kenilworth, UK) and an Electret microphone (Beyerdynamic MC930C, Heilbronn, Germany). To obtain the conditions CLASH and LAPSE without manipulating phonetic parameters, the words of the two naturally spoken and recorded conditions SHIFT and NOSHIFT were cut between A and BC components of the compound. The final part (BC) of each word of one condition was spliced onto the first part (A) of the same word of the other condition and vice versa. Hence, the second part of the compound (BC) of the condition NOSHIFT (e.g., *Bahn-hof*) was combined with the initially stressed first part of the compound (A) of the condition SHIFT (e.g., *Haupt*) to create the compound word for CLASH. For the condition LAPSE, the first parts (A) bearing initial stress (e.g., *Güter*) of the condition NOSHIFT were combined with the shifted forms of second parts (BC; e.g., *Bahn-hof*) of the condition SHIFT to obtain two adjacent unstressed syllables. The compounds of the well-formed conditions SHIFT and NO SHIFT were also spliced between first and second parts to avoid measuring a splicing effect. For these conditions, each sentence of the two control conditions was recorded twice and the first sentence part of Recording 1 was spliced with the final sentence part of Recording 2. This procedure has been employed before in previous studies with similar stimuli (Henrich et al., 2014, 2015; Bohn et al., 2013). All compounds were controlled for and normalized in loudness, that is, the volume of all compounds was adjusted to a uniform level of volume throughout all stories. This loudness adjustment was carried out via auditory inspection using the sound recording and analysis software Amadeus Pro. We also applied a phonetic analysis to the stimuli to ensure that the speaker produced real stress shifts in the condition SHIFT and no shifts in the condition NOSHIFT.

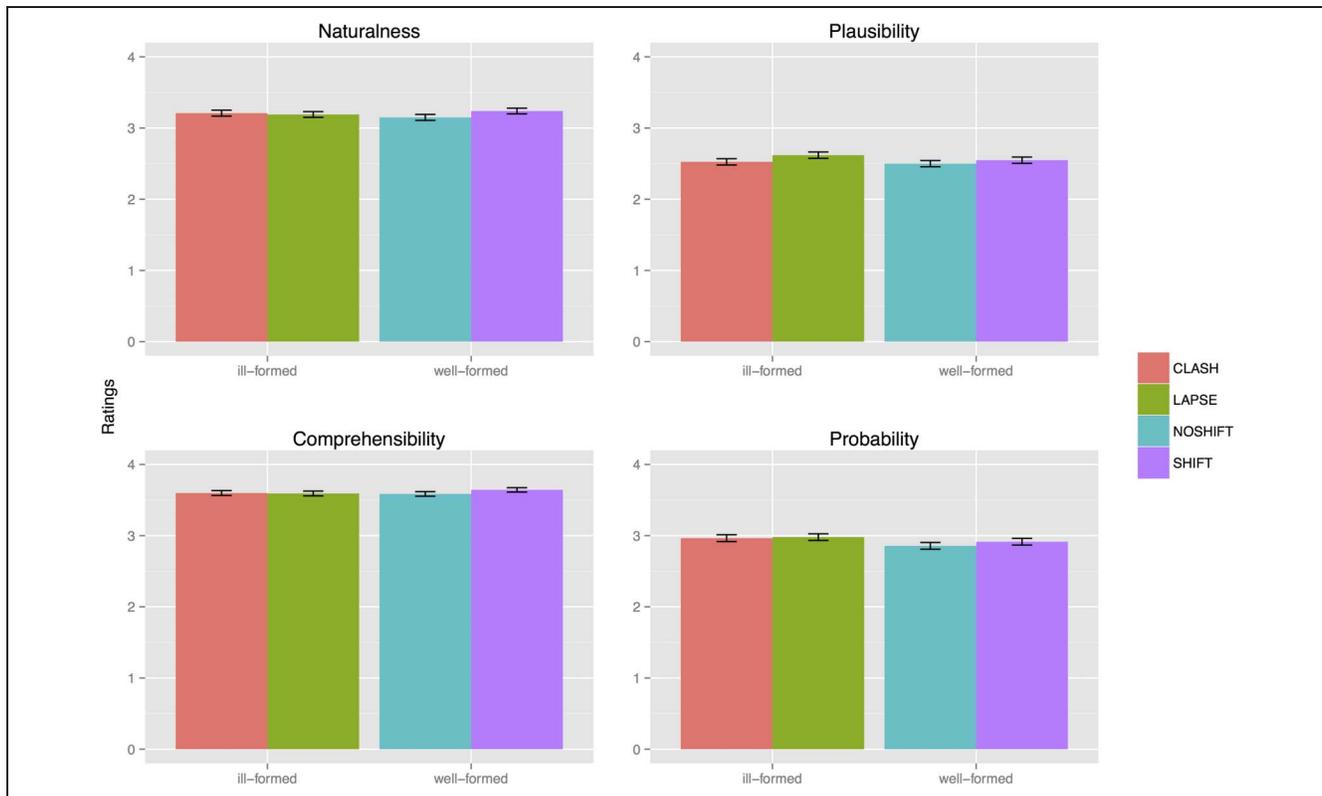
## Pretests

We pretested the stories by means of an online questionnaire for four criteria: (1) naturalness, (2) comprehensibility, (3) plausibility, and (4) probability. The questions in the questionnaire were as follows: (1) How natural was this passage? (2) How comprehensible was this passage? (3) How probable is the event that was described in the passage? (4) How often does this happen? Participants ( $N = 177$ ) answered on a 4-point scale, which was formulated accordingly for each question: 1 = *very natural, rather natural, rather unnatural, unnatural*; 2 = *very well, well, rather less, not at all*; 3 = *very likely, rather likely, rather unlikely, unlikely*; 4 = *very often, once in a while, only sometimes, never*. The ratings scale ranged from 1 to 4, where 1 was minimum and 4 was maximum. Participants rated the passage within which the word appeared, rather than rating the compound word in isolation; the reason for that was that we wanted to capture natural listening similar to the task in the fMRI experiment.

We used generalized linear mixed effects models (package *lme4*, Bates et al., 2014) in R (R Core Team, 2014) to analyze the results of the questionnaire. Figure 1 presents the mean ratings per condition. For the inferential statistics, we first checked the distribution of the data in MATLAB and statistics toolbox release (2016; The MathWorks, Natick, MA) (using the *allfitdistr* function) and then modeled the data accordingly in R. Last, we employed a forward model selection procedure in R, in which we used likelihood ratio tests to compare a base model including an intercept with successively more complex models including rhythmical well-formedness (well-formed vs. ill-formed) and lexical stress (correct vs. incorrect stress of the second compound) as fixed factors. Participants and stories were modeled as random factors; the distribution of the data was binomial, and we took this into account by modeling with the function *glmer* of *lme4* package in R. Only random intercepts by participant and story were included due to convergence problems with more complex random effects structures. We compared the random effects model to the main effects model of Lexical stress and Rhythmical well-formedness (using the ANOVA function in R) and found no significant improvement in the model fit: naturalness:  $p = .408$ , comprehensibility:  $p = .684$ , plausibility:  $p = .138$ , probability:  $p = .399$ . There were thus no systematic differences in ratings of naturalness, comprehensibility, plausibility, and probability varying systematically with regard to our manipulation.

## Imaging Procedure and Behavioral Data Acquisition

Participants went through a training session outside the scanner before the scanning procedure. In the training session, they listened to two stories and answered two questions subsequent to each story (practice stimuli were



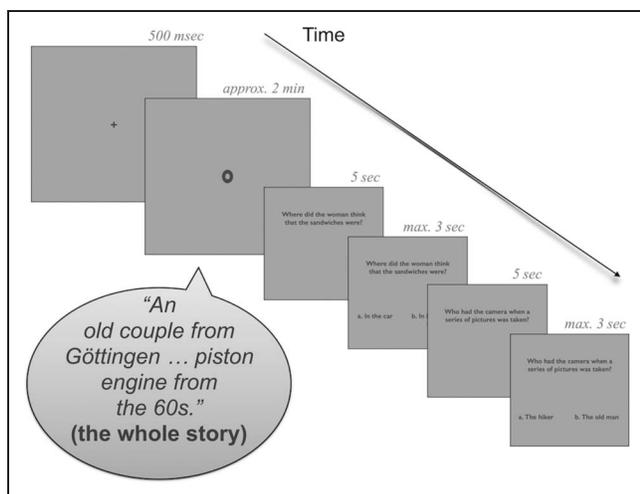
**Figure 1.** Pretest results. Error bars represent standard error.

not used during the subsequent fMRI scan). In the scanner, participants listened to a total of 20 stories and answered to 40 questions (two after each story), spread across four blocks of five stories each (and their two questions after each story). Participants heard the stories through MRI-compatible earphones. Sound loudness was optimized in the scanner before starting the experiment with each participant individually; each participant heard a practice sound and adjusted the volume to their preferred loudness. The order of the stories was assigned randomly for each participant to avoid sequence effects. After the first three blocks of five stories, the participant had a break of 45 sec. The scanner was running during the break, while participants saw the visual message “Short break!” in the middle of the screen.

In the scanner, the stories were presented auditorily, whereas the participant looked at a fixation point in the center of a computer display. After each story, two comprehension questions (referring to the immediately previous story) were presented visually. The questions asked either about the protagonists or about some objects in the story, for example, “How many sandwiches were left for the woman, after the man had finished eating?” Participants were instructed to choose between two possible answers, for example, “One” or “Three.” In this way, we controlled that the participants attended to the content of the stories.

One story trial consisted of the following events: first, a fixation cross was shown in the middle of the screen for 500 msec before the story started. The cross was then

replaced by a fixation point and at the same time the story started. The duration of the story was approximately 2 min ( $\pm 10$  sec). Then the first question was presented visually. The question was presented all at once, centered and toward the top third of the screen for 5 sec. After that, the two possible answers appeared toward the bottom third of the screen, clearly separated from each other; each answer was designated with an index letter, a or b; “a” was always on the left-hand side, and “b” was always on the right-hand side of the participant. The possible answers stayed on the screen until participants made their decision or until a maximal duration of 3 sec (duration pretested, supporting a natural pace throughout the experiment). After the first question, the second question was presented in the same procedure with the same type of content. Participants gave their answers by pressing the middle left or middle right button on a customized response box, which was fixed to their left leg, with their left middle or index finger accordingly. The position of the correct answer was counterbalanced across the experiment. After the second question, a new story trial started, beginning with the fixation cross. All visual stimuli (cross, fixation point, questions and answers, break message) were presented in dark gray on light gray background (see Figure 2). Presentation of stories and questions was time-jittered (jitter duration between 1.5 and 3 sec always assigned randomly) between story and first question and also between first and second question. The procedure was implemented



**Figure 2.** Example trial of one story and its comprehension questions.

and presented with the software package Presentation (Neurobehavioral Systems, Inc., San Francisco, CA).

### fMRI Data Acquisition

During the MR session, a series of EPIs was gathered to record the time course of the participants' brain activity. Measurements were performed on a 3-T MRI system (Trio, A Tim System 3T, Siemens, Erlangen, Germany) with a 12-channel head matrix receive coil. Functional images were acquired using a T2\*-weighted single shot EPI sequence: parallel imaging factor of 2 (GRAPPA), echo time = 25 msec, repetition time = 1450 msec, flip angle = 90°, slice thickness = 4.0 mm and 0.6 mm gap, matrix = 64 × 64, field of view = 224 × 224 mm, in-plane resolution = 3.5 × 3.5 mm<sup>2</sup>, bandwidth = 2232 Hz/pixel, EPI factor of 64, and an echo spacing of 0.53 msec. We gathered 30 transversal slices oriented to the AC–PC line in ascending order.

To avoid saturation and stabilization effects, the initial five images were removed from the analyses of each participant data set. Head movements of the participants were minimized by using foam paddings.

A whole head T1-weighted data set was acquired with a 3d MPRage sequence: parallel imaging factor of 2 (GRAPPA), echo time = 2.26 msec, repetition time = 1900 msec, flip angle = 9°, 1 mm isometrical resolution, 176 sagittal slices, 256 × 256 matrix.

### fMRI Data Analyses

All analyses for the fMRI data were calculated in SPM8 (Wellcome Trust Centre for Neuroimaging), implemented in MATLAB.

A slice time correction (to the 15th slice) was performed first. Then images were realigned to the first image to correct for head movement artifacts. We then normalized the volumes into standard stereotaxic ana-

tomical Montreal Neurological Institute (MNI) space by using the transformation matrix calculated from the first EPI scan of each participant and the EPI template. On the normalized data (resliced voxel size 2 mm<sup>3</sup>), we applied an 8-mm FWHM Gaussian smoothing kernel to compensate for intersubject anatomical variation.

For the single-subject analysis, the design matrix was created separately for each subject, based on the log files from the fMRI session. Our critical events were the whole compound words, which were modeled in seconds (mean duration of event = 0.828 sec, *SD* = 0.095 sec) using the SPM default double gamma hemodynamic response function. As factors of no interest, we modeled separately the rest of the stories (speech processing excluding the critical events), the time for reading of the question (5 sec) and answer (as recorded in the log file), the button presses, and the jitters before each question. The baseline consisted of the 45-sec pauses between blocks. To correct for movement artifacts in each individual session, the realignment parameters were entered as multiple regressors in the first-level analysis.

On the group-level analysis, we modeled a full factorial design of lexical stress and rhythmical well-formedness for the contrasts between the first-level vectors NOSHIFT, CLASH, SHIFT, LAPSE (each against baseline). Brain activations were plotted on the anatomical MRICroN template (the Colin brain). We used the cluster extent thresholding algorithm by Slotnick, Moo, Segal, and Hart (2003), which implements a family-wise error correction using a Monte Carlo simulation to estimate the cluster extent threshold. The desired correction for multiple comparisons was set to  $p < .05$ , and the assumed voxel type I error was set to  $p < .005$ ; after 10,000 iterations, our cluster threshold was estimated at 72 voxels. For all fMRI results reported here, a significance threshold of  $p < .005$  and a cluster extend threshold of 72 voxels were used. For the localization of the clusters, we used the anatomy toolbox of SPM (Eickhoff et al., 2005) and the AFNI tool whereami, which provides anatomical details on the peak voxels, based on four different brain atlases: the standard Talairach–Tournoux Atlas (Talairach & Tournoux, 1988), Eickhoff et al. (2005), Desikan et al. (2006), and Haskins Pediatric Atlas and Template (Molfese, Glen, Mesite, Pugh, & Cox, 2015).

### Contrasts of Interest

In the experimental design of lexical stress and rhythmical well-formedness, we tested the following effects according to our hypotheses:

1. interaction of Lexical stress and Rhythmical well-formedness in an  $F$ -contrast (LS × RW);
2. main effect of Rhythmical well-formedness in an  $F$ -contrast (RW), which tests both SHIFT and NOSHIFT versus CLASH and LAPSE without providing information on the direction of the effect;

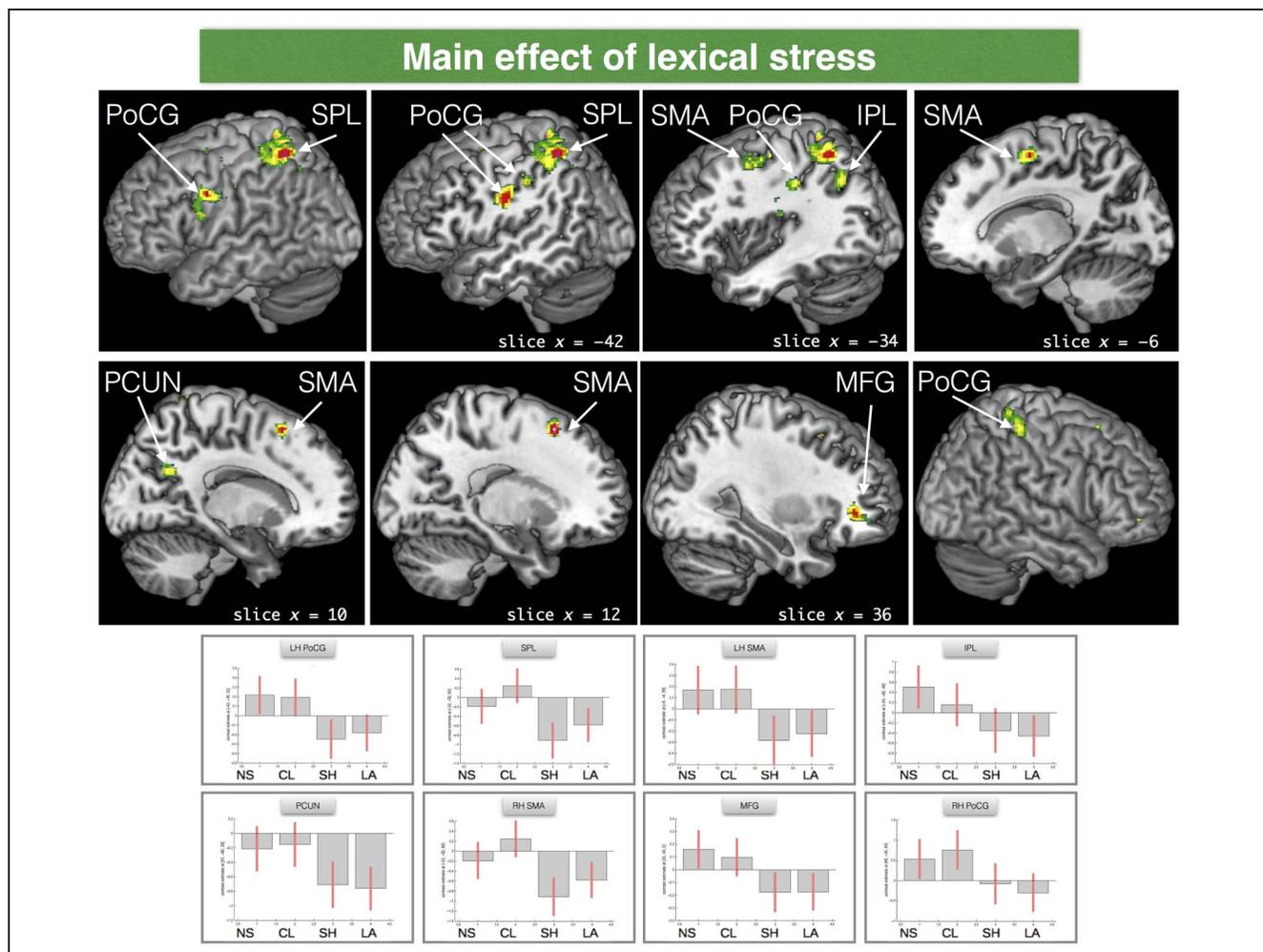
**Table 2.** Peak Voxel Details for the Main Effect of Lexical Stress and for the Interaction between Lexical Stress and Rhythmical Well-formedness

<i>Contrast</i>	<i>Anatomical Region</i>	<i>H</i>	<i>MNI Coordinates</i>	<i>F</i>	<i>Cluster Size in Voxels</i>
LS	SMA	R	12 20 56	20.57	103
	PoCG	L	-42 -16 32	17.37	520
	SPL	L	-32 -52 60	17.30	768
	MFG extending to the orbital part	R	30 40 0	16.48	119
	MFG	L	-26 2 54	14.84	194
	SMA	L	-6 -4 58	14.67	102
	PCUN	R	10 -60 28	12.19	130
	IPL	L	-34 -62 46	11.67	76
	PoCG	R	46 -34 64	11.14	130
LS × RW	Culmen (Lobule IV)	L	-12 -26 -28	12.20	88

<i>Contrast</i>	<i>Anatomical Region</i>	<i>H</i>	<i>MNI Coordinates</i>	<i>t</i>	<i>Cluster Size in Voxels</i>
LAPSE vs. SHIFT	STG	R	64 -26 2	3.71	148

LS = main effect of Lexical stress; LS × RW = interaction between Lexical stress and Rhythmical well-formedness.



**Figure 3.** Localization of the activated clusters for the main effect of Lexical stress, as plotted on the MRICroN template of the Colin brain. ORBmid = orbital part of the middle frontal gyrus; LH = left hemispheric; RH = right hemispheric; NS = NOSHIFT; CL = CLASH; SH = SHIFT; LA = LAPSE.

- main effect of Lexical stress in an  $F$ -contrast (LS), which tests both SHIFT and LAPSE (incorrect lexical stress on BC part of compound) versus NOSHIFT and CLASH (correct lexical stress on BC part of compound) without providing information on the direction of the effect; and
- pairwise comparisons of rhythmically well-formed versus ill-formed conditions (and the opposite) within one level of lexical stress in  $t$  contrasts to identify which conditions elicit the strongest BOLD signal changes. These comparisons were masked inclusively with the interaction mask to identify the activated clusters for the pairwise comparisons given the interaction effect.

## RESULTS

Participants achieved 90% ( $SD = 5.61$ ) correctness in answering the comprehension questions. This performance indicated that the participants listened to the stories attentively.

The results presented in Table 2 showed a main effect of Lexical stress localized in nine clusters with peaks in the bilateral SMA, bilateral postcentral gyrus (PoCG), bilateral middle frontal gyrus (MFG), bilateral IPL, right superior parietal lobule (SPL), and right precuneus (PCUN). Figure 3 shows the localization of these clusters as plotted on the Colin brain template of SPM8 (Wellcome Trust Centre for Neuroimaging) together with the BOLD signal

changes in the peak voxels of the bilateral PoCG, left SPL, bilateral SMA, right IPL, right PCUN, and right MFG. Note that 37% of the right IPL cluster extends to the SPL and to the PCUN. There was no suprathreshold activation for the main effect of Rhythmical well-formedness, but there was one cluster in the brain stem that was activated for the interaction of lexical stress and rhythmical well-formedness (see Table 2 and Figure 4).

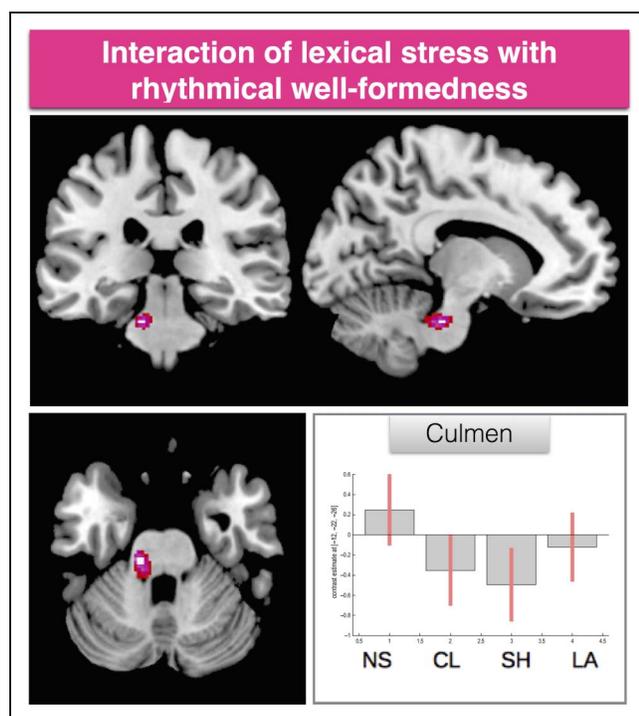
We resolved the interaction by lexical stress to contrast violation of lexical stress in a rhythmically unlicensed context against a rhythmically licensed context (CLASH vs. NOSHIFT and LAPSE vs. SHIFT). Because the mask of the interaction was limited to one cluster, we found suprathreshold clusters when masking with the activation of the interaction effect only for LAPSE versus SHIFT, which was localized in the right STG (see Table 2 for details of the peak voxel).

## DISCUSSION

In this study, we investigated the neural correlates of rhythmic irregularities, as manifested in the interplay between lexical stress and rhythmical well-formedness in natural speech. As stimuli, we used German compounds, which followed the rationale of previous ERP experiments (Henrich et al., 2014, 2015; Bohn et al., 2013), and embedded them into rich narrative context. Thus, participants encountered the stimuli while listening to stories. After each story, they answered two comprehension questions presented visually, which did not draw attention to the prosodic manipulation.

We employed a full factorial design of rhythmical well-formedness (well-formed vs. ill-formed) crossed with lexical stress (correct: *Báhn-hof* vs. incorrect: *Bahn-bòf*). The higher BOLD responses for correct lexical stress (NOSHIFT and CLASH) in comparison with incorrect lexical stress (SHIFT and LAPSE) were largely bilateral and observable in frontoparietal regions, in the SMA and the precuneus. This activation pattern is in line with the findings on lexical stress manipulations by Domahs et al. (2013), in which higher BOLD responses for correct versus incorrect lexical stress were found in the left angular gyrus and right retrosplenial cortex. More importantly, because irregular rhythmical patterns in this study resemble the mild violations of Domahs et al. (2013), our results are in accordance with their findings of bilateral SMA and left angular gyrus activation, which overlaps with our left IPL. In addition, our findings go one step further than those of Aleman et al. (2005) in that the left hemispheric activation of SMA, pre-central gyrus, PoCG, and SPL and the right hemispheric activation in SMA, PoCG, and INS have now been shown to scale up for lexical stress processing, even in natural story listening and without a lexical or rhythmical discrimination task.

The comprehension task of this study was similar to the semantic task of Rothermich and Kotz (2013) in that it did not draw attention to the stress patterns. However,



**Figure 4.** Localization of the activated clusters for the interaction of lexical stress with rhythmical well-formedness, as plotted on the MRICroN template of the Colin brain. NS = NOSHIFT; CL = CLASH; SH = SHIFT; LA = LAPSE.

in our study, the presented story stimuli were longer than Rothermich and Kotz's (2013) single-sentence stimuli: In a 2-min-long story, only small parts contained a manipulation relevant to rhythmical processing (four compound words, each of approximately 800-msec length in each story). In addition, the stories did not follow a completely isochronous rhythmical structure but comprised natural speech with naturally alternating rhythm and natural prosody. The only congruent finding in Rothermich and Kotz's (2013) semantic task was the activation of the right IFG, which may be topographically overlapping with our right orbitofrontal cluster found for the main effect of Lexical stress.

Rothermich and Kotz (2013) found significant results (bilateral IFG, right STG, and left pre-SMA) in their metrical task; they also report that responses to the metrical task were generally stronger compared with the semantic task. A possible explanation for this, according to the authors, was that the prosodic structure of auditory signals is not processed explicitly. In our study, no suprathreshold clusters for the main effect of Rhythmical well-formedness were found. This finding may be attributed to two different reasons or a combination thereof: First, the subtlety of the rhythmical violations, in that stress shifts are usually preferred in production but are not necessarily perceived consciously (Bohn et al., 2011). In contrast, violations of lexical stress patterns, as those studied by Domahs et al. (2013), constitute deviations from the phonological stress encoded in the lexicon, which are not generally licensed. The cited studies revealed that, in contrast to semantic or syntactic violations, stress violations seem to be harder to detect (cf. Domahs et al., 2013; Rothermich & Kotz, 2013; Klein et al., 2011; Geiser et al., 2008). Second, the combination of natural listening to stories and no explicit task drawing attention to the stress patterns of the stimuli, but instead a content-related question related to the plot of the stories. The fact that no main effect for Rhythmical well-formedness could be found in this study is in accordance with the results of Henrich et al. (2015), in which both rhythmical well-formedness and lexical stress deviations were less salient and perceptible if attention was directed away from them.

For the interaction of lexical stress and rhythmical well-formedness, one cluster showed significant activation, the peak of which was localized in the culmen according to the Anatomy Toolbox of SPM8 (Eickhoff et al., 2005). Visual inspection of the plotted activation showed the cluster to be located largely within the brain stem. Because linguistic stress is partly realized in excursions of fundamental frequency, this result is congruent with findings of Krishnan, Gandour, and Bidelman (2012). They recorded brain stem responses to pitch alternations between English nonmusicians, English musicians, and Chinese speakers with EEG. They expected to find a graded effect of brainstem sensitivity based on the experience of pitch processing: Musicians were highly trained to discriminate pitch contours, Chinese speakers are

sensitive to changes of pitch because of their training in a language using lexical tones, and English nonmusicians should show less sensitivity compared with the previous groups, because they speak a stress-timed language and are not trained to discrimination of lexically distinctive pitch. Indeed, group comparisons of spectral  $f_0$  magnitudes showed the hypothesized graded effect, thereby supporting the role of the brain stem in pitch encoding.

### Natural Language Rhythm in a Sensorimotor Network of Speech Perception

Our findings can be interpreted within the recent framework of sensorimotor integration during speech perception (Hickok et al., 2011). In this article, the authors propose that the auditory system of the human brain is heavily involved in speech production as well as that the motor system is involved in speech perception, as part of a feedback control loop. This loop is suggested to be localized in the dorsal auditory stream, in accordance with recent neurobiological theories that highlight its importance in auditory processing (Rauschecker & Scott, 2009; for a recent perspective including higher-order language processing, see Bornkessel-Schlesewsky et al., 2015). More specifically, according to the literature (Hickok et al., 2011; Peelle, Johnsrude, & Davis, 2010), this stream connects the primary auditory cortex through a sensorimotor interface in temporoparietal regions and through articulatory areas (premotor cortex) to the inferior frontal cortex. The activation of this stream for speech processing is assumed to be left lateralized. In this study, we found supporting evidence for this view in the IPL, SMA, and orbitofrontal activations, along with the dorsal auditory stream, as stated in Hypothesis 4. However, our results support the assumed lateralization of the network only partially: The IPL activation was left lateralized, the SMA activation was bilateral, and the orbitofrontal activation was localized in the right hemisphere. Updating this model based on the current findings on naturalistic language stimuli could include the possibility of the activation pattern occurring bilaterally, as a result of speech comprehension, which is not restricted by a highly targeted task.

The dual stream model of Hickok et al. (2011) also states that perception may be modulated by rhythmical patterns of the auditory signal, as a result of predictive processing during speech comprehension. In relation to rhythmical prediction in language, we could not provide evidence supporting the processing of subtle rhythmical irregularities as non-predicted rhythm patterns. The direction of the observed effects showed higher BOLD responses for the pattern of the correctly stressed second constituent (*Bahn-hof* compared with *Bahn-höf*, thus the lexical stress), which would contradict a possible top-down prediction of the stress of the next syllable based on the rhythmical context. Specifically, after

*Haupt-*, if the brain follows the RR, it should expect an unstressed syllable. In contrast to that, and based on our fMRI results, the brain seems to be sensitive only to the lexical pattern of *Bahnbof* and not necessarily make predictions based on the hypothesized harmonic rhythm pattern of stressed and unstressed syllables. Again, this might be due to the given task and attention settings. If attention is not directed toward the prosodic structure, the strength of the included violations is essential for their detection: Whereas subtle rhythmic deviations might not be detected, deviations from lexical stress provide clearer violations may be more easily recognized, leading to higher processing costs and stronger activations possibly due to the need for reanalysis. This view would regard the perception of correct stress patterns and deviations from them as a result of a backward-looking operation, such as the reconstruction of the correct lexical stress pattern. In contrast, the absence of a detection of subtle rhythmic deviations may be explained in two ways: Either participants did not automatically process rhythmic alternations during story comprehension or they did not automatically use this information to form predictions.

Further support for the distinction between strong and subtle stress deviations comes from the fact that the resolution of the interaction by lexical stress showed a significant effect only for LAPSE versus SHIFT, but not for CLASH versus NOSHIFT: when unattended, the combination of two, that is, a lexical as well as a rhythmical deviation in LAPSE, is easier to detect than the single subtle rhythmical deviation in CLASH. This is in line with results by Domahs et al. (2013), who showed that the strength of a violation influences their detection, as subtle and milder violations of the word prosodic structure are more error-prone than more severe and thus clearer violations. Interestingly, lexical stress interacts with rhythmical well-formedness only if a shift and thus a deviation of lexical stress is included. This shows that higher activation is found only if deviations from lexical stress are not rhythmically licensed.

### Lexical Stress Processing within Domain-independent Timing Perception

A recent meta-analysis investigated the role of the SMA (and its subsections pre-SMA and SMA proper) in temporal processing (Schwartz, Rothermich, & Kotz, 2012). Although this study excluded experiments which used complex stimuli such as music or speech, the implications can be transferred to the language domain as testable research hypotheses. The bilateral SMA activation of this study supports its involvement in temporal sequencing of complex acoustic structures. According to Schwartz et al. (2012), the temporal processing network, which includes the pre-SMA and SMA proper, may form the neurobiological basis of temporal structure perception across different modalities such as music or speech perception. Our study highlights the involvement

of the bilateral SMA in lexical stress processing as part of domain-independent temporal processing.

### Conclusions

This study was the first to investigate neural responses to the interplay of lexical stress and rhythmical well-formedness within natural language contexts of auditory stories and in the absence of a task that would draw the participants' attention to the acoustic features of the stimuli. The results pointed to a sensorimotor activation pattern that included the bilateral SMA, the left IPL and SPL, the left precuneus, as well as the right orbitofrontal regions. The present findings offer insights into current neurobiological theories of speech processing and contribute to the neural underpinnings of stress pattern recognition as a domain-independent computation related to timing perception.

Reprint requests should be sent to Katerina D. Kandylaki, Imperial College London, Bioengineering, RSM Building, South Kensington Campus, London SW7 2AZ, United Kingdom, or via e-mail: a.kandylaki@imperial.ac.uk.

### REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., De Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*, 221–228.
- Auer, P., & Uhmann, S. (1988). Silben- und akzentzählende Sprachen: Literaturüberblick und Diskussion. *Zeitschrift für Sprachwissenschaft*, *7*, 214–259.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*.
- Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In *Papers in Laboratory Phonology II* (pp. 68–86). Cambridge: Cambridge University Press.
- Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, *51*, 760–771.
- Bohn, K., Wiese, R., & Domahs, U. (2011). The status of the rhythm rule within and across word boundaries in German. In *Proceedings of ICPHS XVII* (pp. 1–4). Hong Kong.
- Bolinger, D. (1965). Pitch accent and sentence rhythm. In *Forms of English: Accent, morpheme, order* (pp. 139–180). Cambridge, MA: Harvard University Press.
- Bolinger, D. L. M. (1981). *Two kinds of vowels, two kinds of rhythm*. Reproduced by the Indiana University Linguistics Club. Bloomington, IN.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, *19*, 142–150.
- Couper-Kuhlen, E. (1986). *An introduction to English prosody*. London: Edward Arnold.
- Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.
- 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 and Language*, *125*, 272–282.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335.
- Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*, *20*, 541–552.
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In *LREC*. (pp. 759–765).
- Grabe, E., & Warren, P. (1995). Stress shift: Do speakers do it or do listeners hear it. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology (Vol. 4)*, pp. 95–110. New York: Cambridge University Press.
- Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25*, 135–155.
- Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language*, *136*, 19–30.
- Henrich, K., Wiese, R., & Domahs, U. (2015). How information structure influences the processing of rhythmic irregularities: ERP evidence from German phrases. *Neuropsychologia*, *75*, 431–440.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*, 407–422.
- Juszyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.
- Kandylaki, K. D. (2016). *Stories*. Ph.D. thesis. Available at doi:10.6084/m9.figshare.3122515.v1
- Kandylaki, K. D., Nagels, A., Tune, S., Kircher, T., Wiese, R., Schlesewsky, M., et al. (2016). Predicting “when” in discourse engages the human dorsal auditory stream: An fMRI study using naturalistic stories. *Journal of Neuroscience*, *36*, 12180–12191.
- Kandylaki, K. D., Nagels, A., Tune, S., Wiese, R., Bornkessel-Schlesewsky, I., & Kircher, T. (2015). Processing of false belief passages during natural story comprehension: An fMRI study. *Human Brain Mapping*, *36*, 4231–4246.
- Kiparsky, P. (1966). Über den deutschen Akzent. *Studia Grammatica*, *7*, 69–98.
- Klein, E., Domahs, U., Grande, M., & Domahs, F. (2011). Neurocognitive foundations of word stress processing—Evidence from fMRI. *Behavioral and Brain Functions: BBF*, *7*, 15.
- Knaus, J., Wiese, R., & Janßen, U. (2007). The processing of word stress: EEG studies on task-related components. In *Proceedings of the International Congress of Phonetic Sciences 2007* (pp. 709–712). Saarbrücken, Germany.
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2012). Experience-dependent plasticity in pitch encoding: From brainstem to auditory cortex. *NeuroReport*, *23*, 498.
- Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*, 2906–2915.
- Lieberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex*, *17*, 2659–2668.
- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*, 294–305.
- MATLAB and statistics toolbox release. (2016). Natick, MA: The MathWorks, Inc. The MathWorks, Inc., Natick, Massachusetts, United States.
- Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, *62*, 253–265.
- Meyer, L., & Cooper, G. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Molfese, P. J., Glen, D., Mesite, L., Pugh, K., & Cox, R. W. (2015). *The Haskins pediatric brain atlas*. 21st Annual Meeting of the Organization for Human Brain Mapping (OHBM), Honolulu, 2015. Retrieved from <https://afni.nimh.nih.gov/pub/dist/HBM2015/>.
- Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*, 233–243.
- Peelle, J. E., Johnsrude, I., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, *4*, 51.
- Pitt, M. A., & Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language*, *29*, 611–632.
- Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory n400 event-related brain potential. *Cognitive Brain Research*, *1*, 73–86.
- R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna*. <http://www.R-project.org/>.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724.
- Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *Neuroimage*, *70*, 89–100.
- Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport*, *21*, 580–584.
- Schmidt-Kassow, M., & Kotz, S. A. (2009a). Attention and perceptual regularity in speech. *NeuroReport*, *20*, 1643–1647.
- Schmidt-Kassow, M., & Kotz, S. A. (2009b). Event-related brain potentials suggest a late interaction of meter and syntax in the p600. *Journal of Cognitive Neuroscience*, *21*, 1693–1708.
- Schwartze, M., Rothermich, K., & Kotz, S. A. (2012). Functional dissociation of pre-SMA and SMA-proper in temporal processing. *Neuroimage*, *60*, 290–298.
- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory (Blackwell handbooks in linguistics 1)*, pp. 550–569. Oxford: Blackwell.
- Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, *164*, 77–105.
- Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Research. Cognitive Brain Research*, *17*, 75–82.
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: An approach to cerebral imaging*. Stuttgart: Thieme.
- Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, 2, 185–218.
- Vogel, I., Bunnell, H. T., & Hoskins, S. (1995). The phonology and phonetics of the rhythm rule. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology* (Vol. 4, pp. 111–127). New York: Cambridge University Press.
- Wagner, P., & Fischenbeck, E. (2002). Stress perception and production in German stress clash environments. In *Proceedings of Speech Prosody 2002*. Aix en Provence.
- Wernicke, C. (1969). The symptom complex of aphasia. In *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968* (pp. 34–97). Netherlands: Springer.
- Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press.