

Tuning Neural Phase Entrainment to Speech

Simone Falk^{1,2,3}, Cosima Lanzilotti⁴, and Daniele Schön⁴

Abstract

■ Musical rhythm positively impacts on subsequent speech processing. However, the neural mechanisms underlying this phenomenon are so far unclear. We investigated whether carry-over effects from a preceding musical cue to a speech stimulus result from a continuation of neural phase entrainment to periodicities that are present in both music and speech. Participants listened and memorized French metrical sentences that contained (quasi-)periodic recurrences of accents and syllables. Speech stimuli were preceded by a rhythmically regular or irregular musical cue. Our results show that the presence of a

regular cue modulates neural response as estimated by EEG power spectral density, intertrial coherence, and source analyses at critical frequencies during speech processing compared with the irregular condition. Importantly, intertrial coherences for regular cues were indicative of the participants' success in memorizing the subsequent speech stimuli. These findings underscore the highly adaptive nature of neural phase entrainment across fundamentally different auditory stimuli. They also support current models of neural phase entrainment as a tool of predictive timing and attentional selection across cognitive domains. ■

INTRODUCTION

A remarkable discovery of the last few years is that musical experiences can facilitate certain aspects of speech processing. Long-term musical training enhances representations of pitch, timbre, and timing in speech; fosters speech-in-noise recognition, verbal memory, and speech segmentation; and promotes reading skills in typically developing as well as dyslexic children (Flaugnacco et al., 2015; Elmer et al., 2014; François & Schön, 2014; Kraus, Hornickel, Strait, Slater, & Thompson, 2014; François, Chobert, Besson, & Schön, 2013; Strait, Parbery-Clark, Hittner, & Kraus, 2012; Moreno et al., 2009). These effects are attributed to enhanced neural plasticity in auditory and speech-related areas of the brain that have been observed both at a functional and morphological levels (Schlaug, 2015; Kraus et al., 2014; Herholz & Zatorre, 2012; Kraus & Chandrasekaran, 2010; Hyde et al., 2009; Jäncke, 2009). Besides long-term effects of musical training, listening to musical rhythm can shape the encoding of speech at a shorter term. For instance, Przybylski et al. (2013) showed that listening to 30 sec of highly regular music improved the accuracy in a syntactic task in both dyslexic children and children with specific language impairment. Similar results were also found in Parkinson patients (Kotz & Gunter, 2015). At an even shorter timescale (i.e., a few seconds), we found that a rhythmical sequence of tones matching the rhythm of the syllabic and accentual

structure of a following pseudoword or short sentence enhanced speech perception in adults and speech production in hearing impaired children (Cason, Astésano, & Schön, 2015; Cason, Hidalgo, Isoard, Roman, & Schön, 2015; Cason & Schön, 2012). These findings are compatible with recent theories of expectancy-driven speech processing (Kotz & Schwartz, 2010) and indicate that auditory temporal expectancies induced by musical patterns can facilitate speech perception (Schön & Tillmann, 2015). However, the neural processes scaffolding music-to-language transfers during the encoding of musical and speech rhythms are so far unclear. Here, we investigate the possibility that phase-coupling of neural oscillations with the temporal structure of speech increases because of neural resonance with a preceding musical rhythm.

Behaviorally, humans synchronize with musical rhythms when they move their hands, heads, or legs to the beat. Neurally, listening to music stimulates a subcortical and cortical audiomotor network. Periodicities of neural oscillations phase-lock to the musical beat (Tierney & Kraus, 2015; Fujioka, Zedel, & Ross, 2010; Jones, 2009). This coupling of intrinsic (i.e., endogenous) to sensory (i.e., exogenous) rhythms has also been termed “entrainment” (Calderone, Lakatos, Butler, & Castellanos, 2014; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Large & Jones, 1999). The process is described as a tuning of neural excitability toward expected times that may even continue when sensory stimulation has stopped (Calderone et al., 2014; Arnal & Giraud, 2012). Nozaradan, Peretz, and Mouraux (2012) demonstrated that neural entrainment to music is not merely the result of a response to low-level acoustics of the sound envelope but occurs when an internal metrical representation (i.e., nested levels of perceived periodicities) of the sound is built. Using steady

¹Aix-Marseille Univ, LPL, UMR 7309, CNRS, Aix-en-Provence, France, ²Université Sorbonne Nouvelle Paris-3, LPP, UMR 7018, CNRS, Paris, France, ³Ludwig-Maximilians-University, Munich, Germany, ⁴Aix Marseille Univ, INSERM, INS, Marseille, France

state-evoked potentials, the authors showed that the peaks of neural resonance followed the structure of metrical musical patterns. The strongest peak was found at the beat level (e.g., at 1.25 Hz), but peaks in the power spectral density (PSD) were also present at higher and lower harmonics (at 0.625, 2.5, and 5 Hz), representing higher-order beat groupings as well as subdivisions down to the level of single tones. Neural entrainment is not exclusive to music, though. Research has highlighted the fact that speech elicits robust neural entrainment, in particular at low frequencies in the 4–8 Hz (theta) range (Giraud & Poeppel, 2012; Peelle & Davis, 2012; Giraud et al., 2007; Luo & Poeppel, 2007; Ahissar et al., 2001). This frequency range has been associated with the temporal signature of syllable production (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009). For instance, when reducing acuity of the spectral envelope, theta-band activity decreases in the auditory cortex in parallel to speech intelligibility (e.g., Doelling, Arnal, Ghitza, & Poeppel, 2014). Further studies also suggest that nested hierarchies of cortical oscillations entrain to speech, from delta rhythms (<4 Hz), which represent fluctuations in speech prosody such as intonation and accent structure (Giraud & Poeppel, 2012) to gamma rhythms (>30 Hz; e.g., Ding & Simon, 2014; Gross et al., 2013). Eventually, top-down signals from frontal and motor regions can modulate entrainment to speech in the auditory cortex, with larger effects on the delta-band frequencies (Park, Ince, Schyns, Thut, & Gross, 2015). In summary, processing of both music and speech rhythms relies on coupling of brain oscillatory activity to nested levels of perceived event periodicities in these complex sounds.

Importantly, it has been suggested that neural entrainment optimizes the processing of predictable events by providing a way for the sensory systems to encode ongoing information (Calderone et al., 2014). For instance, entrained EEG delta phase for rhythmic but not nonrhythmic stimuli enhances visual contrast sensitivity (Cravo, Rohenkohl, Wyart, & Nobre, 2013). Similarly, in the auditory domain, the phase of entrained oscillations correlates with RT to target stimuli (Stefanics et al., 2010) and target detection accuracy (Henry, Herrmann, & Obleser, 2014) and accelerates the perceptual emergence of task-relevant temporal patterns (Riecke, Sack, & Schroeder, 2015). Similar functional contributions of neural entrainment have been shown for speech stimuli (Luo & Poeppel, 2007; Ahissar et al., 2001), and the tracking of speech envelope information has been suggested to be a good candidate to subserve the function of syllabic parsing (Doelling et al., 2014). Moreover, entrainment to speech seems to go beyond the acoustic characteristics. It is affected by listeners' ability to extract linguistic information (Zoefel & VanRullen, 2015b; Peelle, Gross, & Davis, 2013) and can track grammar-based linguistic structures (Ding, Melloni, Zhang, Tian, & Poeppel, 2016).

In this study, we examine the hypothesis that carryover effects from a preceding musical stimulus to a rhythmically

matching speech stimulus result from a continuation of neural phase entrainment to periodicities that are present in both the musical and speech stimuli. We presented participants with metrical speech stimuli containing quasiperiodic recurrences of syllables (at 5 Hz) and periodic recurrences of accented syllables (at 1.65 Hz). The speech stimuli were preceded either by a rhythmically regular tone sequence matching the metrical structure of the speech pattern or sequences with an irregular temporal structure. We predicted stronger oscillatory activity and inter-trial phase locking in speech following matching musical sequences ("regular cue") compared with mismatching unstructured sequences ("irregular cue"). Moreover, we also examined whether oscillatory activity during cue and speech presentation may relate to behavioral response.

METHODS

Participants

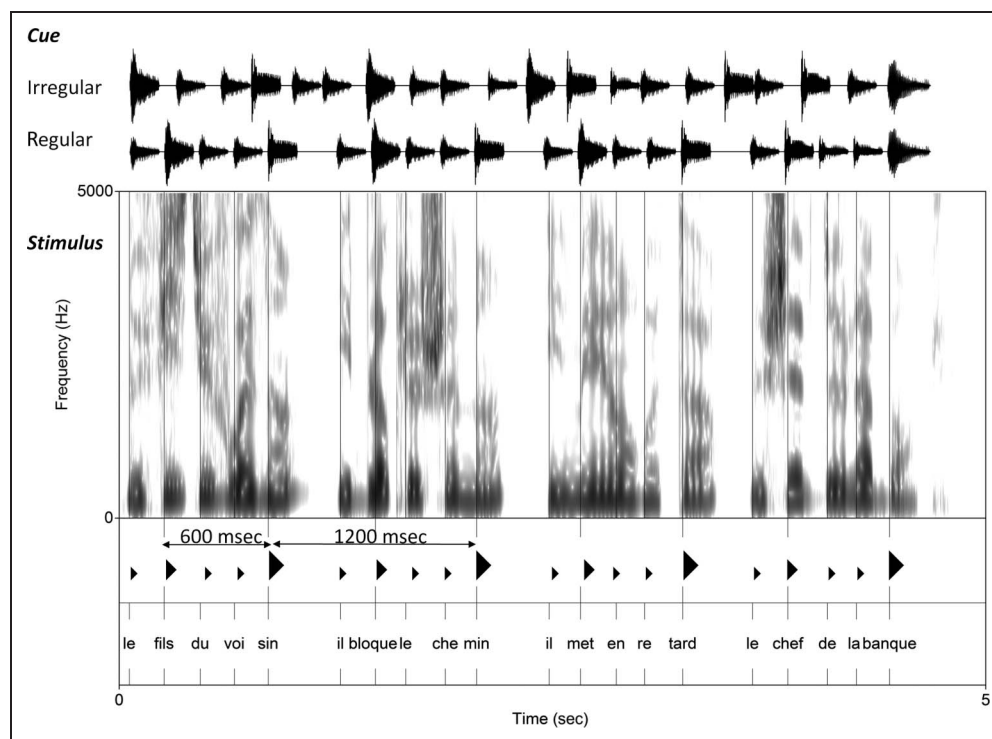
Twenty-four French participants (13 men, mean age = 27, $SD = 5$) took part in this study. All had normal hearing and normal or corrected-to-normal vision. All gave informed consent to participate in the study that was approved by the local ethics committee. Each participant received €20 at the end of the experiment.

Stimuli

The linguistic material consisted of 60 spoken French utterances sharing the same syntactic structure (i.e., consisting of two short sentences similar to the example in Figure 1; see also Falk, Volpi-Moncorger, & Dalla Bella, 2017). Every utterance comprised 20 syllables subdivided in four accentual phrases of five syllables each. The utterances were recorded (at 44,100 Hz, then downsampled to 22,050 Hz) by a female speaker who was trained to pronounce each accentual phrase with a Low-High-Low-High intonation pattern (LHLH; Welby, 2006; Jun & Fougeron, 2000), comprising a prenuclear ("secondary") accent on the second syllable and the nuclear ("primary") accent on the last syllable (see Figure 1).

The utterances were cued before each recording by a metronome to be read at a regular pace (i.e., 600 msec interonset intervals between accented syllables). If necessary, manual corrections were done (using PRAAT; Boersma, 2001) to obtain a highly regular meter with an average interonset interval of 600 msec (± 20 msec) between accented syllables (measured from the syllabic p-center, which was estimated via the algorithm of Cummins & Port, 1998). No further manual adjustments were done on syllabic intervals. The average utterance duration was 4.8 sec (range = 4.7–5 sec). Figure 2B shows the spectral representation of the temporal envelope averaged over all the 60 speech stimuli, with clear peaks at 0.8, 1.65, and 5 Hz, corresponding to the frequency of primary accent recurrence, primary + secondary accent recurrence, and syllable recurrence, respectively.

Figure 1. Oscillograms of the nonverbal cues (top) and an example of a verbal stimulus (bottom) (“the son of the neighbor blocks the road, he delays the bank director”). Big triangles mark primary accents, medium triangles mark secondary accents, and small triangles mark nonaccented syllables.



Two types of musical stimuli were built to precede the speech stimuli in the experiment, regular and irregular. The regular rhythmic cue consisted of four sequences that stylized average pitch and temporal structure of the speech stimuli. The first three sequences were physically identical and consisted of five isochronous cello pizzicato sounds (at 22,050 Hz) lasting 200 msec each (F#, ↑G, G, G, ↑C, arrows indicate ascending or descending intervals). The last sequence had the same temporal structure but different pitches similar to the final intonation pattern of the speech stimuli (G, ↑Bb, Bb, Bb, ↓Eb). The last note lasted 250 msec to mimic sentence final lengthening on the last syllable. The specific pitches were chosen because

they best reflected the average pitch contour present in the spoken sentences, with minor adjustments to fit into a tonal system. Between the four sequences, a silence of 200 msec was inserted to create a regular ternary meter. The ternary meter was also explicitly marked by an increase in intensity on the second and last sound of each sequence (see Figure 1). Four versions of irregular cues were obtained by scrambling the pitches and the silences of the regular cue. This resulted in four different cues that were rhythmically irregular, thus preventing participants to become acquainted with the irregular structures. Figure 2A shows the spectral representation of the envelope of regular and irregular cues. The peaks visible in the

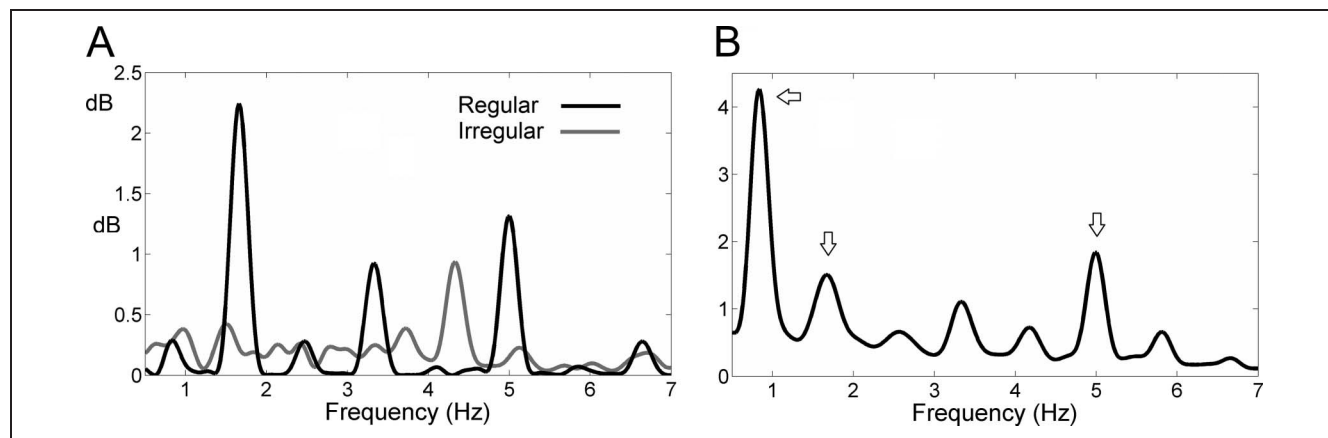


Figure 2. Audio signal. (A) Mean envelope PSD for regular and irregular cues computed after band-pass filtering (300–2500 Hz) and Hilbert transform. Note that the peak at 3.3 Hz for the regular cue is the first harmonic of the 1.65 Hz peak, although this frequency was not present in the metrical structure. (B) Mean envelope PSD for target sentences. The arrows point to the frequencies of interest.

regular cue reflect the temporal structure of the cue, with the ternary meter represented by the peak at 1.65 Hz (=600 msec) and the note level presented by the 5 Hz peak (=200 msec).

Procedure

The 60 speech stimuli were presented twice in two sessions separated by a short break. In a given trial, a cue was presented, regular or irregular, immediately followed by a speech stimulus. The interval between the rhythmic cue and the speech stimulus was manipulated in such a way that the SOA between the last note of the cue and the first accent in the utterance was always 600 msec. At 2140 msec after the end of the speech stimulus, a content word (i.e., noun, verb, or adjective) was presented on the screen, and participants had to press one of two buttons to decide whether the word was present in the previously heard utterance or not. Half of the words were present in the utterance, whereas the other half were not. Of those that were not present, half were synonyms or semantically closely related to a word in the utterance with comparable lexical frequency (e.g., *to shock*–*to upset someone*).

Stimulus presentation was structured in miniblocks of four trials. Within each miniblock, only one type of cue, regular or irregular, was used. Speech stimuli that were preceded by a regular cue in Session 1 were preceded by an irregular cue in Session 2. The distribution of speech stimuli across miniblocks and the order of conditions (regular vs. irregular) of miniblocks were pseudo-randomized and counterbalanced across participants. The software Presentation (Neurobehavioral Systems, Berkeley, CA) was used to program the experiment. A sound Blaster X-Fi Xtreme Audio, an amplifier Yamaha P2040, and Yamaha loud speakers (NS 10M) were used for sound presentation.

Data Analyses

Analysis of Behavioral Data

Participants' answers in the detection task were analyzed as Hits when the participant correctly notified a word as different from the words presented in the utterance. A false alarm (FA) occurred when the participant reported a difference when there was actually none. Detection scores were calculated by subtracting percentage of FAs from percentage of Hits. d' was not used because of the absence of FA in some participants. Detection times were analyzed for Hits only. Trials with values greater than two standard deviations from the mean were excluded from the analysis. Because of a technical problem with the button press setup, we could not collect a complete set of responses from three participants that were thus discarded from behavioral analyses.

Signal Processing and Statistical Analyses

EEG signal was recorded at 1000 Hz sampling rate using a BrainAmp amplifier and 64 preamplified Ag–AgCl electrodes mounted following the 10–10 international system (actiCap) in a soundproof and Faraday cage. The ground electrode was placed at AFz and the reference electrode at FCz.

Signal processing was done using EEGLAB (Delorme & Makeig, 2004, v12.022) and custom MATLAB (The MathWorks, Natick, MA) scripts. Continuous data were filtered using a high-pass filter (0.4 Hz, 12 dB/octave, Hamming windowed sinc FIR filter) and major artifacts rejected by visual inspection. Independent component analysis (infomax) was used to remove physiological artifacts such as eyeblinks and muscular activity. Data were then segmented into epochs of 5 sec starting at the onset of both cues and speech stimuli, separately. Epochs were zero-mean normalized and re-referenced to the algebraic average of all electrodes. Further rejection of remaining artifacts was done on the basis of visual inspection of each epoch and was always lower than 10% of the total number of epochs in a given condition (yielding on average 55 epochs per condition per participant).

The spectral analysis was performed using the Welch's periodogram to compute the PSD over a window of 5000 msec using a Hanning tapering and no overlap. The output consisted of 5000 amplitude estimates with a frequency bin width of 0.1 Hz (zero-padding ratio of 2 to visually improve spectral resolution). The transformation was computed separately for each participant, channel, condition, and over non-averaged epochs (i.e., regular and irregular cues and speech stimuli).

We performed two types of frequency analysis. The analysis of the PSD allowed estimating possible global changes in power at target frequencies. This method does not take into account the phase information. The intertrial coherence (ITC) was used to estimate the signal consistency across trials at each time–frequency point, taking into account both phase and power information. This latter measure has been shown to be more sensitive to stimulus synchronized neural activity than power (Ding & Simon, 2013). To control for the typical $1/f$ -like trend in the low-frequency EEG spectrum (Bedard, Kroeger, & Destexhe, 2006), the PSD was $1/f$ -corrected and values at a target frequency (e.g., 0.8, 1.65, and 5 Hz) were divided by the average value of two lower and two higher neighbors (e.g., for 5 Hz: 4.7, 4.8, 5.2, 5.3 Hz). The validity of this normalization procedure relies on the assumption that, in the absence of entrainment, the signal amplitude at a given frequency should be similar to the signal amplitude of the mean of the surrounding frequencies (Nozaradan et al., 2012; Nozaradan, Peretz, Missal, & Mouraux, 2011). Thus, this normalization ensures that ratios greater than 1 actually indicate the presence of entrainment at specific frequencies. Statistical analyses were run on a ROI (Fz, Cz, F1, F2, FC1, FC2, FC3, FC4) best describing the 5-Hz activity in the rhythmic

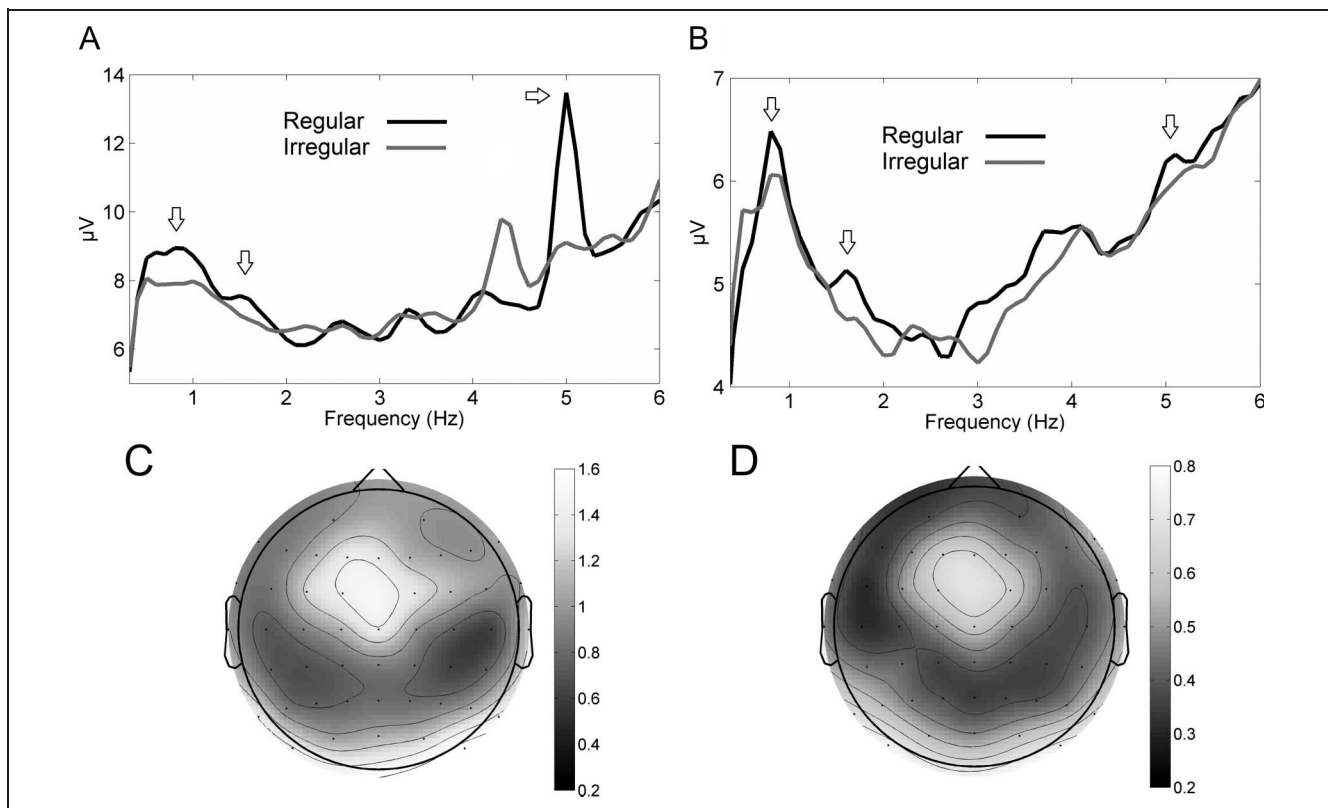


Figure 3. EEG signal. Top row. (A) $1/F$ rectified PSD for regular and irregular cues. (B) $1/F$ rectified PSD for sentences preceded by regular and irregular cues. Bottom row. (C) Topographic maps of the raw PSD at 5 Hz for regular cues. (D) Topographic maps of the raw PSD at 5 Hz for sentences (after regular and irregular cues).

cues, which was at the core of our working hypothesis. The ROI was identified by using a nonparametric cluster-based permutation approach (Maris & Oostenveld, 2007) that allowed to select one single cluster with $p < .01$ (see Figure 3C). As our aim was to study possible changes in oscillatory activity during speech perception, the choice of the ROI on the basis of the 5-Hz topography during the cues was important to prevent a data selection bias for the statistical analyses of the speech stimuli. Statistics were used to verify whether (1) ratios were significantly greater than 1, indicating the presence of a peak (Wilcoxon signed-rank, one-sample), and (2) the difference of ratios in the regular and irregular conditions was different from zero (Wilcoxon signed-rank, paired).

Although not the primary aim of this study, we used sLORETA (Pascual-Marqui, 2002) to localize the source of the response difference to speech stimuli in regular and irregular conditions in the two most important frequency bands (1.65 and 5 Hz). sLORETA method computes the three-dimensional distribution of electrically active neuronal generators within the brain as current density values based on the recorded scalp electric potential differences and then estimates the inverse problem using the assumption that the smoothest of all possible activities is the most plausible. The analysis was conducted with a three-shell spherical head model registered to the Talairach atlas (voxel dimension 5×5 mm). Statistical

significance was assessed by means of a nonparametric randomization test ($p < .05$, corrected for multiple comparisons).

Finally, to measure the phase coupling across trials at the expected frequencies, we computed the ITC between 0 and 7 Hz using a discrete fast Fourier transform and Hanning tapering (using the EEGLAB *newtimef* function with fast Fourier transforms and symmetric Hanning window tapering, *padratio* of 4). The values for each frequency were then averaged across time to yield a phase-locking index for each estimated frequency. Again statistical analyses were run on a ROI (Fz, Cz, F1, F2, FC1, FC2, C1, C2, FC3) best describing the 5-Hz ITC activity and selected in the same way as it was done for the PSD analyses. Nonparametric statistics were used to verify whether the difference of phase-locking index across all estimated frequencies in the regular and irregular condition was different from zero (Wilcoxon signed-rank, paired, FDR-corrected).

RESULTS

Behavioral Task

Overall, participants' detection scores were high ($M = 79\%$, $SD = 10.4\%$), indicating that the task was rather easy. No differences between detections in speech stimuli

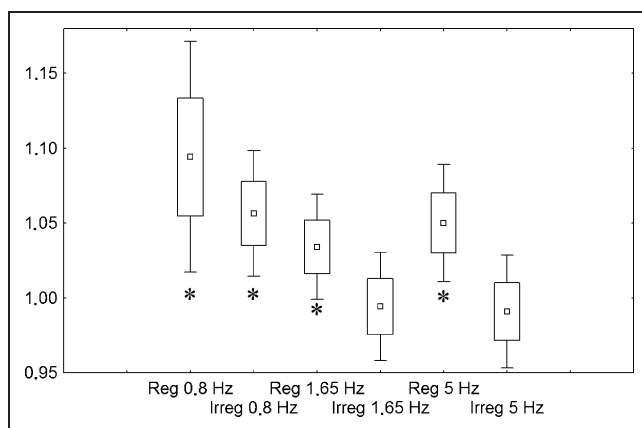


Figure 4. EEG signal. Normalized amplitude ratios at the expected frequencies for speech stimuli preceded by regular and irregular cues. Values larger than 1 indicate a peak in the spectrum. Whiskers represent the within-subject standard error from the mean.

following regular or irregular cues were found (regular cue: mean score = 78%, $SD = 9.1\%$; irregular cue: mean score = 80%, $SD = 12.2\%$; $p > .60$), nor was there a difference in detection times (regular cue: $M = 957$ msec, $SD = 235.1$ msec; irregular cue: $M = 955$ msec, $SD = 244.2$ msec, $p > .80$).

Entrainment to Speech after Regular/Irregular Cue Presentation

First, we examined whether the PSD estimates varied as a function of the preceding musical cues when participants listened to speech (recall that the speech stimuli were always metrical sentences and that the same stimuli were presented repeatedly after different cues varied in regu-

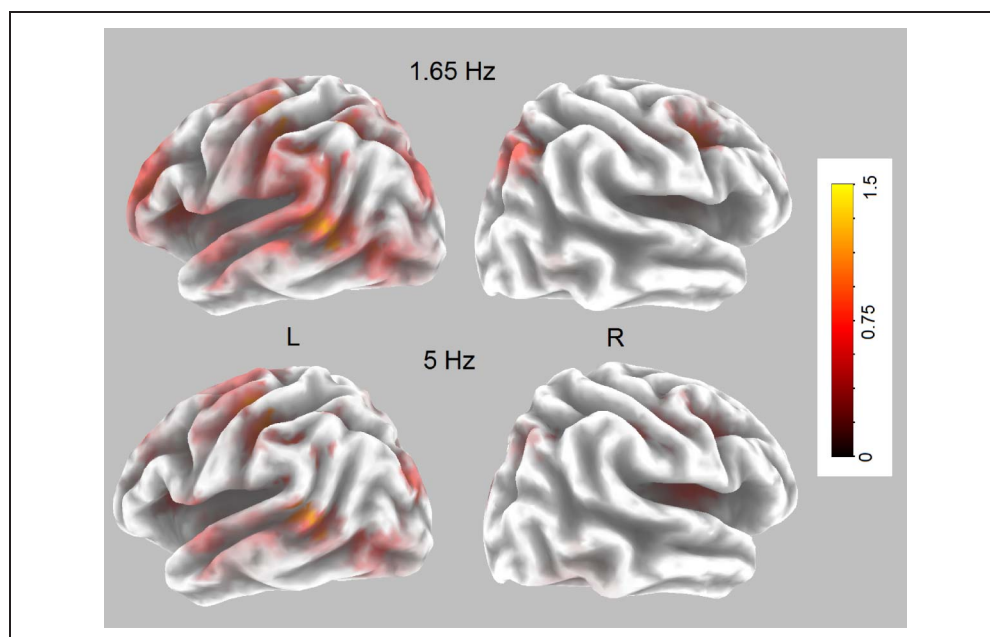
larity). Overall, averaged over cue condition, several peaks were visible in the PSD of speech at expected frequencies of accent and syllable recurrence (5 Hz, 1.65 Hz, and 0.8 Hz; see Figure 3B), reflecting the three largest peaks visible on the acoustic PSD of the speech stimuli (5 Hz, 1.65 Hz, 0.8 Hz; Figure 2B).

Figure 4 shows that the normalized amplitude ratios at the expected frequencies for speech stimuli preceded by regular cues were always significantly above 1 (5 Hz: $z = 2.1$, $p = .02$; 1.65 Hz: $z = 1.6$, $p = .05$; 0.8 Hz: $z = 2.1$, $p = .02$). This means that the peak value was larger than the immediately surrounding values. When speech was preceded by irregular cues, only the peak at 0.8 Hz was larger than the neighboring values ($z = 2.1$, $p = .02$).

Direct comparisons of spectral peaks during speech perception in the regular and irregular conditions showed a trend for larger peaks in the regular condition at 5 and 1.65 Hz, although these differences did not reach significance ($z = 1.8$, $p = .07$; $z = 1.7$, $p = .08$). No differences were present for the 0.8 Hz ($z = 0.8$, $p = .4$). However, the same contrast using nonparametric bootstrapping analyses on current source densities as estimated by sLoreta did show significantly larger responses at both 1.65 and 5 Hz for speech stimuli preceded by a regular cue, compared with those preceded by an irregular cue (Figure 5). These differences reached significance, at the source level, in the left hemisphere over the middle and superior temporal gyri and middle frontal and precentral gyri.

Results of the ITC analyses for the speech stimuli also revealed a clear peak at 5 Hz for speech stimuli preceded by regular cues (Figure 6B, top right). This peak corresponds to the 5-Hz peak visible on the envelope's PSD of the speech stimuli (Figure 2B). Statistical comparisons of ITC for sentences preceded by regular and irregular

Figure 5. Statistical nonparametric maps (t values) of the difference between speech stimuli preceded by regular and irregular cues (after bootstrapping) for the two most important target frequencies (1.65 Hz, top and 5 Hz, bottom) over left and right hemispheres.



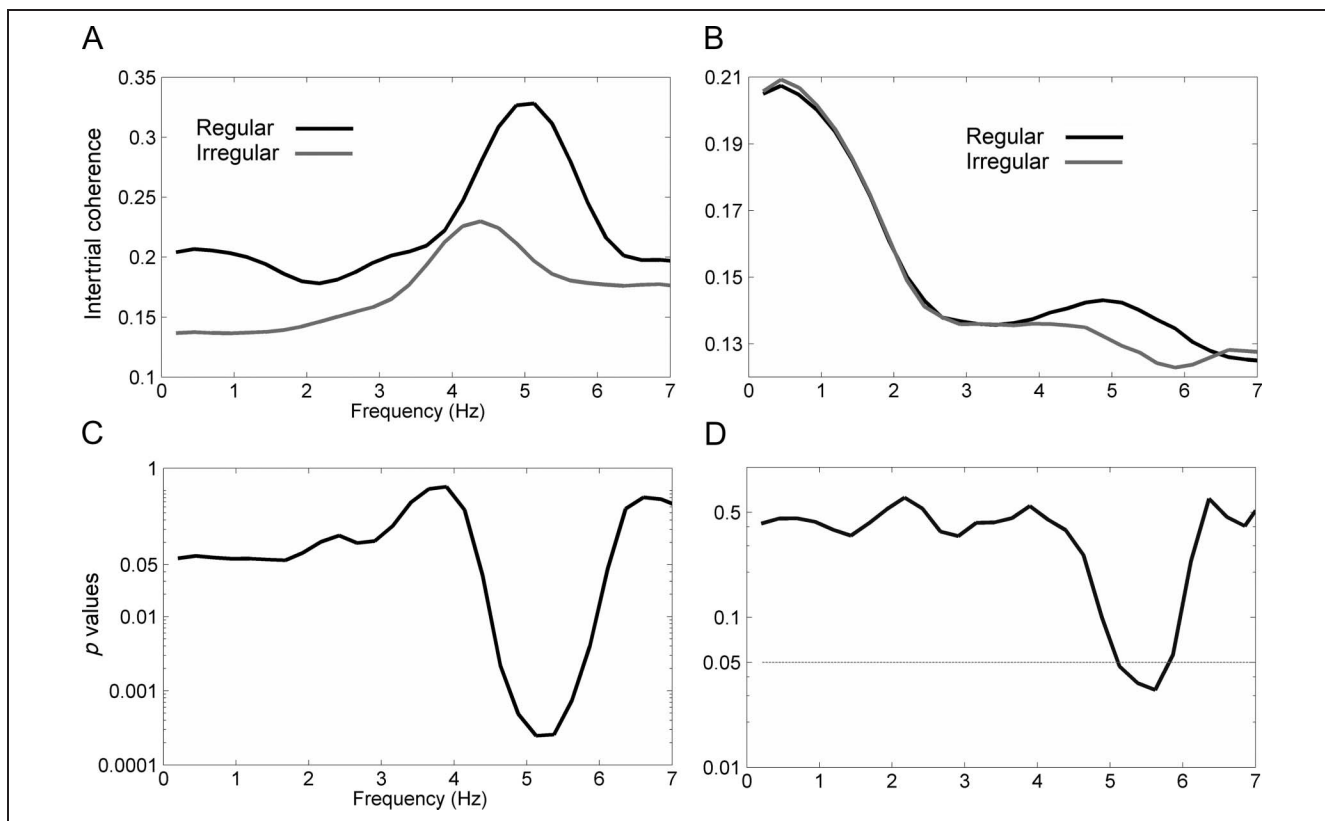


Figure 6. Top row. (A) ITC for regular and irregular cues. (B) ITC for sentences preceded by regular and irregular cues. Bottom row. (C) p Values illustrating the differences between regular and irregular cues as a function of frequency. (D) p Values illustrating the differences between sentences preceded by regular and irregular cues as a function of frequency.

cues showed significantly larger coherence in the regular condition around 5 Hz with p values $<.05$ between 5 and 5.8 Hz (Figure 6D).

Entrainment during Cue Presentation

The joint PSD, ITC, and sLoreta analyses on speech processing revealed stronger neural phase coupling at expected frequencies at 5 Hz and, to some extent, on 1.65 Hz, during the processing of identical sentences after regular compared with irregular cue presentation. These results fit well with our initial hypothesis that a nonverbal musical cue can modify entrainment to a subsequent speech stimulus. However, to underpin the idea of carryover effects from cue to speech, we needed to verify whether neural phase entrainment was already present at the specific target frequencies during listening to the rhythmic structure of the regular musical cue.

As expected from the acoustic PSD analyses (Figure 2A), results of the PSD for musical cues showed clear peaks at 5 Hz for the regular and at 4.3 Hz for the irregular cues (Figure 3A). As to the rhythmical target frequencies, statistical analyses revealed significant peaks at 5, 1.65, and 0.8 Hz for the regular cues ($z = 4.03$, $p < .0001$; $z = 2.9$, $p = .02$, $z = 3.8$, $p = .0001$, respectively). For the irregular cues, the peaks at these frequencies

were smaller and mostly not significant (5 Hz: $z = 1.7$, $p = .05$; 1.65 Hz: $z = 1.6$, $p = .06$; 0.8 Hz: $z = 1.3$, $p = .09$). Direct comparisons of spectral peaks during perception of regular and irregular cues showed larger peaks during the regular condition at 5 and 0.8 Hz ($z = 3.7$, $p < .001$; $z = 2.5$, $p = .01$), whereas the difference at 1.65 Hz did not reach significance ($p = .15$).

Paralleling PSD analyses, ITC analyses also showed a clear peak at 5 Hz for the regular cue and at 4.4 Hz for the irregular cue (Figure 6A). Statistical comparison of ITC for regular and irregular cues showed significantly larger coherence for regular cues between 4.4 and 6 Hz with p values $<.001$ between 4.8 and 5.6 Hz (Figure 6C).

Overall, these results indicate that neural entrainment at target frequencies (in particular 5 Hz) already took place at the stage of regular cue presentation.

Behavioral Responses and Neural Entrainment

Finally, despite the fact that the behavioral task did not in itself yield differences between regularity conditions, we examined whether behavioral responses were associated with neural entrainment in the cue conditions. Consequently, ITC measures for the cues and for speech were correlated with the behavioral response accuracy of participants. Importantly, the ITC at 5 Hz measured during

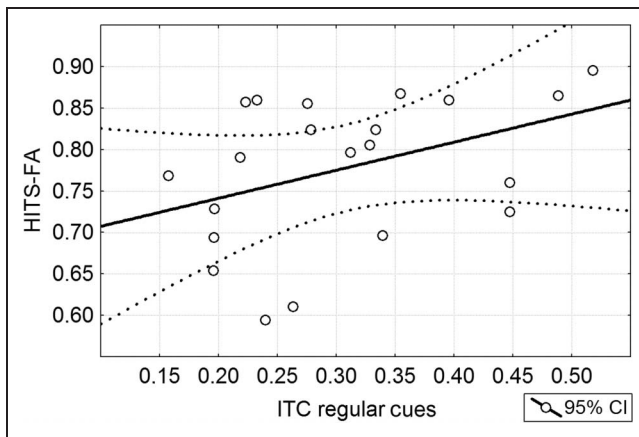


Figure 7. Scatter plot of the performance in the behavioral task (% of Hits minus % of FAs) versus ITC during regular cue processing (preceding the speech stimuli) for each participant. The dark line is the linear fit, and the dotted lines indicate 95% confidence intervals.

regular cue presentation was a good predictor of the overall accuracy in the behavioral task (regular: $r = .45$, $p < .05$, irregular: $r = .3$, *ns*, Spearman nonparametric correlation, see Figure 7). By contrast the ITC measured during speech presentation did not correlate with behavioral responses ($r < .1$).

DISCUSSION

We investigated whether neural phase entrainment to the rhythmic structure of a nonverbal musical cue enhances neural entrainment to a subsequent speech stimulus. Participants listened to rhythmically regular or irregular nonverbal cues preceding the presentation of a spoken utterance that matched the rhythm of the regular cue. Our results show that listening to the regular cue modifies neural response at critical frequencies in the speech stimulus compared with the irregular condition, especially at the syllable level. Analyses on the surface (PSD) and source level jointly indicated stronger phase coupling at 5 Hz and, to a smaller extent, at 1.65 Hz after regular versus irregular cue presentation. This finding was corroborated by analyses of intertrial phase coherence (ITC), which index phase similarities across trials, and revealed stronger phase coupling of neural responses around 5 Hz during regular versus irregular cue processing. Finally, the behavioral task we chose for the study turned out to be very easy for the participants and, therefore, may have failed to show differences in memory for words between the regular and irregular condition. This lack in sensitivity may also be at the origin of the lack of correlation between the behavioural measure and neural activity during speech perception. Yet, individual ITCs for regular cues were indicative of the participants' success in memorizing the exact wording of the following speech stimuli. This suggests that the state of the network at the beginning of the sentence

(induced by the regular cues) seems to play an important role in speech perception. Overall, the result integrates well with findings from other tasks that showed behavioral benefits of nonverbal cueing for speech processing (Falk, Volpi-Moncorger, & Dalla Bella, 2017; Falk & Dalla Bella, 2016; Cason, Hidalgo, et al., 2015; Kotz & Gunter, 2015; Cason & Schön, 2012).

These findings support the hypothesis that the processing of metrical levels in speech can be influenced by metrical levels heard before in a nonverbal (musical) sequence via neural entrainment. Neural oscillations (1) phase-lock to and (2) are amplified at event periodicities defined by the metrical levels that are present in both stimulus streams (Lakatos et al., 2008). Thus, entraining neural oscillations with a temporally clearly structured auditory rhythm results in a facilitated entrainment to speech when the temporal structure of speech and the musical rhythm coincide. The observed carryover effect from the musical cue to speech raises the question if processes underlying neural entrainment for (rhythmically regular) verbal and nonverbal stimuli follow similar principles. Some theories describe the function of neural entrainment to beat and meter as enhancing “predictive timing” (Fujioka, Ross, & Trainor, 2015) and thereby, attending toward salient events in an auditory stimulus (Kotz & Schwartz, 2010; Large & Jones, 1999). In line with the oscillatory selection hypothesis (Schroeder & Lakatos, 2009; see Frey, Ruhnau, & Weisz, 2015, for an overview), it has been proposed that neural phase entrainment is a consequence of a rhythmic processing mode (in contrast to “continuous processing mode”) that the brain adopts by aligning phases of high neural excitability with salient (“attended”) events in a rhythmic signal (e.g., Peelle et al., 2013). In contrast to music, which typically features musical beat structure, such a process is less obvious for spontaneous speech that varies substantially in its temporal regularity and predictability of prominent prosodic events (e.g., accents or syllables; see Cummins, 2015; Arvaniti, 2009). However, intense behavioral and neuroimaging research of the past years has convincingly demonstrated that rhythmically predictable verbal contexts (e.g., through regular metrical patterns such as found in poetry, nursery rhymes, etc.) benefit spoken language processing through enhanced predictive timing and higher attending toward relevant events in speech (Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013; Rothermich, Schmidt-Kassow, & Kotz, 2012; Schwartz, Rothermich, Schmidt-Kassow, & Kotz, 2011; Schmidt-Kassow & Kotz, 2009). Furthermore, beneficial carryover effects between nonverbal and verbal rhythmic contexts have been observed. A nonverbal but temporally predictable auditory cue mapping onto metrical structure in speech can induce temporal predictions that direct attending toward upcoming events in speech at predicted times (Falk, Volpi-Moncorger, & Dalla Bella, 2017; Falk & Dalla Bella, 2016; Kotz & Gunter, 2015). As a consequence, facilitation of speech perception and production processes have been observed (Cason, Astésano,

et al., 2015; Cason, Hidalgo, et al., 2015; Cason & Schön, 2012). However, to our knowledge, our study provides first evidence of how these carryover effects are implemented in terms of brain dynamics by increasing both the amplitude and the phase coupling strength at linguistically salient periodicities.

As for music, our results replicated previous findings showing phase-locking in the theta and delta range to metrical levels of musical signals (e.g., Nozaradan et al., 2012; Schaefer, Vlek, & Desain, 2011; Fujioka et al., 2010; Brochard, Abecasis, Potter, Ragot, & Drake, 2003). For example, Nozaradan et al. (2011) found that a periodic stream of pure tones with a 2.4-Hz frequency elicited steady-state evoked potentials at this frequency and at subharmonic resonances at 1.2 or 0.8 Hz. The phase-locking could be altered depending on imagined binary or ternary groupings of the sound. Thus, beat perception was largely based on predictions generated through the participants' imaginary groupings. In the present study, we also found enhanced spectral peaks in neural response (at 5 and 0.8 Hz) as well as substantially higher ITC (at 5 Hz), when participants listened to the regular musical cue compared with the irregular cue. Interestingly, although our stimuli were much shorter (i.e., 5 sec) than in the aforementioned study (i.e., 33 sec), their metrical levels (0.8, 1.65, and 5 Hz) were still clearly visible in the neural responses. This was possibly due to the fact that they exhibited acoustic information pointing to metrical (i.e., enhanced intensity of tones in a ternary meter) and musical structure (i.e., pitch variations, phrase-final lengthening, timbre of a musical instrument). At least when using an ecological cue, even a short exposure may be sufficient to entrain neural response to metrical levels. Note that the presence of spectral peaks in the EEG response to the musical cue could also be accounted for by a regular succession of evoked responses. Interestingly, the oscillatory activity did not wear off immediately when the musical stimulus stopped. Neural response to the following speech stimulus was also enhanced at points in time that were predictable from the rhythmically regular cue. This maintenance of regular neural activity across time and stimuli would be unusual for evoked activity but could be explained in terms of entrainment of oscillatory activity. This finding indicates that neural oscillatory dynamics seem to be maintained, even when the stimulus quality changes from nonverbal to verbal.

In this study, we used highly metrical and temporally predictable structures in both verbal and musical stimuli. Despite the close metrical matching of both types of stimuli, the carryover effect from the musical cue to the speech stimulus via sustained neural entrainment is not trivial, as speech was still substantially more variable than the musical cue in several respects. First, the speech stimuli varied considerably in semantic and intonational structure. Second, the speech stimuli displayed higher temporal variability at the basic metrical level of syllable durations than the tones in the cue, as no temporal

adjustments were done on syllabic intervals. This may explain why we found enhanced ITC at precisely 5 Hz in the regular cue, but an enhancement in a broader range around 5 Hz in speech. Third, although the musical beat level was marked by higher intensity in the cue (at 1.65 Hz), this acoustic marker was not present in the French speech stimuli in which accents are marked by pitch and duration variations rather than intensity (Welby, 2006). Nevertheless, a peak is visible in the EEG power spectrum (Figures 3B and 4) at 1.65 Hz in speech (i.e., the accentual level) only after regular cue presentation. These are important findings that highlight the adaptive mechanisms of neural oscillatory behavior (e.g., Gross et al., 2013). Flexible adaptation requires at least two properties of the system. First, the neural entrainment shown at different frequencies during nonverbal stimuli is robust to small frequency detuning during speech and can be quickly retuned to new frequencies. Second, the phase of the oscillatory system keeps track of previous oscillatory states (cue) when adapting to new oscillatory dynamics (speech), thus explaining the greater entrainment to speech when preceded by regular cues.

Given two nonidentical oscillators, frequency entrainment will take place depending on frequency detuning, namely the difference between two oscillators. The greater the difference, the smaller should be the coupling (although the relation is not linear). For example, the 4.3 Hz generated by the irregular cues in our study did not seem to enhance the 5 Hz of subsequent speech (Figure 3A). By contrast, if the mismatch is not very large, the frequencies of two systems become equal or entrained, that is, synchronization takes place. Generally, the width of the synchronization region will increase with coupling strength (Pikowsky, Rosenblum, & Kurths, 2001). Future research should further examine the possible regions of synchronization in the case of auditory cueing.

It is also a remarkable finding, that a PSD peak was present at 1.65 Hz in the neural response to speech following the regular cue, but not in the response to the regular cue itself. This result may point to a process of oscillatory adaptation to metrical levels that increases in strength over time, provided that the metrical level is sufficiently salient and interpretable in the signal (such as the regular accent structure in speech). It is indeed an open question how stimulus interpretation may affect the process of neural entrainment to speech and music (see Zoefel & VanRullen, 2015a, for a discussion). There is an ongoing debate whether the brain merely tunes in to salient acoustic events (i.e., constituted by the peaks in the spectral amplitude envelope) or if oscillatory activity is modulated by higher-order linguistic structure in speech (i.e., sustained by linguistic predictions from the phonological, syntactic, semantic domain; Zoefel & VanRullen, 2015b, 2016). Evidence for a “more-than-acoustics” account in speech comes from studies showing that neural phase entrainment with speech in the 5-Hz frequency range is influenced by speech intelligibility (Doelling

et al., 2014; Gross et al., 2013; Peelle et al., 2013; Peelle & Davis, 2012; Luo & Poeppel, 2007; Ahissar et al., 2001). In music research, there is also evidence that neural phase entrainment is modulated by higher-order hierarchical structure and its interpretation (e.g., Fujioka et al., 2015; Nozaradan et al., 2011, 2012). Future studies should further investigate the conditions under which neural entrainment is altered by higher-order processes and predictions specific to language and music.

In summary, cross-domain effects of neural phase entrainment from music to speech such as observed in this study require a highly flexible mechanism driving oscillatory adaptation. The boundaries of flexibility of the mechanism await further clarification. For example, it is an open issue whether neural phase entrainment will be observed when the cue and subsequent speech stimuli are less regular and predictable. Note that the regular cues in our study provided optimal matching conditions with speech as they closely followed the melodic contour and tempo of the sentences. They were exactly timed with the speech stimuli such that the metrical pattern of speech directly resumed at expected times derived from the metrical pattern of the cue. Future investigations should clarify whether different versions of the cue (e.g., larger differences in pitch, tempo, or timing) would yield similar or attenuated neural entrainment. So far, behavioral results using auditory cues matching conversational (i.e., rhythmically less regular) speech excerpts point toward benefits for verbal processing (Cason, Astésano, et al., 2015). On the other hand, these and other results (e.g., Falk & Dalla Bella, 2016) revealed that benefits for speech processing can be boosted when motor activity is combined with the auditory cue. These findings point toward accounts of neural entrainment to speech and music that underscore the role of the motor system in generating temporal predictions (Chemin, Mouraux, & Nozaradan, 2014; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Chen, Penhune, & Zatorre, 2008; see Morillon, Hackett, Kajikawa, & Schroeder, 2015, for a review). The theta range in neural phase entrainment was often interpreted as relating to sensorimotor cycles in syllable production, such as jaw kinematics, which by close connections between auditory and motor systems would become relevant for speech perception (Fujioka et al., 2015; Peelle & Davis, 2012; Kotz & Schwartz, 2010; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985). Involvement of the motor cortex is also suggested in our study by higher activations found in the precentral and middle frontal gyri during speech processing (both at 1.65 and 5 Hz) after regular cue presentation.

Conclusions

The results of this study provide fresh insights into the mechanisms underlying neural entrainment and the link between music and speech rhythm processing. They offer an explanation of how benefits of rhythmic nonverbal

cueing on subsequent speech processing can be generated via neural phase entrainment. These issues may be of particular relevance for future research into therapeutical approaches using musical stimulation in speech rehabilitation such as musical training in dyslexia (Flaugnacco et al., 2015; Overly, 2003) or rhythmic cueing in hearing impairment (Cason, Hidalgo, et al., 2015; Thaut, McIntosh, & Hoemberg, 2014).

Acknowledgments

We thank Deirdre Bolger, Virginie Epting, Patrick Marquis, and Chloé Volpi-Moncorger for help with stimulus recording and data analyses. This work was supported by the European Union Seventh Framework Program (FP7-PEOPLE-2012-IEF, No. 327586 to S. F.), the Brain and Language Research Institute (BLRI, ANR-11-LABX-0036 to S. F. and D. S.), and the LMUexcellent program within the framework of the German excellence initiative (to S. F.).

Reprint requests should be sent to Simone Falk, Laboratoire Phonétique et Phonologie, Université Sorbonne Nouvelle, Paris-3, 19 Rue des Bernardins, 75005 Paris, France, or via e-mail: simone.falk@univ-paris3.fr, simone.falk@germanistik.uni-muenchen.de.

REFERENCES

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 13367–13372.
- Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*, 390–398.
- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*, 46–63.
- Bédard, C., Kroeger, H., & Destexhe, A. (2006). Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Physical Review Letters*, *97*, 118102.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*, 341–345.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The “ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, *14*, 362–366.
- Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*, 300–309.
- Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50.
- Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology*, *29*, 102–107.
- Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*, 2652–2658.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436.
- Chemin, B., Mouraux, A., & Nozaradan, S. (2014). Body movement selectively shapes the neural representation of musical rhythms. *Psychological Science*, *25*, 2147–2159.

- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*, 2844–2854.
- Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience*, *33*, 4002–4010.
- Cummins, F. (2015). Rhythm and speech. In M. A. Redford (Ed.), *Blackwell handbooks in linguistics. The handbook of speech production* (1st ed., pp. 158–177). Hoboken, NY: Wiley.
- Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.
- Ding, N., & Simon, J. Z. (2013). Power and phase properties of oscillatory neural responses in the presence of background activity. *Journal of Computational Neuroscience*, *34*, 337–343.
- Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311.
- Doelling, K. B., Arnal, L. H., Ghitz, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768.
- Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., & Jäncke, L. (2014). Music and language expertise influence the categorization of speech and musical sounds: Behavioral and electrophysiological measurements. *Journal of Cognitive Neuroscience*, *26*, 2356–2369.
- Falk, S., & Dalla Bella, S. (2016). It is better when expected: Aligning verbal and motor rhythms enhances verbal processing. *Language, Cognition, and Neuroscience*, *31*, 699–708.
- Falk, S., Volpi-Moncorger, C., & Dalla Bella, S. (2017). Auditory-motor rhythms and speech processing in French and German listener. *Frontiers in Psychology*, *8*, 395.
- Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One*, *10*, e0138715.
- François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*, 2038–2043.
- François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research*, *308*, 122–128.
- Frey, J. N., Ruhnau, P., & Weisz, N. (2015). Not so different after all: The same oscillatory processes support different types of attention. *Brain Research*, *1626*, 183–197.
- Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *Journal of Neuroscience*, *35*, 15187–15198.
- Fujioka, T., Zendel, B. R., & Ross, B. (2010). Endogenous neuromagnetic activity for mental hierarchy of timing. *Journal of Neuroscience*, *30*, 3458–3466.
- Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*, 1127–1134.
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*, e1001752.
- Henry, M. J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 14935–14940.
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*, 486–502.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*, 3019–3025.
- Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience*, *27*, 521–538.
- Jones, M. R. (2009). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford library of psychology. The Oxford handbook of music psychology* (pp. 81–92). Oxford, NY: Oxford University Press.
- Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 209–242). Boston: Kluwer.
- Kelso, J. A., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *The Journal of the Acoustical Society of America*, *77*, 266–280.
- Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, *1337*, 62–68.
- Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*, 392–399.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*, 599–605.
- Kraus, N., Hornickel, J., Strait, D. L., Slater, J., & Thompson, E. (2014). Engagement in community music classes sparks neuroplasticity and language development in children from disadvantaged backgrounds. *Frontiers in Psychology*, *5*, 1403.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*, 712–723.
- Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology*, *31*, 230–238.
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*, 10234–10240.

- Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*, 17572–17581.
- Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497–505.
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*, 1649–1653.
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, *24*(Suppl. D), 5–12.
- Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*, 1378–1387.
- Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization. A universal concept in nonlinear sciences*. New York: Cambridge University Press.
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., et al. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology*, *27*, 121–131.
- Riecke, L., Sack, A. T., & Schroeder, C. E. (2015). Endogenous delta/theta sound-brain phase entrainment accelerates the buildup of auditory streaming. *Current Biology*, *25*, 3196–3201.
- Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One*, *8*, e56000.
- Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*, 232–244.
- Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Decomposing rhythm processing: Electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological Research*, *75*, 95–106.
- Schlag, G. (2015). Musicians and music making as a model for the study of brain plasticity. *Progress in Brain Research*, *217*, 37–55.
- Schmidt-Kassow, M., & Kotz, S. A. (2009). Attention and perceptual regularity in speech. *NeuroReport*, *20*, 1643–1647.
- Schön, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: Perspectives for language rehabilitation. *Annals of the New York Academy of Sciences*, *1337*, 32–39.
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*, 9–18.
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*, 172–176.
- Schwartz, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology*, *87*, 146–151.
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience*, *30*, 13578–13585.
- Straut, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*, 191–201.
- Thaut, M. H., McIntosh, G. C., & Hoemberg, V. (2014). Neurobiological foundations of neurologic music therapy: Rhythmic entrainment and the motor system. *Frontiers in Psychology*, *5*, 1185.
- Tierney, A., & Kraus, N. (2015). Neural entrainment to the rhythmic structure of music. *Journal of Cognitive Neuroscience*, *27*, 400–408.
- Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, *34*, 343–371.
- Zoefel, B., & VanRullen, R. (2015a). The role of high-level processes for oscillatory phase entrainment to speech sound. *Frontiers in Human Neuroscience*, *9*, 651.
- Zoefel, B., & VanRullen, R. (2015b). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *Journal of Neuroscience*, *35*, 1954–1964.
- Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, *124*, 16–23.