

Western University

Scholarship@Western

Brain and Mind Institute Researchers'
Publications

Brain and Mind Institute

7-1-2020

Music as a scaffold for listening to speech: Better neural phase-locking to song than speech

Christina M. Vanden Bosch der Nederlanden
The University of Western Ontario

Marc F. Joanisse
The University of Western Ontario

Jessica A. Grahn
The University of Western Ontario

Follow this and additional works at: <https://ir.lib.uwo.ca/brainpub>

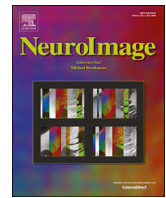


Part of the [Neurosciences Commons](#), and the [Psychology Commons](#)

Citation of this paper:

Vanden Bosch der Nederlanden, Christina M.; Joanisse, Marc F.; and Grahn, Jessica A., "Music as a scaffold for listening to speech: Better neural phase-locking to song than speech" (2020). *Brain and Mind Institute Researchers' Publications*. 506.

<https://ir.lib.uwo.ca/brainpub/506>



Music as a scaffold for listening to speech: Better neural phase-locking to song than speech



Christina M. Vanden Bosch der Nederlanden^{*}, Marc F. Joanisse, Jessica A. Grahn

Department of Psychology & Brain and Mind Institute, The University of Western Ontario, Canada

ARTICLE INFO

Keywords:

Music
Language
Rhythm
Entrainment
Phase-locking
Theta

ABSTRACT

Neural activity synchronizes with the rhythmic input of many environmental signals, but the capacity of neural activity to entrain to the slow rhythms of speech is particularly important for successful communication. Compared to speech, song has greater rhythmic regularity, a more stable fundamental frequency, discrete pitch movements, and a metrical structure, this may provide a temporal framework that helps listeners neurally track information better than the rhythmically irregular rhythms of speech. The current study used EEG to examine whether entrainment to the syllable rate of linguistic utterances, as indexed by cerebro-acoustic phase coherence, was greater when listeners heard sung than spoken sentences. We assessed listeners phase-locking in both easy (no time compression) and hard (50% time-compression) utterance conditions. Adults phase-locked equally well to speech and song in the easy listening condition. However, in the time-compressed condition, phase-locking was greater for sung than spoken utterances in the theta band (3.67–5 Hz). Thus, the musical temporal and spectral characteristics of song related to better phase-locking to the slow phrasal and syllable information (4–7 Hz) in the speech stream. These results highlight the possibility of using song as a tool for improving speech processing in individuals with language processing deficits, such as dyslexia.

1. Introduction

The human brain adeptly tracks temporal regularities in speech (e.g., Ahissar et al., 2001; Ahissar and Ahissar, 2005; Luo and Poeppel, 2007; Ghitza et al., 2012; Peelle et al., 2013; Keitel et al., 2018). Particularly, the degree of entrainment, or the synchronization of neural activity with stimulus input, at slow delta (1–4 Hz) and theta (4–7 Hz) frequencies, is related to an individual's ability to understand speech (e.g., Peelle et al., 2013; Doelling et al., 2014; Park et al., 2015). Oscillatory dynamics may be particularly important for helping listeners segment the continuous speech stream into multiple timescales that correspond to linguistic units in speech, like phonemes (gamma, 30–50 Hz), words/syllables (theta, 4–7 Hz), or phrases (delta, 1–4 Hz; Poeppel, 2003; Ghitza, 2011; Giraud and Poeppel, 2012). The delta and theta frequency bands appear particularly important for syllable segmentation in language listening. In the amplitude envelope of a speech stream, the prominent “edges” of the envelope, related to syllable onsets, help to reset the phase of ongoing oscillations (Luo and Poeppel, 2012). Acoustic edges are related to peaks in amplitude within the speech stream, in particular, edges are rapid increases in amplitude from the syllable onset to the peak of the amplitude envelope (Oganian and Chang, 2019). Neural phase resetting to

coincide with acoustic edges is thought to improve language segmentation as well as the parsing of temporal fine structure happening on much faster timescales. In particular, the phase reset at these acoustic edges aligns periods of high neuronal excitability to coincide with important segments of the speech stream, such as formant transitions or voice onset times (Doelling et al., 2014). Thus, theta frequency oscillations seem particularly important for neural processing of speech and intelligibility, a link that has already been shown behaviorally (Ghitza and Greenberg, 2009).

The relationship between neural speech tracking and intelligibility is also supported by studies that have found poor speech envelope tracking for individuals with reading deficits both in the theta (Abrams et al., 2009; Power, Mead, Barnes, & Goswami, 2012, 2013; Power et al., 2016) and gamma (Lehongre et al., 2011; Lehongre et al., 2013) bands. Clinically, improving neural entrainment to a stimulus may have complementary effects on intelligibility and perhaps comprehension. There are at least two ways that enhanced entrainment could be accomplished without external stimulation (i.e., tACS or tDCS, Zoefel and Davis, 2017), 1) through improved top-down attention to the stimulus, or 2) by altering stimulus features to create a clearer signal for the brain to track. In the current study, we examined neural tracking of the amplitude envelope in

^{*} Corresponding author. The Brain and Mind Institute, The University of Western Ontario, 1151 Richmond St, London, Ontario, Canada, N6A 3K7.
E-mail address: cdernede@uwo.ca (C.M. Vanden Bosch der Nederlanden).

sung versus spoken sentences, to determine whether the acoustic features and structure of music would induce better tracking for sung compared to spoken materials.

Music and language share many similarities that allow for direct comparisons. Both speech and song unfold in time and are composed of discrete units that subscribe to hierarchical rules (Patel, 2003; Jackendoff, 2009). However, music and language temporally unfold in different ways (Patel and Daniele, 2003). The rhythms of song are dictated by note durations, typically related by integer multiples, which allows simple (e.g., 4:1, 2:1, or 3:1) integer-ratios to describe how the smaller temporal units fit into larger portions of a musical phrase. The durations in speech are not related by simple integer-ratios, which increases the durational contrast between long and short syllables (Grabe and Low, 2002) and leads to a more irregular rhythmic structure than song.

Even though speech and song both rely on fundamental frequency (F0) information to convey meaning, F0 in song is more stable and moves discretely from one perceived pitch to the next, while the F0 of speech glides continuously from syllable to syllable (Zatorre and Baum, 2012; Vanden Bosch der Nederlanden et al., 2015a). This F0 stability may make sung syllables more stable perceptual targets, and the discrete changes in pitch may more clearly delineate syllable onsets. Differences in the way that F0 and syllable onsets in speech are processed in the brain may also lead to hemispheric differences in the strength of phase-locking for the left and right hemispheres (Zatorre and Baum, 2012). The right hemisphere seems to be particularly well-suited for extracting features of the auditory stream that require longer timescales (Giraud et al., 2007). Consistent with this, multiple studies show greater activation in the right auditory areas for steady state pitches or vowels compared to rapid formant transitions common in speech, especially at syllable onsets (e.g., Joannisse and Gati, 2003; Zatorre et al., 1994). Together, temporal and spectral differences between speech and song may highlight syllable level information in sung versus spoken speech streams, with greater phase-locking in the right hemisphere for song compared to speech. Furthermore, other factors such as melodic and harmonic knowledge in song may help listeners better predict the onset of the next syllable in a continuous utterance.

Regularity can benefit behavioral and neural performance in speech through rhythmic priming. In priming studies, listeners are given anywhere from 3 to 30 s of either very regular music or music that matches the rhythmic stress patterns of the upcoming speech stream. This listening period improves syntactic (Przybylski et al., 2013) and speech perception (Cason and Schön, 2012; Cason et al., 2015) for spoken utterances compared to priming with an irregular sequence. Priming with rhythmic scaffolding that matches the upcoming speech sequence not only improves behavior, but may even improve neural entrainment to the stimulus (Falk et al., 2017). Given the clear rhythmic context sung lyrics offer over spoken text, these studies highlight the potential for enhanced entrainment for sung compared to spoken utterances.

Most cultures have music, and people engage with song in both situationally unstructured or structured ways, such as singing with their children or attending a musical. The benefit of studying phase-locking to song is that songs tap into the rich structure of musical knowledge learned through experience, whereas artificially regularized utterances spoken to a metronome or recited with more stable fundamental frequencies do not. The goal of the current study is to examine whether phase-locking to language can be enhanced by making speech information in the stimulus more salient through song, and, in doing so, guiding attention to relevant events in the speech stream. Therefore, it is important to use a natural stimulus in which the musical structure highlights syllable information. Song is also commonly used in the language learning classroom, which makes it a good candidate for language remediation.

Although several studies examine neural entrainment to rhythmic stimuli (e.g., Nozaradan et al., 2011, 2012; Fujioka et al., 2009, 2012; Henry and Obleser, 2012), only a few studies have demonstrated entrainment to natural music. In piano music, phase-locking to the note

rate is greatest in the delta and theta band, similar to the syllable rate in speech (Doelling and Poeppel, 2015). Moreover, greater neural response amplitude is found at the beat frequency of unfamiliar compared to well-known piano pieces, demonstrating neural activity is modulated at beat-related frequencies for natural music (Meltzer et al., 2015). Entrainment also occurs for pop songs (Tierney and Kraus, 2015). A recent study compared rhythmically similar spoken utterances and piano melodies (Harding et al., 2019), showing greater cortical tracking of rhythmic information for music compared to speech. However, greater entrainment was significantly related to musical expertise and no intermediate effect was observed for non-musicians, suggesting that musical training predicted greater phase-locking, not the musical structure itself. Of course, both these factors could have an effect, or interacting effects, on neural entrainment. To our knowledge, there are no studies that examine entrainment to vocal music, and also none that directly compare vocal music and language. Directly comparing entrainment to the same utterance when spoken and sung offers the unique opportunity to observe whether musical features, like rhythmic regularity, melodic contour, or other temporal or spectral factors illustrated in Table 1, aid phase-locking, and perhaps especially in difficult listening settings.

In our daily lives, speech is not presented in clear listening conditions, and factors such as speech-in-noise or rapidly presented speech can make it difficult to extract relevant information from the speech stream. In the lab, degraded speech has illustrated how poor readers show deficits in comprehension and discrimination of time-compressed speech (Abrams et al., 2009; Ahissar et al., 2001; Watson et al., 1990). Time-compressed sentences can be difficult even for typically developing readers, especially when sentences are compressed to 0.5 times their original rate (e.g., Abrams et al., 2008), but comprehension remains quite good above 50%-time-compression for typically developing readers (Beasley et al., 1980). In order to examine whether song helps listeners phase-lock to syllable information in speech, we decided to include both easy and difficult listening conditions to determine whether the potential for song to highlight syllable-level information would be particularly beneficial when speech is time-compressed and difficult to track.

Given that we were hoping to observe differences based on the subtle acoustic differences that occur in sung versus spoken lyrics (Zatorre and Gandour, 2008; Vanden Bosch der Nederlanden et al., 2015a; Vanden Bosch der Nederlanden et al., 2015b), we wanted to ensure that the differences were unique to the differences between music and language, and not the result of other features that also co-vary regardless of the domain. For instance, speech and song can differ in their average F0, with higher F0 recorded for song than speech, but speech and song can also be produced at a variety of pitch heights. Pitch height has been associated with arousal (Jaquet et al., 2014; Huron, 2015), which may increase attention. In addition, lower frequency sequences elicit greater entrainment to beat or meter related frequencies than higher frequency sequences (Lenc et al., 2018). The peak modulation rate, or rate of change over time, is slower in music than in language (Ding et al., 2017). This slower modulation rate indicates that the durations of sung syllables are

Table 1

Three temporal and three spectral characteristics of the uncompressed spoken and sung utterances. Despite similar syllable lengths, song had more variable durations between onsets and shorter onset durations (time to amplitude peak). Although the average F0 was similar, song had more stable F0 over the length of each syllable and had great harmonicity than speech. Abbreviations: ms = milliseconds, Hz = Hertz, St = semitones, dB = decibels, a.u. = arbitrary units. Mean (standard deviation).

Acoustic Features	Song	Speech
Syllable Duration (ms)	258 (4)	258 (3)
Onset-to-Onset Variability (a.u.)	.46 (.2)	.44 (.2)
Syllable Onset Duration (ms)	51 (7)	67 (15)
F0 (Hz)	129.6 (10)	130.1 (10)
F0 Instability (St)	.7 (.2)	1.3 (.5)
Harmonicity (dB)	11.6 (1.2)	9.3 (.85)

longer. This difference in modulation rate could also mean that the total durations of songs may be longer than in an otherwise identical spoken utterance. Slower utterances may allow for enhanced entrainment, as amplitude envelope fluctuations, which are thought to be the main driver of phase-locking to speech, may be clearer and more isolated in a slowly unfolding stimulus. It is also possible that faster or slower temporal rates for spoken and sung materials would lead to different patterns of lateralization in phase-locking, unrelated to the processing of spoken or sung information. Thus, average F0 and duration, at a minimum, must be controlled when comparing entrainment to speech and song. If there is a benefit to highlighting syllable onsets using musical features like meter, rhythmic regularity, discrete and stable pitch movements, then, after controlling for overall duration and average F0, we expect that sung exemplars will result in greater phase-locking values in the delta and theta bands. These bands correspond to the syllable rate of speech, so we predict greater phase-locking to song than speech in both the compressed and non-compressed conditions, at frequencies corresponding to the syllable rate. We anticipate a larger effect from highlighting syllable features in the more difficult, time-compressed conditions. As described above, this difficult condition may be uniquely suited to show a difference between conditions.

In the current study, we examined how individuals tracked temporal information in speech for sentences that were spoken and sung. We used the measure of cerebro-acoustic phase coherence¹ to estimate how well individuals track temporal information in speech. Listeners heard four matched sentences presented at a regular speaking rate (uncompressed) and a 50% time-compressed speaking rate (compressed). As described above, time-compressed speech significantly degrades comprehension of the stimulus (Ahissar et al., 2001) and previous studies have found that poor readers have particular difficulty understanding time-compressed speech (Watson et al., 1990; Abrams et al., 2009). Because the goal of the study is to examine whether sung stimuli alter phase-locking to language, with the potential to affect comprehension, we wanted to include a condition that was difficult for typically developing listeners and listeners who are known to have issues with language processing.

2. Methods

2.1. Participants

We recruited twenty-three right-handed adults (13 females) from the University's Undergraduate Psychology participant pool. Four additional participants were not included in the current sample because of experimenter error ($N = 1$), excessive movement during the EEG portion of the study ($N = 2$), or failing to complete the EEG session ($N = 1$). Participants ranged in age from 17 to 27 with a mean of 19.48 ($SD = 2.56$) years of age. All participants were either native English speakers ($N = 18$) or were fluent in English ($N = 3$, other first languages: Bengali, Farsi, and Hindi).² There was a range in musical training from 0 to 11 (Mean = 4.07; $SD = 3.24$) years, with only 3 individuals reporting that they still played regularly. On average, participants had normal IQ (WASI non-verbal IQ = 102.08, $SD = 11.44$), and normal phonological awareness (CTOPP = 103.13, $SD = 11.56$). All materials were approved by the University's Ethics Committee.

¹ interchangeably referred to throughout the text as coherence, phase coherence, phase-locking.

² It is important to note that bilinguals have exposure to different speech rhythms and have better rhythmic processing abilities than monolinguals (Roncaglia-Denisson, Palmer). As such, we performed these analyses without 3 non-native English speakers and results were nearly identical. As in the full analysis, the difference between song and speech was only found for compressed speech in the same frequency band (3.667–5.33 Hz).

2.2. Stimuli

Overtly spoken and sung utterances were recorded by having two male English speakers (American accent) speak and sing several Harvard Sentences (see Table 1; IEEE subcommittee, 1969). For sung utterances, the melody was provided for the singer. The melody was created by the first author to have a clear Western melodic structure while also having a speech-like contour. In fact, to create these stimuli, they were first spoken, then a melody was created from the spoken contour and moved to better fit the diatonic musical scale. Singers were only given a melodic representation of the sentence. Singers were encouraged to use the same overall pitch, duration, and contour of the sung utterance when they also spoke the sentence in their own way. These speech/song pairs are part of a larger corpus of matched pairs created for a previous study in our lab. Four matched pairs were selected for the current study and there were no differences in average pitch or total duration (Table 1). The same singer/speaker was selected for both the spoken and sung version of a single text. There were 4 texts used for speech and song, which read: "Both brothers wear the same size", "Clap as the curtain slides back", "Sew the button on the vest", and "The wide road shimmered" (to hear sample stimuli see <https://sites.google.com/view/christinavbdn/stimuli>). Listeners heard 100 repetitions of each stimulus for a total of 400 sung and 400 spoken trials. These same utterances were then time-compressed using Adobe Audition and the same number of stimuli were presented for the compressed utterance condition. In total, listeners heard 1600 trials of 4 texts across spoken/sung and uncompressed/compressed conditions. The 16 stimuli ranged from 620 ms to 2000 ms, the compressed stimuli were 848 ms on average and the uncompressed stimuli were 1695 ms. Stimuli were presented in blocks (compressed and non-compressed presented separately) and the study lasted 1 h.

Stimulus descriptives revealed that, despite the average syllable duration and average pitch of spoken and sung utterances being nearly identical, there were differences between spoken and sung utterances in both temporal and spectral features. **Onset-to-onset variability** is a normalized measure of how variable the duration between syllable onsets were while taking into account the length of the syllable duration (onset-to-onset standard deviation/onset-to-onset average duration). There was more variability for song than speech in this measure, which could be reflective of the alternation of longer and shorter integer multiple related intervals in music compared to language. **Syllable onset duration** measured the time from minimum to maximum amplitude within pre-defined syllable duration windows. It took less time on average to reach the peak of the amplitude envelope in song compared to speech, that is, the onset, or rise time, is steeper for song than speech. **F0 instability** was the average of the standard deviation for F0 in semitones for each syllable (Vanden Bosch der Nederlanden et al., 2015b). This F0 instability measure gives a sense of whether listeners perceived the pitch to be largely stable over the course of a syllable, which was evident in our song stimuli, or whether the pitch moved throughout the syllable, which was evident in our speech stimuli. Finally, **harmonicity** is the ratio of the periodic information compared to noise in an acoustic signal measured in decibels, which provides a measure for pitch salience (Gygi et al., 2007; Vanden Bosch der Nederlanden et al., 2016). Therefore, song had more salient pitch contours than speech in the current stimulus set.

We measured the coherence of neural and acoustic envelope signals at frequency bands specifically related to prosodic, syllabic, and phonemic speech rates. To identify whether or where these speech-related frequencies occurred in our stimulus set, we obtained power spectra for all excerpts (see Fig. 1). Power spectra were obtained using the same process for extracting amplitude envelopes below (section 2.3), but without down-sampling and by doing a fast Fourier transform (FFT) with a rectangular window padded to make all stimuli 10 s. Clear peaks can be seen in the uncompressed utterances for what likely corresponds to prosodic (2–3.7 Hz), syllabic (5.1–7.1 Hz), and perhaps phonemic (10–13.2 Hz) rates. These rates are very similar to the rates found in previous studies (Keitel et al., 2018). For the difficult listening condition,

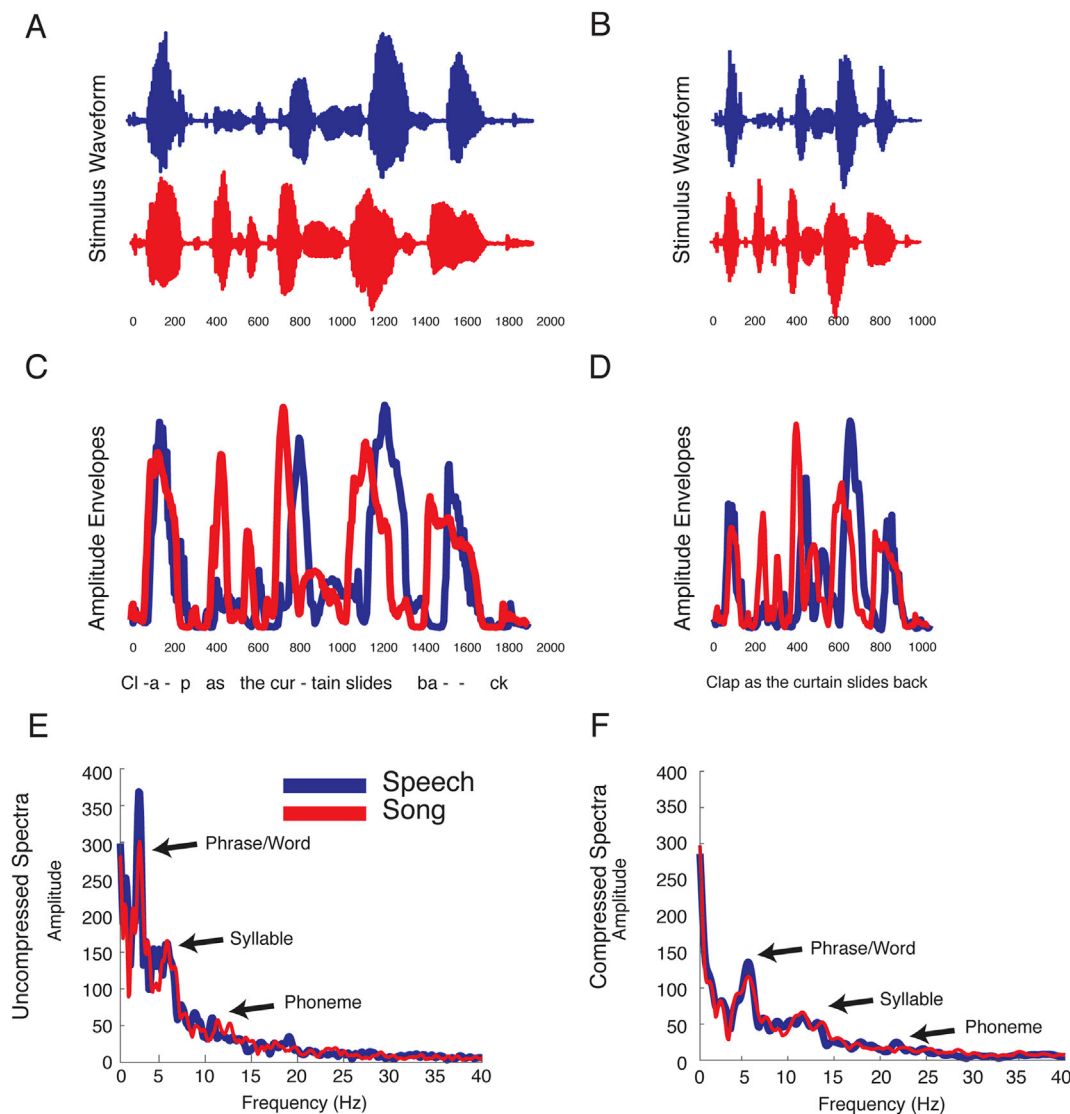


Fig. 1. Schematic of stimulus waveforms for A) uncompressed and B) compressed versions of the same utterance, C & D) Hilbert transformed amplitude envelopes of the same sentences exhibited in A & B, and amplitude spectra of all E) uncompressed and F) compressed stimuli, demonstrating that there are peaks in frequency roughly corresponding to the phrasal, syllable, and perhaps phoneme level. See [Supplemental Fig. 1](#) for example of another sentence and [Supplementary Fig. 2](#) for detrended stimuli spectra illustrating more pronounced peaks.

listeners also heard 50% time-compressed versions of the spoken and sung stimuli. Importantly, 50% time-compressed speech was still intelligible, resulting in 89% accuracy across nine naïve listeners who typed out the text. The compressed utterances show peaks at double the frequency of uncompressed utterances (4.1–6.7; 10.1–14.4; 21–22.8) as a result of 50% time-compression.

2.3. EEG recording and preprocessing

Continuous EEG data was recorded using BioSemi ActiveTwo system. EEG data was recorded from 32 active scalp electrodes (10/20 grid system), a Common Mode Sense active electrode and Driven Right Leg passive electrode as ground, and six additional active flat electrodes placed on both mastoids, outer canthi of each eye, and inferior and superior orbit of the left eye. Data were recorded in the frequency range of 0.1–100 Hz at a 512 Hz sampling rate (104 Hz Bandwidth), with voltage offsets below ± 20 mV.

All offline data was processed using EEGLAB (Delorme and Makeig, 2004), ERPLAB (Markley et al., 2011), and FieldTrip Software (Oostenveld et al., 2011). Data were re-referenced to the average of left and

right mastoids. EEG data were high-pass filtered at 1 Hz (order = 1690) and low-pass filtered at 40 Hz (order = 170) both using a Hamming windowed sinc FIR (finite impulse response) filter. Eye blinks, horizontal eye movements, noisy electrodes, and other physiological artefacts (heartbeat, muscle) were removed using the *fastica* independent components analysis (ICA) toolbox in MATLAB. Offline EEG data was down-sampled to 200 Hz. Continuous EEG data was epoched into 3000 ms bins, which included a 900 ms pre-stimulus period. Epochs were baseline corrected using 900 ms pre-stimulus intervals and noisy epochs were rejected using $\pm .75$ μ V threshold. Thus, 3000 ms epochs were submitted to the time-frequency analyses. All participants had less than 10% of trials rejected, with an average of 1532 trials (95% of trials) left in the analysis per person. To obtain the amplitude envelopes of the acoustic stimuli for later cerebro-acoustic phase coherence calculations, the frequency spectrum of each utterance was separated into 10 bands of equal width along the basilar membrane using Liberman's cochlear frequency map from 100 to 10000 Hz (Liberman, 1982) using a custom Matlab script. The Hilbert transform was used to extract the amplitude envelope of each band, which were then low-pass filtered at 40 Hz and High-pass filtered at 1 Hz. The average of the 10 filtered bands was then

down-sampled to 200 Hz to match the resolution of the pre-processed EEG data. Stimuli were zero-padded to equal the 3000 ms EEG epochs. This epoch duration is sufficient for estimating syllable-level information in the speech stream as our Nyquist rate, the slowest frequency we can estimate with accuracy, was 0.666 Hz, below our 1 Hz filter.

Individual EEG epochs and stimulus epoch were then converted to time-frequency domain using FieldTrip's Wavelet Convolution using Morlet wavelets that were linearly spaced from 3 to 7 cycles over 1–40 Hz with a 0.25 Hz frequency resolution and 5 ms temporal resolution. All of the stimulus amplitude envelopes were submitted to the same Wavelet Convolution and the resulting complex values from EEG data and stimuli were used to estimate cerebro-acoustic phase coherence (see Harding et al., 2019 for formula). This measure of phase coherence was calculated by comparing how aligned the phase of the pre-processed EEG signal is compared to the amplitude envelope of the corresponding utterance. Cerebro-acoustic phase coherence ranges from 0 (no coherence) to 1 (complete coherence). Before wavelet convolution, we introduced a 150 ms time lag between stimulus and EEG data, because this time lag resulted in the peak cerebro-acoustic phase coherence in the theta band on average. We determined the 150 ms time lag by performing analyses for a subset of participants with 5–300 ms time lags in 5 ms time-lag increments. Cerebro-acoustic coherence is compared here using only frontal and central electrodes (Fp1/2, AF3/4, F7/7, F3/4, FC1/2, FC5/6, C3/4, CP1/2, Fz, Cz), which is where the auditory response is typically observed (Luck, 2014). We averaged coherence over these fronto-central electrodes after calculating coherence for all electrodes individually. Cerebro-acoustic coherence was averaged over time ranges corresponding to the length of the stimulus presented. We only considered coherence for these time ranges as averaging over the entire range would simply lower overall coherence values due to the zero-padding of the stimulus. To estimate a null distribution for cerebro-acoustic phase coherence in our study, we used a bootstrap technique in which we calculated phase coherence for 100 random pairings of EEG data and stimulus envelopes, separately within the uncompressed and compressed conditions. Data are then reported as z-score transforms comparing true and random pairings.

2.4. Procedure

After participants provided consent, they were fitted with EEG caps and picked out a movie ("Wall-E" or "Up") to watch with audio playing during the EEG study. Participants wore ER-1 insert ear Etymotic headphones that attenuate the room noise by 30+ dB and were told to ignore the utterances being presented in their ears and to just watch the movie. Spoken and sung utterances were presented using a custom E-Prime 2.0 script. Stimuli were presented at a 1050 ms inter stimulus interval with a 500 ms jitter (800–1300 ms ISI) sampled from a uniform distribution in 1 ms increments. Listeners heard utterances in 4 blocks, 2 compressed and 2 uncompressed, and were allowed to take breaks between blocks. Blocks were counterbalanced across participants and spoken and sung utterances were randomly presented within each block. The EEG listening portion of the study lasted approximately 1 h. Participants also completed a number of behavioural tasks assessing nonverbal IQ, phonological awareness, reading, executive function, and beat perception and production skills. Results from these tasks are not reported here and were collected for a separate study.

2.5. Analysis

Neural data was examined across all frequencies (1–40 Hz) by creating a z-score transformation of cerebro-acoustic phase coherence for uncompressed speech and song and random permutations of stimulus and EEG pairings (see Methods, and see Supplemental Fig. 3 for non-transformed data). Separate z-score transformed coherence values (henceforth, coherence values) for speech and song were compared to chance using a two-tailed one-sample *t*-test across all frequencies, which

we corrected for multiple comparisons using False Discovery Rate (FDR). Average coherence values were taken over the left (Fp1, AF3, F7, F3, FC1, FC5, C3, CP1) and right (Fp2, AF4, F8, F4, FC2, FC6, C4, CP2) hemispheres and entered into a 2 Hemisphere (Left, Right) x 2 Utterance (Speech, Song) ANOVA.

3. Results

Contiguous frequency regions of significance were found for speech from 3.67 to 8.67 Hz and for song from 4.0 to 10.33 Hz (Fig. 2A), suggesting phase-locking to syllabic information in the current utterances. There were no significant differences between coherence values for speech and song after correcting for multiple comparisons using the FDR correction (corrected *p* values all greater than 0.128). Visual inspection of the topographies (Fig. 2A) suggested that there may have been differences in the pattern of lateralization of phase coherence for speech compared to song. There was no statistically significant difference in lateralization for uncompressed utterances, $F(1,22) = .206, p = .655, \eta_p^2 = .009$, and no difference between speech and song, $F(1, 22) = .006, p = .940, \eta_p^2 < 0.001$.

For compressed utterances (Fig. 2B), there was again significant phase coherence compared to chance for spoken (1.33–1.67 Hz and 4.33–8.33 Hz) and sung (3.67–7.33 Hz) utterances, with one significant region for speech (11.33–15.0 Hz) having significantly less phase coherence compared to chance. This finding illustrates that random pairings of speech and EEG had more coherence than true pairings. This frequency band corresponded to the compressed syllable rate of speech (see Fig. 1D), so it is possible that participants were actively inhibiting entrainment to syllables, which were too fast to process, instead chunking utterances into larger units at the phrasal level. There was also a significant difference when comparing z-score transformed phase-coherence values for speech and song from 3.67 to 5 Hz, with greater coherence for song than speech in this frequency region, which corresponds to slower phrasal or syllabic information in the spoken or sung streams. As is clear in Fig. 3, there was a difference between speech and song, $F(1,22) = 7.374, p = .012, \eta_p^2 = .251$, and an interaction between hemisphere and utterance type, $F(1,22) = 4.925, p = .037, \eta_p^2 = .183$. Simple effects revealed that there was a difference between speech and song in the right hemisphere, $F(1,22) = 11.125, p = .003, \eta_p^2 = .336$, but not the left hemisphere, $F(1,22) = 3.952, p = .059, \eta_p^2 = .152$.

4. Discussion

This study was designed to determine whether the musical characteristics of song, including rhythmic regularity and discrete and stable F0, would be related to better neural phase-locking to speech when it was sung compared to spoken. Using cerebro-acoustic phase coherence to index entrainment, we found that phase-locking to the slowly unfolding syllable information in compressed speech is greater when an utterance is sung than spoken. This finding was only evident for the more difficult listening condition, when speech and song were time-compressed by fifty percent. A key part of the design was the careful control of overall F0 and duration to determine whether the different manner in which the utterance was presented, and not overall slower syllable rate, led to differences in phase-locking. The careful matching of acoustic features for spoken and sung utterances is evident in the indistinguishable patterns of phase-locking to uncompressed speech. Only when the stimuli were more difficult to comprehend are the musical features of song related to increased phase coherence to phrasal linguistic information. Our results are the first to directly compare phase-locking to speech and song and examine the potential for perceptual advantages when listening to song compared to speech.

It is not entirely surprising that phase-locking differences were only observed for time-compressed utterances. Other studies looking at phase-locking to speech show similar findings, such that effects were more pronounced for difficult listening conditions compared to easy or

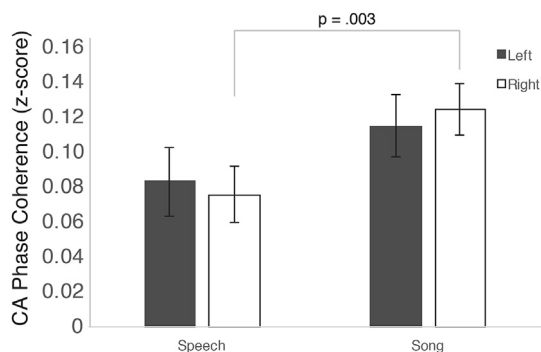
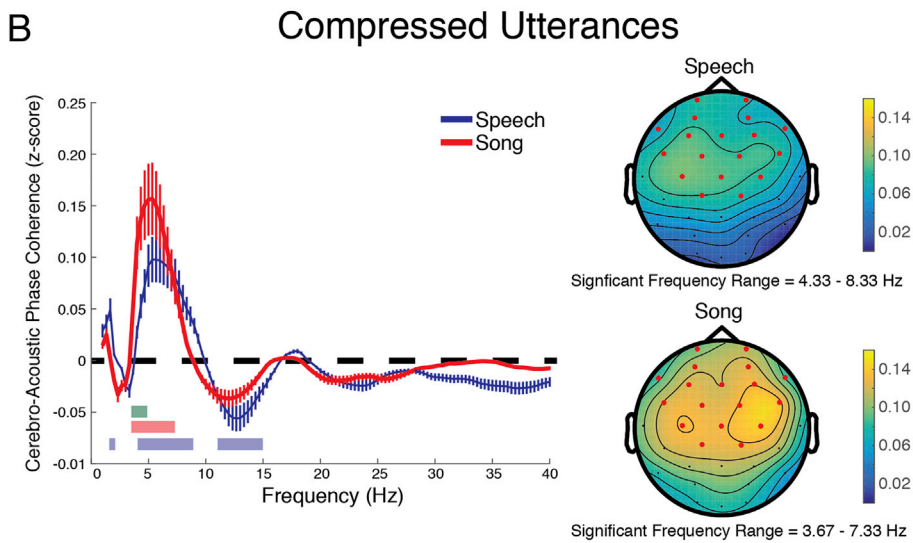
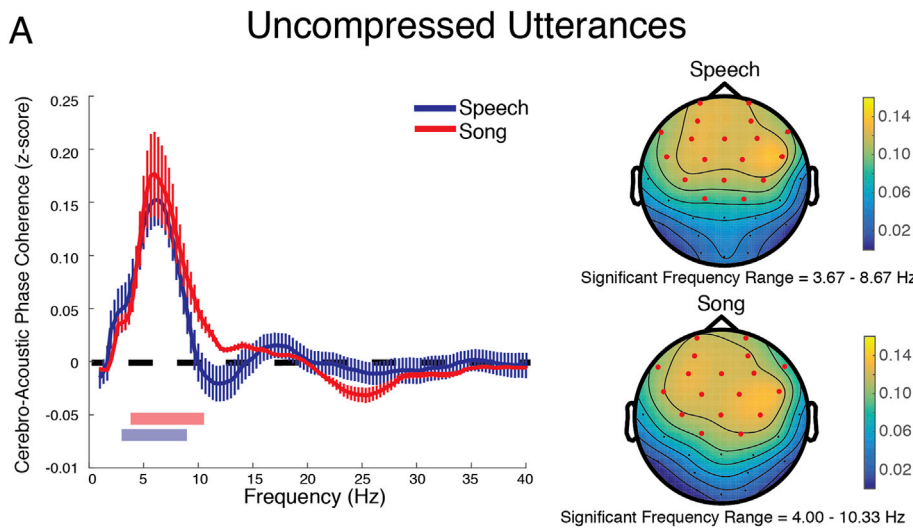


Fig. 3. Average cerebro-acoustic phase coherence for compressed speech and song stimuli, over all right and left fronto-central electrodes. Song had greater phase coherence in the right hemisphere than speech, but no difference was found between hemispheres for speech and song.

conversational listening conditions (Abrams et al., 2009). Time compressing speech obfuscates key features of the speech stream that are important for speech processing, including acoustic edges. It is possible that the melodic structure, steep (and perhaps clearer) acoustic edges, F0, and rhythmic structure combined may have uniquely benefitted

Fig. 2. Z-score transformed cerebro-acoustic phase coherence values for A) uncompressed and B) compressed utterances for all spoken and sung exemplars. Right panels show the topographies for phase coherence averaged over significance regions (shown in shaded bars at the bottom of each figure). Red dots highlight the location of the fronto-central electrodes used in the phase coherence analyses. Shaded bars indicate frequencies that are significantly different from chance (blue = speech and red = song) or from each other (green = song > speech).

phase-locking to song compared to speech. Finally, EEG is less sensitive to neural fluctuations than other tools like MEG, so future work will use MEG to examine whether there might be differences in uncompressed speech and song as well. We also used a limited number of sentences in the current study, which makes it possible that listeners became adapted to the spoken and sung versions presented here, leading to less robust phase-locking to both speech and song. It is also possible that listeners began to perceive spoken utterances transform to song, although pilot data from our lab suggests that the overtly spoken and sung utterances used here remain stable in their percept and do not transform from speech to song when it is repeated directly 10 times. Future work should use more spoken and sung sentences so that the differences between sung and spoken utterances can be examined without the potential interference for neural adaptation or speech-to-song transformations.

Qualitatively, time-compressed utterances appeared to have reduced amplitude overall compared to uncompressed utterances. Raw data (Supplemental Fig. 3) illustrate that compressed utterances had larger phase-locking values than uncompressed utterances, which means that the random pairings for the baseline resulted in similarly large phase-locking values for compressed utterances, thus reducing the overall z-score transformed data. This is consistent with more noise in the neural signal due to the difficulty of the condition. We also found a difference in the compressed speech condition where less phase coherence was observed than chance for speech in the higher frequency band of

11.33–15 Hz. It is possible that, as in [Ghitza and Greenberg \(2009\)](#), this rate was too fast to be intelligible and so listeners chunked speech into larger, phrasal units for ease of comprehension. More research is needed to determine whether listeners phase-lock to larger units of speech (lower frequencies) when it is time-compressed.

We also found greater phase coherence in right fronto-central electrodes for compressed song compared to compressed speech, but no difference between song and speech for the left electrodes. These findings are consistent with a wealth of research showing generally greater activation over the right hemisphere for music-related tasks ([Rogalsky et al., 2011](#)) compared to speech and language-related tasks. It is important to note that these findings did not show right lateralized responses within the song condition, that is, phase-locking within speech and within song showed statistically indistinguishable patterns of activation for the right and left auditory electrodes. Equal phase coherence for left and right fronto-central electrodes within each condition is surprising given previous studies that have shown left lateralized coherence for speech ([Peelle et al., 2013](#)) and right lateralized coherence for music ([Doelling and Poeppel, 2015](#)). The observed bilateral activation in the current study is in line with research suggesting that, at earlier levels of language processing, speech is largely processed bilaterally ([Hickok and Poeppel, 2007](#)). The addition of sung information may have led to greater coherence overall, with a particular increase in the right hemisphere compared to speech, which may have been related to the additional weight of melodic cues reliant on right hemisphere dominant timescales (see [Poeppel, 2003](#)).

Using a natural musical stimulus, we observed better phase-locking to sung than spoken time-compressed utterances. Although we were intentional about the use of a natural musical stimulus that is both richly structured and is already used in the language learning classroom, the current study was not designed to examine which features or characteristics of song were related to the observed increase in phase-locking for sung stimuli. It is possible that increased phase-locking was related to faster time to peak amplitude at the onset of a syllable for song, creating acoustic edges that were easier to track than relatively slower changes in the amplitude envelope of speech. Indeed, electrocorticography studies show that the steeper the acoustic edge in the amplitude envelope, the greater the observed cortical response in the superior temporal gyrus ([Oganian and Chang, 2019](#)). Acoustic edge processing could have important implications for studying speech processing in developmental language disorders, as well. Children with dyslexia are worse at discriminating the steepness of acoustic edges in the amplitude envelope of syllables, which are also known as rise times in speech ([Goswami, 2011](#)). For instance, the rise time for a canonical syllable /ba/ is around ~30 ms while the rise time for the syllable /wa/ is much longer at ~125 ms. Children with dyslexia are able to discriminate phonemes based on duration, but are worse at discriminating phonemes based on envelope cues and have particular difficulty discriminating protracted rise times, like /wa/ ([Goswami et al., 2011](#)). Faster rise times for sung utterances could be related to the presence of discrete movements in F0—instead of gliding from syllable to syllable as in speech, clearer syllable onsets may be present for each stable pitch and the singer must move quickly to the vowel to sing and hold the note.

Of course, attentional factors could drive phase-locking differences beyond the acoustic features described above. For instance, the presence of a beat in music allows listeners to focus their attention on relevant moments in time ([Large and Jones, 1999](#)). Perhaps the presence of a metrical framework or beat enables listeners to better predict when a syllable will arrive, improving phase-locking. Similarly, as all sung utterances were sung in accordance with the Western diatonic musical scale, listeners' melodic knowledge could provide additional top-down information as to whether or when another syllable would arrive. That is, enculturated listeners would not expect a melody to end on a leading tone, but would rather wait for a resolution to the tonic or the dominant of the musical scale ([Krumhansl, 1991](#); [Margulis, 2005](#); [Pearce and Wiggins, 2012](#)). The nonlinear dynamics of intrinsic neuronal oscillations

can also explain how we attend to musical relationships, highlighting the pull of the listener's expectations toward more stable notes in the tonal hierarchy ([Large et al., 2016](#)). Thus, neurophysiological dynamics of listening to music may also contribute to better phase-locking to song than speech. Future research should characterize the contributions of top-down or bottom-up influences of music on entrainment to the delta and theta frequencies in speech, and examine whether a combination of these factors are important for bringing about greater phase-locking to song than speech.

These findings provide some of the first evidence that, under difficult listening conditions, the brain can track low-frequency information in the speech stream better when that utterance is sung compared to spoken. Music and language are both important tools for human communication. The similarities between domains make song an ideal tool for the systematic investigation of how language and music are processed in the brain, and the playfulness of song may make it ideal for interventions in language processing disorders. Of particular interest is the potential benefit to individuals who have shown deficits in phase-locking to speech, such as those with dyslexia (for review see [Goswami, 2011](#)). There is widespread evidence for temporal processing deficits to speech in children and adults with reading deficits ([Abrams et al., 2009](#); [Power et al., 2013](#); [Lehongre et al., 2011](#); [Woodruff Carr et al., 2014](#)). Evidence for this temporal deficit consistently indicates aberrant processing in the delta and theta bands, which is relevant given that our findings suggest a specific benefit in tracking speech information within this slow phrase/word-related frequency band. Further studies should examine how adults and children with dyslexia entrain to the slow syllable-related frequency bands when listening to speech and song. It is possible that individuals with dyslexia would show the canonical deficit in tracking low frequency information in speech, but may show statistically indistinguishable phase-locking to song if additional cues to syllable onsets bootstrap their ability to track speech syllables. Our findings are the first to demonstrate that the musical features of sung speech are related to better phase-locking to syllable level information in difficult listening conditions. These findings have the potential to guide interventions for facilitating speech processing in the real world, suggesting possible acoustic manipulations for in-ear assistive technologies or for aiding learning in the classroom.

Funding

This work was supported by Natural Sciences and Engineering Research Council of Canada awards to MFJ and JAG, and by a post-doctoral award in Cognitive Neuroscience from Western's Brain and Mind Institute to CMVBdN.

Declaration of competing interest

None.

CRediT authorship contribution statement

Christina M. Vanden Bosch der Nederlanden: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Project administration. **Marc F. Joannis:** Methodology, Resources, Writing - review & editing, Supervision, Conceptualization. **Jessica A. Grahn:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision.

Acknowledgments

Many thanks to Dr. Molly Henry for providing feedback and guidance on time-frequency analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2020.116767>.

References

- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *The Journal of Neuroscience* 28 (15), 3958–3965.
- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2009. Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* 29 (24), 7686–7693. <https://doi.org/10.1523/JNEUROSCI.5242-08.2009>.
- Ahissar, E., Ahissar, M., 2005. Processing of the temporal envelope of speech. In: *The Auditory Cortex a Synthesis of Human and Animal Research*, Peter Heil, Henning Scheich, Eike Budinger, and Reinhard Konig. Psychology Press, Routledge, USA, pp. 295–314.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. Unit. States Am.* 98 (23), 13367–13372. <https://doi.org/10.1073/pnas.201400998>.
- Beasley, D.S., Bratt, G.W., Rintelmann, W.F., 1980. Intelligibility of time-compressed sentential stimuli. *J. Speech Lang. Hear. Res.* 23 (4), 722–731.
- Cason, N., Schön, D., 2012. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50 (11), 2652–2658. <https://doi.org/10.1016/j.neuropsychologia.2012.07.018>.
- Cason, N., Astésano, C., Schön, D., 2015. Bridging music and speech rhythm: rhythmic priming and audio-motor training affect speech perception. *Acta Psychol.* 155, 43–50. <https://doi.org/10.1016/j.actpsy.2014.12.002>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., Poeppel, D., 2017. Temporal modulations in speech and music. *Neurosci. Biobehavioural Rev.* 81, 181–187. <https://doi.org/10.1016/j.neubiorev.2017.02.011>.
- Doelling, K.B., Poeppel, D., 2015. Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. U.S.A.* 112 (45), E6233–E6242. <https://doi.org/10.1073/pnas.1508431112>.
- Doelling, K., Arnal, L., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85 (2), 761–768. <https://doi.org/10.1016/j.neuroimage.2013.06.035>.
- Falk, S., Lanzilotti, C., Schön, D., 2017. Tuning neural phase entrainment to speech. *J. Cognit. Neurosci.* 29 (8), 1378–1389. https://doi.org/10.1162/jocn_a.01136.
- Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2009. Beta and gamma rhythms in human auditory cortex during musical beat processing. *Ann. N. Y. Acad. Sci.* 1169, 89–92. <https://doi.org/10.1111/j.1749-6632.2009.04779.x>.
- Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2012. Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32 (5), 1791–1802. <https://doi.org/10.1523/JNEUROSCI.4107-11.2012>.
- Ghitza, O., 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2, 130. <https://doi.org/10.3389/fpsyg.2011.00130>.
- Ghitza, O., Greenberg, S., 2009. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66 (1–2), 113–126. <https://doi.org/10.1159/000208934>.
- Ghitza, O., Giraud, A.L., Poeppel, D., 2012. Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front. Hum. Neurosci.* 6 (340), 1–4. <https://doi.org/10.3389/fnhum.2012.00340>.
- Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15 (4), 511–517. <https://doi.org/10.1038/nn.3063>.
- 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.
- Goswami, U., 2011. A temporal sampling framework for developmental dyslexia. *Trends Cognit. Sci.* 15 (1), 3–10. <https://doi.org/10.1016/j.tics.2010.10.001>.
- Goswami, U., Fosker, T., Huss, M., Mead, N., Szucs, D., 2011. Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. *Dev. Sci.* 14 (1), 34–43. <https://doi.org/10.1111/j.1467-7687.2010.00955.x>.
- Grabe, E., Low, E.L., 2002. Durational variability in speech and the rhythm class hypothesis. *Lab. Phonol.* 7, 515–546.
- Gygi, B., Kidd, G.R., Watson, C.S., 2007. Similarity and categorization of environmental sounds. *Percept. Psychophys.* 69 (6), 839–855.
- Harding, E.E., Sammler, D., Henry, M.J., Large, E.W., Kotz, S.A., 2019. Cortical tracking of rhythm in music and speech. *NeuroImage* 185, 96–101.
- Henry, M.J., Obleser, J., 2012. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. Unit. States Am.* 109 (49), 20095–20100. <https://doi.org/10.1073/pnas.1213390109>.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. <https://doi.org/10.1038/nrn2113>.
- Huron, D., 2015. Affect induction through musical sounds: an ethological perspective. *Phil. Trans. Biol. Sci.* 370 (1664) <https://doi.org/10.1098/rstb.2014.0098>, 20140098–20140098.
- Jackendoff, R., 2009. Music Percept. UME, 195–204. <https://doi.org/10.1525/MP.2009.26.3.195>, 26.
- Jaquet, L., Danuser, B., Gomez, P., 2014. Music and felt emotions: how systematic pitch level variations affect the experience of pleasantness and arousal. *Psychol. Music* 42 (1), 51–70. <https://doi.org/10.1177/0305735612456583>.
- Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage* 19, 64–79.
- Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biol.* 16 (3), e2004473 <https://doi.org/10.1371/journal.pbio.2004473>.
- Krumhansl, C.L., 1991. Music psychology: tonal structures in perception and memory. *Annu. Rev. Psychol.* 42, 277–303. <https://doi.org/10.1146/annurev.ps.42.020191.001425>.
- Large, E.W., Jones, M.R., 1999. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106 (1), 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>.
- Large, E.W., Kim, J.C., Flaig, N.K., Bharucha, C.J., Krumhansl, C.L., 2016. A neurodynamic account of musical tonality. *Music Percept.* 33 (3), 319–331.
- Lehongre, K., Morillon, B., Giraud, A.-L., Ramus, F., 2013. Impaired auditory sampling in dyslexia: further evidence from combined fMRI and EEG. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2013.00454>.
- Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., Giraud, A.-L., 2011. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* 72 (6), 1080–1090. <https://doi.org/10.1016/j.neuron.2011.11.002>.
- Lenc, T., Keller, P.E., Varlet, M., Nozaradan, S., 2018. Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (32), 8221–8226. <https://doi.org/10.1073/pnas.1801421115>.
- Liberman, A.M., 1982. On finding that speech is special. *Am. Psychol.* 37 (2), 148–167. <https://doi.org/10.1037/0003-066X.37.2.148>.
- Luck, S.J., 2014. *An Introduction to the Event-Related Potential Technique*, second ed. A Bradford Book: The MIT Press, Cambridge, Massachusetts.
- Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54 (6), 1001–1010. <https://doi.org/10.1016/j.neuron.2007.06.004>.
- Luo, H., Poeppel, D., 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3 (170), 1–10.
- Margulis, E.H., 2005. A model of melodic expectation. *Music Percept.* 22, 663–714.
- Markley, C., Luck, S.J., Lopez-Calderon, J., 2011. ERPLAB Toolbox 2.0 User's Manual. University of California, Sacramento, CA. Davis.
- Meltzer, B., Reichenbach, C.S., Braiman, C., Schiff, N.D., Hudspeth, A.J., Reichenbach, T., 2015. The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention. *Front. Hum. Neurosci.* 9, 436. <https://doi.org/10.3389/fnhum.2015.00436>.
- Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31 (28), 10234–10240. <https://doi.org/10.1523/JNEUROSCI.0411-11.2011>.
- Nozaradan, S., Peretz, I., Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32 (49), 17572–17581. <https://doi.org/10.1523/JNEUROSCI.3203-12.2012>.
- Oganian, Y., Chang, E.F., 2019. A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances* 5 (11). <https://doi.org/10.1126/sciadv.aay6279>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.*
- Park, H., Ince, R.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. *Curr. Biol.* 25 (12), 1649–1653. <https://doi.org/10.1016/j.cub.2015.04.049>.
- Patel, A.D., 2003. Rhythm in language and music: parallels and differences. *Ann. N. Y. Acad. Sci.* 999, 140–143. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14681127>.
- Patel, A.D., Daniele, J.R., 2003. Stress-timed vs. Syllable-timed music? A comment on Huron and Ollen (2003). *Music Percept.* 21 (2), 273–276. <https://doi.org/10.1525/mp.2003.21.2.273>.
- Pearce, M.T., Wiggins, G.A., 2012. Auditory expectation: the information dynamics of music perception and cognition. *Top Cognit. Sci.* 4 (4), 625–652.
- Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebr. Cortex* 23 (6), 1378–1387. <https://doi.org/10.1093/cercor/bhs118>.
- Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* 41 (1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3).
- Power, A.J., Mead, N., Barnes, L., Goswami, U., 2013. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 27 (7), 777. <https://doi.org/10.3389/fnhum.2013.00777>.
- Power, A.J., Colling, L.J., Mead, N., Barnes, L., Goswami, U., 2016. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain Lang.* 160, 1–10. <https://doi.org/10.1016/j.bandl.2016.06.006>.
- 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 (1), 121–131. <https://doi.org/10.1037/a0031277>.
- Rogalsky, C., Rong, F., Saberi, K., Hickok, G., 2011. Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31 (10), 3843–3852. <https://doi.org/10.1523/JNEUROSCI.4515-10.2011>.
- Tierney, A., Kraus, N., 2015. Evidence for multiple rhythmic skills. *PLoS One* 10 (9), e0136645. <https://doi.org/10.1371/journal.pone.0136645>.
- Vanden Bosch der Nederlanden, C.M., Hannon, E.E., Snyder, J.S., 2015a. Everyday musical experience is sufficient to perceive the speech-to-song illusion. *J. Exp. Psychol. Gen.* 144 (2), e43–e49. <https://doi.org/10.1037/xge0000056>.
- Vanden Bosch der Nederlanden, C.M., Hannon, E.E., Snyder, J.S., 2015b. Finding the music of speech: Musical knowledge influences pitch processing in speech. *Cognition* 143, 135–140. <https://doi.org/10.1016/j.cognition.2015.06.015>.
- Vanden Bosch der Nederlanden, C.M., Snyder, J.S., Hannon, E.E., 2016. Children use object-level category knowledge to detect changes in complex auditory scenes. *Dev. Psychol.* 52 (11), 1867–1877.
- Watson, M., Stewart, M., Krause, K., Rastatter, M., 1990. Identification of timecompressed sentential stimuli by good vs poor readers. *Percept. Mot. Skills* 71, 107–114.
- Woodruff-Carr, K., White-Schwoch, T., Tierney, A.T., Strait, D.L., Kraus, N., 2014. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. Unit. States Am.* 111 (40), 14559–14564.
- Zatorre, R.J., Baum, S.R., 2012. Musical melody and speech intonation: singing a different tune. *PLoS Biol.* 10 (7), e1001372 <https://doi.org/10.1371/journal.pbio.1001372>.
- Zatorre, R.J., Gandour, J.T., 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. *Phil. Trans. Biol. Sci.* 363 (1493), 1087–1104. <https://doi.org/10.1098/rstb.2007.2161>.
- Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14 (4), 1908–1919.
- Zoefel, B., Davis, M.H., 2017. Transcranial electric stimulation for the investigation of speech perception and comprehension. *Lang. Cognit. Neurosci.* 32 (7), 910–923. <https://doi.org/10.1080/23273798.2016.1247970>.