



Contents lists available at ScienceDirect

NeuroImage

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

Cortical tracking of rhythm in music and speech

Eleanor E. Harding^a, Daniela Sammler^{b,1}, Molly J. Henry^{c,d,1}, Edward W. Large^e,
Sonja A. Kotz^{a,g,*}

^a Department of Neuropsychology, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^b Otto Hahn Group “Neural Bases of Intonation in Speech and Music”, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^c Max Planck Research Group “Auditory Cognition”, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^d Department of Psychology, University of Connecticut Storrs, Connecticut, USA

^e Brain and Mind Institute, Department of Psychology, The University of Western Ontario, London, Ontario, Canada

^g Faculty of Psychology and Neuroscience, Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, the Netherlands

ARTICLE INFO

Keywords:

EEG
Entrainment
Rhythm
Music
Speech
Music training

ABSTRACT

Neural activity phase-locks to rhythm in both music and speech. However, the literature currently lacks a direct test of whether cortical tracking of comparable rhythmic structure is comparable across domains. Moreover, although musical training improves multiple aspects of music and speech perception, the relationship between musical training and cortical tracking of rhythm has not been compared directly across domains. We recorded the electroencephalograms (EEG) from 28 participants (14 female) with a range of musical training who listened to melodies and sentences with identical rhythmic structure. We compared cerebral-acoustic coherence (CACoh) between the EEG signal and single-trial stimulus envelopes (as measure of cortical entrainment) across domains and correlated years of musical training with CACoh. We hypothesized that neural activity would be comparably phase-locked across domains, and that the amount of musical training would be associated with increasingly strong phase locking in both domains. We found that participants with only a few years of musical training had a comparable cortical response to music and speech rhythm, partially supporting the hypothesis. However, the cortical response to music rhythm increased with years of musical training while the response to speech rhythm did not, leading to an overall greater cortical response to music rhythm across all participants. We suggest that task demands shaped the asymmetric cortical tracking across domains.

1. Introduction

Music and speech both contain salient auditory information on multiple timescales. Cortical activity in both music and speech domains phase-locks (Doelling and Poeppel, 2015; Luo and Poeppel, 2007; Doelling et al., 2014), or entrains (Nozaradan et al., 2011; Nozaradan et al., 2012; Buiatti et al., 2009), to rhythm. However, there has thus far been no direct comparison of cortical tracking of rhythm in the two domains. This is surprising, considering the theoretical domain-general applicability of models of perception of environmental rhythms (e.g., Schroeder and Lakatos, 2009) as well as entrainment models such as neural resonance (Large and Kolen, 1994; Large, 2008; Large and Snyder, 2009; Large et al., 2015) and dynamic attending theories (Jones, 1976, 2009; Large and Jones, 1999). Consequently, the current study explored neural oscillatory responses to rhythmic structure across the domains of music

and speech.

Further, musical training improves auditory perception of temporal features in both music (e.g., Geiser et al., 2010; Palmer and Krumhansl, 1990) and speech (e.g., Magne et al., 2007; Marie et al., 2011). Recent evidence suggests that musical expertise improves cortical tracking of music (Doelling and Poeppel, 2015). Moreover, musicians outperform non-musicians at speech-in-noise tasks, suggesting that their expertise translates to improved ability to track speech rhythm against a noisy background (Du and Zatorre, 2017; Slater and Kraus, 2016). Overlap of music and speech rhythm processing in the auditory and motor cortices (Chen et al., 2006; Golumbic et al., 2013; Grahn and Brett, 2007; Keitel et al., 2018; see Kotz et al., 2018, for a comprehensive review) additionally supports the potential for a cross-domain effect of musical training (Patel, 2011). Thus, we investigated cortical tracking of rhythm in participants with a range of formal musical training and related years

* Corresponding author. Department of Neuropsychology, Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

E-mail addresses: kotz@cbs.mpg.de, sonja.kotz@maastrichtuniversity.nl (S.A. Kotz).

¹ authors contributed equally.

<https://doi.org/10.1016/j.neuroimage.2018.10.037>

Received 22 August 2018; Received in revised form 8 October 2018; Accepted 13 October 2018

Available online 15 October 2018

1053-8119/© 2018 Elsevier Inc. All rights reserved.

of training to the cortical response to rhythm in both music and speech.

A recent corpus analysis of the rhythms of music and speech revealed important differences between domains, such as the presence of more spectral energy at somewhat lower frequencies in music than in speech (Ding et al., 2017). Moreover, rhythm is generally considered to be more regular, i.e. more temporally predictable, in music than in speech (Patel, 2008; Cummins, 2012; Kotz et al., 2018). Given rhythmic differences between domains, it may be misleading to directly compare cortical tracking of natural music and speech. Therefore we composed music and speech stimuli with carefully matched regular rhythmic structure and timed predictable downbeats to a metronome during recording. This allowed us to directly compare cortical tracking of rhythm across domains.

The current electroencephalography (EEG) study attempted to show whether, in accordance with current beat perception (e.g., Large et al., 2015) and neural oscillatory frameworks (e.g., Schroeder and Lakatos, 2009), neural activity is similarly phase-locked to music and speech rhythms when those rhythms are comprised of comparable temporal structure. We measured cerebral-acoustic coherence (CACoh; Peelle et al., 2013; Doelling et al., 2014) with the stimulus envelopes in a sample of participants with a range of musical training. We hypothesized that CACoh would be similar for music and speech and that CACoh in both domains would correlate positively with years of musical training (e.g., Doelling and Poeppel, 2015). However, we observed a stronger CACoh for music than speech, driven by an increase of music CACoh with greater amount of musical training that was not found for speech.

2. Materials & methods

2.1. Experimental design

Experimental stimuli were 120 Western tonal melodies composed by a professional musician and 120 German sentences written by a native German speaker. Items were designed to have hierarchical rhythmic structure, such that naturally occurring melodic-contour accents (music) or naturally occurring prosodic stress (speech) defined groups of three notes or syllables, giving rise to the perception of a triple meter or “waltz” (Fig. 1). A pianist and speaker timed these metrical downbeats to a 600 ms (1.667 Hz) metronome during recording. A 1.667 Hz downbeat was intended to also implicitly create a 5-Hz note/syllable rate based on a triple subdivision. Thus, experimental stimuli had similar temporal predictability across domains. Stimuli were presented in blocks with 50% fillers that were not included in the analysis. Filler items differed in respective harmonic and syntactic content of melodies and sentences to provide participants with a task but retained the same rhythm and temporal predictability (i.e. downbeats timed to the same metronome).

2.1.1. Participants

Twenty-eight participants (14 female; mean age = 24.8 ± 2.5 SD) were paid seven Euros per hour for their participation. They were all right-handed native German speakers according to the Edinburgh

handedness inventory (Oldfield, 1971) with self-reported normal hearing and no neurological history. *Musical training:* All participants had formal musical training, with a minimum of 3 and a maximum of 18 years (mean = 10.7 ± 4.3 SD). The participants' musical involvement spanned from casual enthusiasts who practiced a few hours per week to professional-track conservatory students, who practiced 40 hours a week. A year of formal musical training was defined as a year during which the participant underwent a structured lesson schema of any instrument or voice, either self-taught or by an instructor, and practiced an average of at least 1 hour per week. No participants were pianists to avoid motor-related responses to recorded piano melodies. Participants provided written informed consent. The experiment was approved by the local ethics committee of the University of Leipzig, Germany.

2.1.2. Procedure

Items were presented via Sony MDR-XD100 stereo headphones (Sony Corporation, Tokyo, Japan) at a comfortable volume while participants fixated on a white cross on a black background. They were instructed to blink between trials or during short self-paced breaks between the 6-min blocks. Items were separated by a 400-ms inter-stimulus interval (ISI). Participants were asked to mentally answer a question regarding the syntactic content of each melody and sentence (“Does the second phrase contain the same tonal key/agent as the first phrase?”) and were not informed about the rhythmic structure of the materials. In order to ensure their attention, after 10% of the items, a prompt indicated that they should provide their answer with a button press. Preparation and testing took approximately 90 min per session; two sessions occurred on different days (one during which all music stimuli were presented and one during which all speech stimuli were presented) in counterbalanced order across participants. Within sessions, items were presented in discrete blocks of approximately 6 min (60 items per block, 50% fillers). Experimental blocks were alternated with 6-min blocks of music or speech content from a different study.

2.2. Recording and preprocessing

2.2.1. Stimuli

One conservatory-trained pianist (30 years, female, 13 years of classical piano training) and one professionally trained native German speaker (30 years, female) were hired to produce the melodies and sentences, respectively. Musical scores and sentences were given to the pianist and speaker several days before the recording sessions, in order to allow them to become familiar with the material. During recording sessions, a 1.667 Hz (100 bpm) metronome was provided to the pianist and speaker via headphones with the instruction to time the downbeat (strong notes and syllables) to the metronome. Melodies were recorded on a Yamaha Clavinova CLP 150 (Yamaha Corporation, <https://en.wikipedia.org/wiki/Hamamatsu>, https://en.wikipedia.org/wiki/Shizuoka_Prefecture, <https://en.wikipedia.org/wiki/Japan>) electric MIDI keyboard (44.1 kHz sampling rate) using Finale 2008 (Boulder, CO, USA). Sentences were recorded with a Rode NT55 microphone (Silverwater, Australia) at 16-bit resolution

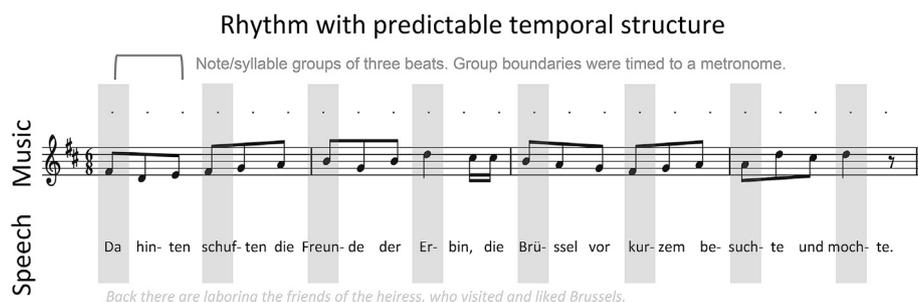


Fig. 1. Stimulus design. Melodies and sentences contained rhythm with predictable temporal structure, defined by the three-count individual notes or syllables (dots) between group boundaries (shaded). Group boundaries were realized dynamically through melodic contour or natural syllable weight and timed to a 1.667 Hz metronome during recording, which in turn facilitated a 5 Hz note or syllable rate.

and a sampling rate of 44.1 kHz using Cool Edit Pro 2.0 (Sibiu, Romania). Recorded stimuli ranged between 4 and 5 seconds (s) per item (sentences: 4.25–4.72 s range, mean 4.46 s, SD 0.99; melodies: 4.44–5.02 s range, mean 4.66 s, SD 1.18) and peak amplitude (intensity) was normalized to 70 dB using Audacity 1.3 and Praat 5.2 (Boersma and Weenink, 2011). Offline using MATLAB 8.0 (2012, The MathWorks, Inc., Natick, Massachusetts, United States), audio signals were passed through a cochlear filterbank using the auditory toolbox (Slaney, 1998). Per item, 128 filters were used between 180 Hz and Nyquist, and the Hilbert transform was taken of each of the 128 outputs. The outputs were then summed per item. Per item, summed envelopes were then low-pass filtered (50 Hz cutoff, zero-phase 6th-order Butterworth IIR to avoid aliasing) and downsampled by a factor of 441 to obtain a 100 Hz sampling rate (f_s). The filtered, downsampled stimuli envelopes were then cut into 4-s epochs (+460 ms to +4460 ms with respect to item onset) to line up with EEG trial epochs.

2.2.2. EEG

EEG was recorded with 64 Ag/AgCl electrodes placed in an elastic cap (Electro Cap Inc., Eaton, OH, USA) according to the extended 10–20 system (Sharbrough et al., 1991) using a 24 bit Brainvision QuickAmp 72 amplifier (Brain Products GmbH, Gilching, Germany). Sampling rate was 500 Hz. Impedances were kept below 5 k Ω throughout the experiment. Eye movements were monitored by bipolar horizontal and vertical electrooculograms (EOG) recorded from electrodes placed beneath the canthus of both eyes as well as above and below the right eye. Additionally, electrodes were placed on the left and right mastoid bones (M1 and M2) and a ground was placed on the sternum. The reference during recording was M1. Offline, EEG data were re-referenced to linked mastoids and high-pass filtered at 0.4 Hz to remove excessive drifts (-3 dB cutoff, fir, 3333 points, Blackman window) with EEProbe (ANT neuro, Enschede, NL).

Each trial was determined by the length of the corresponding stimulus (approximately 4.8 s). In order to prepare the dataset for independent component analysis (ICA), large artifacts were automatically removed whenever the standard deviation of the signal exceeded 20 μ V within a 200 ms sliding window and remaining visible artifacts were removed manually with the EEProbe graphic user interface. With FieldTrip, ICA was used to remove blinks and eye-movements. Five frontal electrodes, which were a source of noise in several participants, were then removed (FPz, FP1, FP2, AF3, AF7). Cleaned data were visually inspected and trials with artifacts were discarded. ICA-cleaned trials were low-pass filtered (50 Hz cutoff, zero-phase 6th-order Butterworth IIR to avoid aliasing) and downsampled by a factor of five to obtain a 100 Hz sampling rate. Data for four participants were excluded based on an a priori rejection criterion of >33% rejected trials, and one additional participant was excluded for incorrectly performing the task during the experiment (final N = 23). The average number of trials per condition (music or speech) that entered the analysis was 106 (6.6 SD). Trials were epoched into 4-s segments (+460 ms to +4460 ms with respect to item onset). Note that the epoch started after stimulus onset to avoid contamination of neural responses by the N1–P2 complex; 4460 ms corresponded to the shortest average stimulus length.

An auditory region of interest (ROI) was defined based on the auditory N1 event-related-potential (ERP) component. First, data for music and speech were averaged and the significance of the auditory N1 response time-locked to stimulus onset was evaluated separately for each electrode based on Monte Carlo FWE-corrected cluster-based permutation tested against a 100 ms zero-corrected baseline. The resulting 46-electrode ROI included: AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, F17, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4.

2.3. EEG analysis

Signal processing was performed with MATLAB 8.0 and the FieldTrip toolbox (Oostenveld et al., 2011). In order to explore phase locking to

music and speech by taking into account phase consistency over trials, we calculated cerebral-acoustic coherence (CACoh) between the single-trial stimulus envelope and the corresponding EEG signal (Peelle et al., 2013).

In order to improve the signal-to-noise ratio, cleaned 4-s EEG epochs were averaged over the region of interest (ROI) per trial. Then, epochs per trial were “reflected” (flipped horizontally and concatenated at beginning and end of forward time series; Cohen, 2014) to create 12-s epochs in order to obtain accurate low-frequency power data for the entire 4-s forward epoch.

Single-trial reflected data were transformed to the time–frequency domain using the Fieldtrip-implemented Morlet wavelets method (sliding window length capturing 7 cycles of each evaluated frequency, from 0.25 to 30 Hz) with a final frequency resolution of 0.5 Hz and final time resolution of 10 ms. The same time–frequency transformation was applied to the single-trial stimulus envelopes.

The coherence between single-trial EEG data and stimulus-envelope was calculated for each of 11 frequency bins between 1 and 6 Hz according to the following formula:

$$CACoh_{(f)} = \frac{1}{t} \left| \sum_{k=1}^t e^{i\Delta\theta(f,k)} \right|$$

where $\Delta\theta(f,k)$ is the angular distance between stimulus envelope and EEG at a single time point (k), e is the base of natural logarithms (~ 2.71), i is the imaginary unit that satisfies $i^2 = -1$, and t is time.

CACoh values were then averaged over trials separately for music and speech. As cortical tracking of speech envelope lags the acoustic signal (Di Liberto et al., 2015; Ding and Simon, 2011), CACoh was calculated for time lags up to 600 ms (one full beat cycle, 10-ms steps) between the stimulus envelope and the EEG signal. For each participant, the optimal lag resulting in the greatest CACoh value per domain was determined. A t -test evaluated differences in the CACoh at optimal lags across music and speech domains.

To test whether CACoh differed significantly from chance, a permutation strategy was adopted (Peelle et al., 2013), whereby single-trial coherence was calculated for each of 100 random stimulus–EEG pairings per participant and domain (using the optimal lag). CACoh values were averaged over trials and iterations, creating a single random-permutation CACoh value per person per domain. t -tests evaluated whether the real CACoh in music and speech was greater than the random-permutation values (alpha corrected to 0.025 for multiple comparisons).

2.4. Effects of musical training

In order to assess relationships between amount of musical training and cortical tracking of music and speech rhythm, we recorded years of formal musical training, defined as the total years that any instrument or voice was studied under formal or auto-didactic instruction, and entered Years-of-training as a covariate into a repeated measures analysis of covariance (RM ANCOVA) with factor Domain (music, speech). Years-of-training was mean-centered to avoid weakening of the main Domain effect (Delaney and Maxwell, 1981; Thomas et al., 2009). An interaction between Domain and Years-of-training was followed up by correlating Years-of-training with CACoh for both music and speech (Pearson correlation, 2-tailed, alpha corrected to 0.025 for multiple comparisons) and comparing music and speech CACoh after a median-split among participants with fewer and greater years of musical training (t -tests, alpha corrected to 0.025 for multiple comparisons).

3. Results

The comparison between real and randomly permuted CACoh data confirmed a significant phase-locked neural response to rhythm in both music and speech domains: Real CACoh at optimal lags per person per

condition was significantly greater than CACoh in random stimulus-EEG pairings ($t(22) < 11.8, p < .0001, d > 2.4$) (Fig. 2A).

Notably, music CACoh was significantly greater than speech CACoh (Fig. 2A) as shown by main effect Domain in the RM ANCOVA (mean CACoh (SD): music = 0.66 (0.008), speech = .65 (0.004); $F(1,22) = 26.1, p < .0001$, partial eta-squared = 0.60). However, this effect was qualified by a significant Domain x Years-of-training interaction ($F(1,22) = 5.0, p = .037$, partial eta-squared = 0.20). Follow-up Pearson's correlations between Years-of-training and CACoh in music and speech showed that CACoh in music—but not speech—increased with years of musical training (Fig. 3; music: $r = 0.500, p = .01$; speech: $r = 0.100, p = .60$). To further follow this up, participants were divided into high and low musical training groups based on a median split, and CACoh was compared for music vs. speech separately for the two groups (Fig. 2B). CACoh to musical stimuli was significantly higher than for speech for individuals with the most musical training ($t(11) = 6.5, p < .001$), but this was not true for the low training group ($t(10) = 1.7, p > .100$).

Optimal lags did not differ between domains (mean lag (SD): music = 162.6 (134.0), speech = 206.0 (113.3); $t(22) = -1.20, p = .200$).

4. Discussion

The current study investigated cortical tracking of music and speech with comparable rhythmic structure and temporal predictability and the impact that musical training has on the cortical response to rhythm. While proposed frameworks describing neuronal oscillatory entrainment to rhythmic environmental stimuli generally (e.g., Schroeder and Lakatos, 2009; Jones, 1976) and beat-based auditory rhythms specifically (e.g., Large et al., 2015) are theoretically applicable to both music and speech domains, the literature, up to now, has been lacking a direct test of whether cortical tracking is comparable across domains when rhythm is comparable. Moreover the effect of musical training has not been compared directly across domains. We hypothesized that EEG cerebral-acoustic coherence (CACoh) measures would reflect comparable cortical tracking of rhythm across domains and that musical training would be associated with neural activity more strongly phase-locked with rhythm in both domains. We found that participants with only a few years of musical training had a comparable cortical response to music and speech rhythm, in line with theoretical predictions. However, the cortical response to music rhythm increased with years of musical training while the response to speech rhythm did not, leading to an overall greater cortical response to music rhythm across all participants.

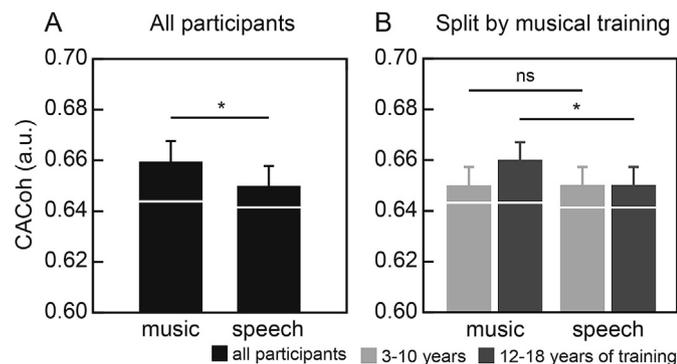


Fig. 2. Cerebral acoustic coherence (CACoh). In both domains, CACoh was significantly higher than CACoh of random stimulus-EEG pairings (white line). (A) Music CACoh was greater than speech CACoh (* indicates $p < .05$) among all participants. (B) Music and speech CACoh were similar among participants with fewer years of musical training (pale gray) while music was greater than speech CACoh (* indicates $p < .025$) among participants with more years of musical training (dark gray). Error bars represent ± 1 SEM.

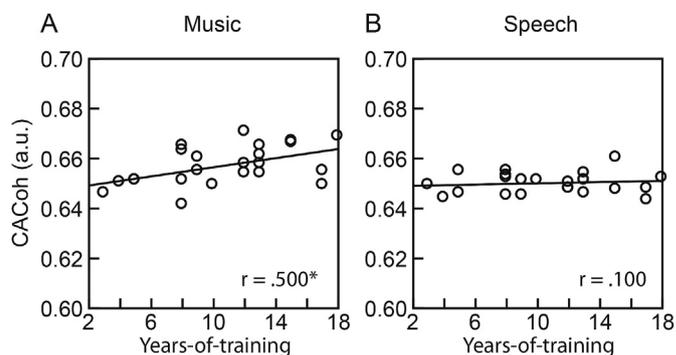


Fig. 3. Correlation with musical training. Years of musical training correlated positively with (A) music CACoh ($r = 0.500, p = .01$) but (B) not at all with speech CACoh ($r = 0.100, p = .600$) (* indicates $p < .025$).

4.1. Cortical tracking of music and speech rhythm

Cerebral-acoustic coherence, indexing single-trial synchronization of brain responses to individual items, showed a significantly more phase-locked cortical response to both music and speech compared to randomly permuted stimulus-EEG pairings. These findings replicate previous observations of cortical tracking of rhythm in music (e.g., Doelling and Poeppel, 2015) and speech (e.g., Doelling et al., 2014; Peelle et al., 2013) and are consistent with observations of entrained neuronal oscillations in music (Nozaradan et al., 2011) and speech (Buiatti et al., 2009; Ding et al., 2016). Neural activity phase-locked to the naturally spoken, temporally predictable hierarchical rhythm (waltz-like meter) is a novel finding in the speech domain, considering that previous paradigms investigating neuronal entrainment to speech used either natural stimuli with little-to-no hierarchical rhythmic organization (e.g. Peelle et al., 2013; Doelling et al., 2014) or artificial speech with temporal isochrony (Ding et al., 2016, 2017; Buiatti et al., 2009). Moreover, to our knowledge we provide the first direct comparison of cortical envelope tracking across the domains of music and speech. In line with theories that predict similar neural entrainment across domains (provided the rhythmic structure has similar temporal predictability; e.g., Large et al., 2015; Schroeder and Lakatos, 2009), we observed similarly strong cortical responses to music and speech, in particular for individuals with less musical training.

Our results for individuals with less musical training are generally in line with behavioral (e.g., Ellis and Jones, 2010; Jones et al., 2002), EEG (e.g., Fitzroy and Sanders, 2015; Brochard et al., 2003; Stefanics et al., 2010) and MEG (e.g., Snyder and Large, 2005; Iversen et al., 2009) findings that converge to support the idea that attention underpinned by low-frequency neuronal oscillations (see Henry and Hermann, 2014) is specifically guided to temporally predictable (musical) beat locations. Here, we extend the scope of those findings to the speech domain. Moreover, our ROI in both domains was based off of the auditory N1, which is attributed to generators in the auditory (Näätänen and Picton, 1987; Picton et al., 1999), and motor cortices (Giard et al., 1994). This is in line with literature that consistently cites auditory and motor cortices as generators of the neuronal oscillations underpinning entrainment to rhythms in speech (regularities in at multiple timescales; Keitel et al., 2018; Golumbic et al., 2013) and music (Chen et al., 2006, 2008; Grahn and Brett, 2007). We interpret these aspects of our findings as supporting the presence of domain-general mechanisms responsible for cortical entrainment to rhythm.

One limitation of the current methodology is that we are unable to distinguish between entrainment of neuronal oscillations by the rhythm of music and speech versus phase locking resulting from repeated evoked potentials caused by acoustic edges in the stimulus (Zoefel et al., 2018). Indeed, this is a limitation of a large majority of studies examining cortical tracking of naturalistic auditory stimuli (Alexandrou et al., 2018). The current study was not designed to address this old debate (see

Capilla et al., 2011; Makeig et al., 2002; David and Friston, 2003). However, we cautiously interpret our findings in the context of entrainment of neural oscillations, following much of the published literature on cortical envelope tracking (Ding and Simon, 2011; Doelling and Poeppel, 2015; Doelling et al., 2014; Di Liberto et al., 2015; Peelle et al., 2013; c. f. Rimmele et al., 2018). Moreover, we suggest that an important step forward in the immediate future of the field will be to design experiments that directly examine the theoretical and mechanistic underpinnings of the envelope-tracking response.

4.2. Musical training predicts cortical tracking of music, but not speech rhythm

Cortical tracking of musical rhythm has been previously demonstrated to depend on musical training (Doelling and Poeppel, 2015; Chen et al., 2008) and musicians show superior perception of speech in noise compared to nonmusicians (e.g., Slater and Kraus, 2016; Kraus and Chandrasekaran, 2010; Kraus and Anderson, 2014; Parbery-Clark et al., 2012). Thus, we hypothesized that musical training would be positively correlated with CACoh for both music and speech. We observed the predicted relationship for musical stimuli: among highly trained musicians (>12 years of musical training), the cortical response to the rhythm was greater in music compared to speech. We suggest that the positive finding in the music domain can be attributed to neural plasticity that accompanies many years of musical training (e.g., Herholz and Zatorre, 2012) and which likely increased the cortical response to rhythm specifically in the trained domain.

However, despite identical rhythmic structure of the music and speech stimuli, we showed an absence of the impact of musical training on neural response to speech rhythm. This initially seems to diverge from previous observations that musical training has transfer effects on various aspects of speech perception (e.g., Gordon et al., 2015; Magne et al., 2007; Marie et al., 2011; Jentschke and Koelsch, 2009; Wong et al., 2007). Yet, we may speculate on two reasons for this. First, the studies that have previously observed an impact of musical training on neuronal entrainment to speech have done so in speech-in-noise paradigms (Slater and Kraus, 2016; Du and Zatorre, 2017). That is, musical training seems to boost entrainment during difficult listening situations, when speech is presented against a noisy background. We presented clear speech against a quiet background with no added noise. Thus, our paradigm may not have presented a challenging enough listening situation to observe advantages based on training. Second, perhaps the enhanced speech perception in highly trained musicians (Patel, 2011) made the regular speech distractingly “non-natural,” which could have required suppression of oscillatory activity related to the regularized speech rhythm in order to better understand the content of the sentences. Indeed, it has been shown that the perception of rhythm in regularized speech (i.e., poetry) can come at the cost of the comprehension of its content (Meninghaus et al., 2015), and that certain types of continuous auditory perception (perhaps, related to phoneme perception; Giraud and Poeppel, 2012) can lead to suppression of rhythm-tracking low-frequency oscillations (Schroeder and Lakatos, 2009). Thus our highly trained musicians doing a “who-did-what-to-whom” sentence comprehension task on rhythmically regularized speech may have suppressed entrainment to speech in order to meet task demands, occluding musical-training-related effects. Future work is needed to firmly assess these possible explanations, and to appropriately fit highly trained musicians’ asymmetric cortical response to music and speech rhythm into theoretical accounts of the neural oscillatory underpinnings of rhythm perception.

5. Conclusion

The current study investigated the EEG neural oscillatory response to music and speech rhythms with comparable hierarchical and temporal predictability, along with the impact of musical training. While

participants with less musical training showed no difference between domains, greater cortical tracking of musical rhythm as opposed to speech rhythm was driven specifically by highly trained musicians’ increased cortical response to music. We suggest that the asymmetric response linked to musical training reflects active top-down adjustment of entrainment to rhythm based on listening conditions and task demands. Our results support models that propose neural oscillatory activity to underpin top-down percepts of rhythm.

Conflicts of interest

We declare no conflict of interest.

Peer reviewers openness initiative

In accordance with the Peer Reviewers Openness Initiative (<https://opennessinitiative.org/the-initiative/>), we declare that it is not possible to publicly archive the anonymized EEG data because we do not have consent from our participants to do so. Materials and summary data from Figs. 2 and 3 may be found at <https://band-lab.com/v-resources/stimuli/>.

Acknowledgements

We thank Burkhard Maess for advice regarding data preprocessing. This research was funded by: the Initial Training Network (ITN) European Grant: “Europe, Brain and Music: New perspectives for stimulating cognitive and sensory processes (EBRAMUS)” [grant agreement 238157]; the French Agence Nationale de la Recherche and the Deutsche Forschungsgemeinschaft joint project (ANR-DFG) “On the origins of grammar: from structural complexity in auditory sequences to syntactic structure”; Otto Hahn Award of the Max Planck Society to Daniela Sammler.

References

- Alexandrou, A.M., Saarinen, T., Kujala, J., Salmelin, R., 2018. Cortical entrainment: what we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience* 1–13.
- Boersma, P., Weenink, D., 2011. Praat: doing phonetics by computer [Computer program]. Version 5.2. retrieved 2 June 2011 from. <http://www.praat.org/>.
- 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. *Psychol. Sci.* 14, 362–366.
- Buiatti, M., Peña, M., Dehaene-Lambertz, G., 2009. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509–519.
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., Gross, J., 2011. Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One* 6 (1), e14543.
- Cohen, M.X., 2014. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.
- Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cognit. Neurosci.* 20 (2), 226–239.
- Chen, J.L., Zatorre, R.J., Penhune, V.B., 2006. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage* 32 (4), 1771–1781.
- Cummins, F., 2012. Looking for rhythm in speech. *Empirical Musicology Review* 7, 28–35.
- David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* 20 (3), 1743–1755.
- Delaney, H.D., Maxwell, S.E., 1981. On using analysis of covariance in repeated measures designs. *Multivariate Behav. Res.* 16, 105–123.
- Ding, N., Simon, J.Z., 2011. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol. (Bethesda)* 107 (1), 78–89.
- Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164.
- Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., Poeppel, D., 2017. Temporal Modulations Reveal Distinct Rhythmic Properties of Speech and Music. *bioRxiv*, 059683.
- Doelling, K.B., Poeppel, D., 2015. Cortical entrainment to music and its modulation by expertise. *P Natl Acad Sci USA* 112, E6233–E6242.
- Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768.

- Du, Y., Zatorre, R.J., 2017. Musical training sharpens and bonds ears and tongue to hear speech better. In: Proceedings of the National Academy of Sciences, 201712223.
- Ellis, R.J., Jones, M.R., 2010. Rhythmic context modulates foreperiod effects. *Atten. Percept. Psychophys.* 72, 2274–2288.
- Fitzroy, A.B., Sanders, L.D., 2015. Musical meter modulates the allocation of attention across time. *J. Cognit. Neurosci.* 27, 2339–2351.
- Geiser, E., Sandmann, P., Jäncke, L., Meyer, M., 2010. Refinement of metre perception—training increases hierarchical metre processing. *Eur. J. Neurosci.* 32, 1979–1985.
- Giard, M.H., Perrin, F., Echallier, J.F., Thevenet, M., Froment, J.C., Pernier, J., 1994. Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials Sect.* 92 (3), 238–252.
- Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15 (4), 511.
- Golumbic, E.M.Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., et al., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77 (5), 980–991.
- Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. *J. Cognit. Neurosci.* 19 (5), 893–906.
- Gordon, R.L., Shivers, C.M., Wieland, E.A., Kotz, S.A., Yoder, P.J., Devin McAuley, J., 2015. Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* 18, 635–644.
- Henry, M.J., Hermann, B., 2014. Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Percept.* 2, 62–86.
- Herholz, S.C., Zatorre, R.J., 2012. Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron* 76 (3), 486–502.
- Iversen, J.R., Repp, B.H., Patel, A.D., 2009. Top-down control of rhythm perception modulates early auditory responses. *Ann NY Acad Sci* 1169, 58–73.
- Jones, M.R., 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83, 323–355.
- Jones, M.R., 2009. Musical time. In: Hallam, S., Cross, I., Thaut, M. (Eds.), *The Oxford Handbook of Music Psychology*, second ed. Oxford UP, Oxford, pp. 81–92.
- Jones, M.R., Moynihan, H., MacKenzie, N., Puente, J., 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* 13, 313–319.
- Jentschke, S., Koelsch, S., 2009. Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744.
- 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.
- Kotz, S.A., Ravignani, A., Fitch, T.C., 2018. The evolution of rhythm processing. *Trends Cognit. Sci.* 22, 896–910.
- Kraus, N., Anderson, S., 2014. Music benefits across the lifespan: enhanced processing of speech in noise. *Hear. Rev.* 21, 18–21.
- Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Large, E.W., 2008. Resonating to musical rhythm: theory and experiment. In: Grondin, S. (Ed.), *The Psychology of Time*. Emerald, West Yorkshire, UK, pp. 189–232.
- Large, E.W., Herrera, J.A., Velasco, M.J., 2015. Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* 9.
- Large, E.W., Jones, M.R., 1999. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106 (119), 159.
- Large, E.W., Kolen, J.F., 1994. Resonance and the perception of musical meter. *Connect. Sci.* 6, 177–208.
- Large, E.W., Snyder, J.S., 2009. Pulse and meter as neural resonance. *Ann NY Acad Sci* 1169, 46–57.
- Di Liberto, G.M., O’Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25 (19), 2457–2465.
- Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010.
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., Besson, M., 2007. Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cerebr. Cortex* 17, 2659–2668.
- Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. *Science* 295 (5555), 690–694.
- Marie, C., Magne, C., Besson, M., 2011. Musicians and the metric structure of words. *J. Cognit. Neurosci.* 23, 294–305.
- Menninghaus, W., Bohrn, I.C., Knoop, C.A., Kotz, S.A., Schlotz, W., Jacobs, A.M., 2015. Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition* 143, 48–60.
- Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24 (4), 375–425.
- Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240.
- Nozaradan, S., Peretz, I., Mouraux, A., 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- 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.*, 156869.
- Palmer, C., Krumhansl, C.L., 1990. Mental representations for musical meter. *J. Exp. Psychol. Human* 16, 728–741.
- Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219, 111–119.
- Patel, A.D., 2008. *Music, Language, and the Brain*. Oxford UP, Oxford.
- Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The opera hypothesis. *Front. Psychol.* 2.
- 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, 1378–1387.
- Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., et al., 1999. Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology* 4, 64–79.
- Rimmele, J.M., Morillon, B., Poeppel, D., Arnal, L., 2018. Proactive sensing of periodic and aperiodic auditory patterns. *Trends Cognit. Sci.* 22, 870–882.
- Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18.
- Sharbrough, F., Chatrian, G.-E., Lesser, R.P., Lüders, H., Nuwer, M., Picton, T.W., 1991. American-electroencephalographic-society guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol.* 8, 200–202.
- Slaney, M., 1998. Auditory toolbox. Interval research corporation. Tech. Rep. 10, 1998.
- Slater, J., Kraus, N., 2016. The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognit. Process.* 17, 79–87.
- Snyder, J.S., Large, E.W., 2005. Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognit. Brain Res.* 24, 117–126.
- 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. *J. Neurosci.* 30, 13578–13585.
- Thomas, M.S.C., Annaz, D., Ansari, D., Serif, G., Jarrold, C., Karmiloff-Smith, A., 2009. Using developmental trajectories to understand developmental disorders. *J. Speech Lang. Hear. Res.* 52, 336–358.
- Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10 (4), 420.
- Zoefel, B., Ten Oever, S., Sack, A.T., 2018. The involvement of endogenous neural oscillations in the processing of rhythmic input: more than a regular repetition of evoked neural responses. *Front. Neurosci.* 12, 95.