



HAL
open science

Tuning neural phase entrainment to speech

Simone Falk, Cosima Lanzilotti, Daniele Schön

► **To cite this version:**

Simone Falk, Cosima Lanzilotti, Daniele Schön. Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience*, 2017, 29 (8), pp.1378-1389. 10.1162/jocn_a_01136 . hal-03583578

HAL Id: hal-03583578

<https://hal.science/hal-03583578>

Submitted on 21 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tuning neural phase entrainment to speech

Journal:	<i>Journal of Cognitive Neuroscience</i>
Manuscript ID	JOCN-2016-0268.R2
Manuscript Type:	Original
Date Submitted by the Author:	n/a
Complete List of Authors:	Falk, Simone; Ludwig-Maximilians-University, German Language and Literature Lanzilotti, Cosima; INSERM, Institut des Neurosciences des systèmes Schön, Daniele; INSERM, UMR1106
Keywords:	Rhythm, Prediction, Attention, Speech perception, Musical stimulation

1
2
3 Tuning neural phase entrainment to speech4
5 Simone Falk^{1,2,3}, Cosima Lanzilotti^{4,5}, Daniele Schön^{4,5}6
7 ¹ Laboratoire Parole et Langage, UMR 7309, Aix-Marseille Université, Aix-en-Provence,
8
9 France10
11 ² Laboratoire Phonétique et Phonologie, UMR 7018, CNRS, Université Sorbonne Nouvelle,
12
13 Paris-3, Paris, France14
15 ³ Institute of German Language and Literature, Ludwig-Maximilians-University Munich,
16
17 Germany18
19 ⁴ INSERM, UMR 1106, Marseille, France20
21 ⁵ Institut de Neurosciences des Systèmes, Aix-Marseille University, Marseille, France22
23
24
25
26
27
28
29
30
31
32 Running title: Tuning neural entrainment to speech33
34
35
36
37
38
39
40
41 Corresponding author: Simone Falk42
43 simone.falk@univ-paris3.fr44
45 Laboratoire Phonétique et Phonologie, Université Sorbonne Nouvelle,46
47 Paris-3, 19 Rue des Bernardins, 75005 Paris, France

Abstract

1
2
3 1
4
5 2
6
7
8 3 Musical rhythm positively impacts on subsequent speech processing. However, the neural
9
10 4 mechanisms underlying this phenomenon are so far unclear. We investigated whether carry-
11
12 5 over effects from a preceding musical cue to a speech stimulus result from a continuation of
13
14 6 neural phase entrainment to periodicities that are present in both music and speech.
15
16 7 Participants listened and memorized French metrical sentences that contained (quasi-)periodic
17
18 8 recurrences of accents and syllables. Speech stimuli were preceded by a rhythmically regular
19
20 9 or an irregular musical cue. Our results show that the presence of a regular cue modulates
21
22 10 neural response as estimated by EEG Power spectral density, Inter-trial coherence (ITC) and
23
24 11 source analyses at critical frequencies during speech processing compared to the irregular
25
26 12 condition. Importantly, ITCs for regular cues were indicative of the participants' success in
27
28 13 memorizing the subsequent speech stimuli. These findings underscore the highly adaptive
29
30 14 nature of neural phase entrainment across fundamentally different auditory stimuli. They also
31
32 15 support current models of neural phase entrainment as a tool of predictive timing and
33
34 16 attentional selection across cognitive domains.
35
36
37
38
39
40 17
41
42

Keywords

43
44
45
46 19 Rhythm, Prediction, Attention, Speech perception, Musical stimulation
47
48
49 20
50
51
52 21
53
54
55 22
56
57
58
59
60

1 A remarkable discovery of the last years is that musical experiences can facilitate certain
2 aspects of speech processing. Long-term musical training enhances representations of pitch,
3 timbre and timing in speech, fosters speech-in-noise recognition, verbal memory, speech
4 segmentation and promotes reading skills in typically developing as well as dyslexic children
5 (Elmer et al., 2014; Flaunacco et al., 2015; François, Chobert, Besson, & Schön, 2013;
6 François & Schön, 2014; Kraus, Hornickel, Strait, Slater, & Thompson, 2014; Moreno et al.,
7 2009; Strait, Parbery-Clark, Hittner, & Kraus, 2012). These effects are attributed to enhanced
8 neural plasticity in auditory and speech-related areas of the brain that have been observed
9 both at a functional and morphological levels (Herholz & Zatorre, 2012; Hyde et al., 2009;
10 Jäncke, 2009; Kraus & Chandrasekaran, 2010; Kraus et al., 2014; Schlaug, 2015). Besides
11 long-term effects of musical training, listening to musical rhythm can shape the encoding of
12 speech at a shorter term. For instance, Przybylski et al. (2013) showed that listening to 30 s of
13 highly regular music improved the accuracy in a syntactic task in both dyslexic and SLI
14 children. Similar results were also found in Parkinson patients (Kotz & Gunter, 2015). At an
15 even shorter time scale (i.e., a few seconds), we found that a rhythmical sequence of tones
16 matching the rhythm of the syllabic and accentual structure of a following pseudo-word or
17 short sentence enhanced speech perception in adults and speech production in hearing
18 impaired children (Cason, Astésano, & Schön, 2015; Cason, Hidalgo, Isoard, Roman, &
19 Schön, 2015; Cason & Schön, 2012). These findings are compatible with recent theories of
20 expectancy-driven speech processing (Kotz & Schwartze, 2010) and indicate that auditory
21 temporal expectancies induced by musical patterns can facilitate speech perception (Schön &
22 Tillmann, 2015). However, the neural processes scaffolding music-to-language transfers
23 during the encoding of musical and speech rhythms are so far unclear. Here, we investigate
24 the possibility that phase-coupling of neural oscillations with the temporal structure of speech
25 increases due to neural resonance with a preceding musical rhythm.

1 Behaviorally, humans synchronize with musical rhythms when they move their hands, heads
2 or legs to the beat. Neurally, listening to music stimulates a subcortical and cortical audio-
3 motor network. Periodicities of neural oscillations phase-lock to the musical beat (Fujioka,
4 Zendel, & Ross, 2010; Jones, 2009; Tierney & Kraus, 2015). This coupling of intrinsic (i.e.,
5 endogenous) to sensory (i.e., exogenous) rhythms has also been termed “entrainment”
6 (Calderone, Lakatos, Butler, & Castellanos, 2014; Lakatos, Karmos, Mehta, Ulbert, &
7 Schroeder, 2008; Large & Jones, 1999). The process is described as a tuning of neural
8 excitability towards expected times that may even continue when sensory stimulation has
9 stopped (Arnal & Giraud, 2012; Calderone et al., 2014). Nozaradan, Peretz, and Mouraux
10 (2012) demonstrated that neural entrainment to music is not merely the result of a response to
11 low-level acoustics of the sound envelope, but occurs when an internal metrical representation
12 (i.e., nested levels of perceived periodicities) of the sound is built. Using steady state-evoked
13 potentials, the authors showed that the peaks of neural resonance followed the structure of
14 metrical musical patterns. The strongest peak was found at the beat level (e.g., at 1.25 Hz),
15 but peaks in the power spectral density (PSD) were also present at higher and lower
16 harmonics (at 0.625, 2.5 and 5 Hz) representing higher-order beat groupings as well as
17 subdivisions down to the level of single tones. Neural entrainment is not exclusive to music,
18 though. Research has highlighted the fact that speech elicits robust neural entrainment, in
19 particular at low frequencies in the 4-8 Hz (theta) range (Ahissar et al., 2001; Giraud et al.,
20 2007; Giraud & Poeppel, 2012; Luo & Poeppel, 2007; Peelle & Davis, 2012). This frequency
21 range has been associated with the temporal signature of syllable production
22 (Chandrasekaran, Trubanova, Stillitano, Caplier, & Ghazanfar, 2009). For instance, when
23 reducing acuity of the spectral envelope, theta-band activity decreases in the auditory cortex
24 in parallel to speech intelligibility (e.g., Doelling, Arnal, Ghitza, & Poeppel, 2014). Further
25 studies also suggest that nested hierarchies of cortical oscillations entrain to speech, from

1
2
3 1 delta rhythms (< 4Hz), which represent fluctuations in speech prosody such as intonation and
4
5 2 accent structure (Giraud & Poeppel, 2012), to gamma rhythms (> 30Hz; e.g., Ding & Simon,
6
7 3 2014; Gross et al., 2013). Eventually, top-down signals from frontal and motor regions can
8
9 4 modulate entrainment to speech in the auditory cortex, with larger effects on the delta-band
10
11 5 frequencies (Park, Ince, Robin, Schyns, Thut, & Gross, 2015). In sum, processing of both
12
13 6 music and speech rhythms relies on coupling of brain-oscillatory activity to nested levels of
14
15 7 perceived event periodicities in these complex sounds.

16
17
18
19 8 Importantly, it has been suggested that neural entrainment optimizes the processing of
20
21 9 predictable events by providing a way for the sensory systems to encode ongoing information
22
23 10 (Calderone et al., 2014). For instance entrained EEG delta phase for rhythmic but not non-
24
25 11 rhythmic stimuli enhances visual contrast sensitivity (Cravo, Rohenkohl, Wyart & Nobre,
26
27 12 2013). Similarly, in the auditory domain, the phase of entrained oscillations correlates with
28
29 13 response time to target stimuli (Stefanics et al., 2010), target-detection accuracy (Henry,
30
31 14 Herrmann, & Obleser, 2014) and accelerates the perceptual emergence of task-relevant
32
33 15 temporal patterns (Riecke, Sack, & Schroeder, 2015). Similar functional contributions of
34
35 16 neural entrainment have been shown for speech stimuli (Ahissar et al., 2001, Luo & Poeppel,
36
37 17 2007) and the tracking of speech envelope information has been suggested to be a good
38
39 18 candidate to subserve the function of syllabic parsing (Doelling et al., 2014). Moreover,
40
41 19 entrainment to speech seems to go beyond the acoustic characteristics. It is affected by
42
43 20 listeners' ability to extract linguistic information (Peelle, Gross and Davies, 2013; Zoefel &
44
45 21 van Rullen, 2015b) and can track grammar-based linguistic structures (Ding, Melloni, Zhang,
46
47 22 Tian, & Poeppel, 2016).

48
49
50
51
52
53 23 In the present study, we examine the hypothesis that carry-over effects from a preceding
54
55 24 musical stimulus to a rhythmically matching speech stimulus result from a continuation of
56
57 25 neural phase entrainment to periodicities that are present in both the musical and the speech
58
59
60

1 stimulus. We presented participants with metrical speech stimuli containing quasi-periodic
2 recurrences of syllables (at 5 Hz) and periodic recurrences of accented syllables (at 1.65 Hz).
3 The speech stimuli were preceded either by a rhythmically regular tone sequence matching
4 the metrical structure of the speech pattern or sequences with an irregular temporal structure.
5 We predicted stronger oscillatory activity and inter-trial phase locking in speech following
6 matching musical sequences ('regular cue') compared to mismatching unstructured sequences
7 ('irregular cue'). Moreover, we also examined whether oscillatory activity during cue and
8 speech presentation may relate to behavioral response.

9 **Materials and Methods**

10 **Participants**

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

15 **Stimuli**

16 The linguistic material consisted of 60 spoken French utterances sharing the same syntactic
17 structure (i.e., consisting of two short sentences similar to the example in Fig. 1). Every
18 utterance comprised 20 syllables subdivided in four accentual phrases of five syllables each.
19 The utterances were recorded at 22050 Hz by a female speaker who was trained to pronounce
20 each accentual phrase with a Low-High-Low-High-intonation pattern (LHLH, Jun &
21 Fougeron, 2000; Welby, 2006) comprising a prenuclear ("secondary") accent on the second
22 syllable and the nuclear ("primary") accent on the last syllable (see Fig. 1).

23 -----Fig.1 ABOUT HERE-----

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

11 -----Fig.2 ABOUT HERE-----

12 Two types of musical stimuli were built to precede the speech stimuli in the Experiment,
13 regular and irregular. The regular rhythmic cue consisted of four sequences that stylized
14 average pitch and temporal structure of the speech stimuli. The first three sequences were
15 physically identical and consisted of five isochronous cello pizzicato sounds (at 22050 Hz)
16 lasting 200 ms each (F#, \uparrow G, G, G, \uparrow C, arrows indicate ascending or descending intervals).
17 The last sequence had the same temporal structure but different pitches similar to the final
18 intonation pattern of the speech stimuli (G, \uparrow Bb, Bb, Bb, \downarrow Eb). The last note lasted 250 ms to
19 mimic sentence final lengthening on the last syllable. The specific pitches were chosen
20 because they best reflected the average pitch contour present in the spoken sentences, with
21 minor adjustments to fit into a tonal system. Between the four sequences, a silence of 200 ms
22 was inserted in order to create a regular ternary meter. The ternary meter was also explicitly
23 marked by an increase in intensity on the second and last sound of each sequence (see Fig. 1).
24 Four versions of irregular cues were obtained by scrambling the pitches and the silences of the

1 regular cue. This resulted in four different cues that were rhythmically irregular, thus
2 preventing participants to become acquainted with the irregular structures. Figure 2A shows
3 the spectral representation of the envelope of regular and irregular cues. The peaks visible in
4 the regular cue reflect the temporal structure of the cue, with the ternary meter represented by
5 the peak at 1.65 Hz (=600 ms) and the note level presented by the 5 Hz peak (=200 ms).

6 **Procedure**

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

17 Stimulus presentation was structured in mini blocks of four trials. Within each mini block,
18 only one type of cue, regular or irregular, was used. Speech stimuli that were preceded by a
19 regular cue in session one, were preceded by an irregular cue in session two. The distribution
20 of speech stimuli across mini blocks and the order of conditions (regular vs. irregular) of mini
21 blocks were pseudo-randomized and counterbalanced across participants. The software
22 Presentation (Neurobehavioral Systems) was used to program the Experiment. A sound
23 Blaster X-Fi Xtreme Audio, an amplifier Yamaha P2040 and Yamaha loud-speakers (NS
24 10M) were used for sound presentation.

1 **Data analyses**

2 *Analysis of behavioral data*

3 Participants' answers in the detection task were analyzed as Hits when the participant
4 correctly notified a word as different from the words presented in the utterance. A False
5 Alarm occurred when the participant reported a difference when there was actually none.
6 Detection scores were calculated by subtracting % of False Alarms from % of Hits. d' was not
7 used to the absence of FA in some participants. Detection times were analyzed for Hits only.
8 Trials with values greater than two Standard deviations from the mean were excluded from
9 the analysis. Due to a technical problem with the button press setup we could not collect a
10 complete set of responses from three participants that were thus discarded from behavioral
11 analyses.

12 *Signal processing and statistical analyses*

13 EEG signal was recorded at 1000 Hz sampling rate using a BrainAmp amplifier and 64 pre-
14 amplified Ag-AgCl electrodes mounted following the 10-10 international system (actiCap) in
15 a sound-proof and Faraday cage. The ground electrode was placed at AFz and the reference
16 electrode at FCz.

17 Signal processing was done using EEGLAB (Delorme & Makeig, 2004, v12.022) and custom
18 Matlab scripts. Continuous data were filtered using a high-pass filter (0.4 Hz, 12 dB/octave,
19 Hamming windowed sinc FIR filter) and major artifacts rejected by visual inspection.

20 Independent Component Analysis (ICA, infomax) was used to remove physiological artifacts
21 such as eye-blinks and muscular activity. Data were then segmented into epochs of 5 s
22 starting at the onset of both cues and speech stimuli, separately. Epochs were zero-mean
23 normalized and re-referenced to the algebraic average of all electrodes. Further rejection of

1 remaining artifacts was done on the basis of visual inspection of each epoch and was always
2 lower than 10% of the total number of epochs in a given condition (yielding on average 55
3 epochs per condition per subject).

4 The spectral analysis was performed using the Welch's periodogram to compute the power
5 spectral density (PSD) over a window of 5000 ms using a Hanning tapering and no overlap.

6 The output consisted of 5000 amplitude estimates with a frequency bin width of 0.1 Hz (zero-
7 padding ratio of 2 to visually improve spectral resolution). The transformation was computed
8 separately for each participant, channel, condition and over non-averaged epochs (i.e., regular
9 and irregular cues and speech stimuli).

10 We performed two types of frequency analysis. The analysis of the power spectral density
11 (PSD) allowed estimating possible global changes in power at target frequencies. This method
12 does not take into account the phase information. The inter-trial coherence was used to
13 estimate the signal consistency across trials at each time-frequency point, taking into account
14 both phase and power information. This latter measure has been shown to be more sensitive to
15 stimulus synchronized neural activity than in power (Ding & Simon, 2013). In order to
16 control for the typical 1/F-like trend in the low-frequency EEG spectrum (Bédard et al.,
17 2006), the PSD was 1/F corrected and values at a target frequency (e.g., 0.8, 1.65 and 5 Hz)
18 were divided by the average value of two lower and two higher neighbors (e.g., for 5Hz: 4.7,
19 4.8, 5.2, 5.3 Hz). The validity of this normalization procedure relies on the assumption that, in
20 the absence of entrainment, the signal amplitude at a given frequency should be similar to the
21 signal amplitude of the mean of the surrounding frequencies (Nozaradan, Peretz, Missal, &
22 Mouraux, 2011; Nozaradan et al., 2012). Thus, this normalization ensures that ratios greater
23 than one actually indicate the presence of entrainment at specific frequencies. Statistical
24 analyses were run on a Region of Interest (ROI: Fz, Cz, F1, F2, FC1, FC2, FC3, FC4) best
25 describing the 5 Hz activity in the rhythmic cues, which was at the core of our working

1
2
3 1 hypothesis. The ROI was identified by using a non-parametric cluster-based permutation
4
5 2 approach (Maris & Oostenveld, 2007) that allowed to select one single cluster with $p < 0.01$
6
7 3 (see Figure 3C). As our aim was to study possible changes in oscillatory activity during
8
9 4 speech perception, the choice of the ROI on the basis of the 5Hz topography during the cues
10
11 5 was important in order to prevent a data selection bias for the statistical analyses of the speech
12
13 6 stimuli. Statistics were used to verify whether 1) ratios were significantly greater than 1,
14
15 7 indicating the presence of a peak (Wilcoxon signed-rank, one-sample), 2) the difference of
16
17 8 ratios in the regular and irregular condition was different from zero (Wilcoxon signed-rank,
18
19 9 paired).

20
21
22
23
24 10 Although not the primary aim of the present study, we used sLORETA (Pascual-Marqui,
25
26 11 2002) to localize the source of the response difference to speech stimuli in regular and
27
28 12 irregular conditions in the two most important frequency bands (1.65 and 5 Hz). sLORETA
29
30 13 method computes the three dimensional distribution of electrically active neuronal generators
31
32 14 within the brain as current density values based on the recorded scalp electric potential
33
34 15 differences and then estimates the inverse problem using the assumption that the smoothest of
35
36 16 all possible activities is the most plausible. The analysis was conducted with a three-shell
37
38 17 spherical head model registered to the Talairach atlas (voxel dimension 5 x 5 mm). Statistical
39
40 18 significance was assessed by means of a nonparametric randomization test ($p < 0.05$,
41
42 19 corrected for multiple comparisons).

43
44
45
46
47 20 Finally, in order to measure the phase coupling across trials at the expected frequencies, we
48
49 21 computed the Inter-Trial-Coherence (ITC) between 0 and 7 Hz using a discrete FFT and
50
51 22 Hanning tapering (using the eeglab newtimef function with FFTs and symmetric Hanning
52
53 23 window tapering, padratio of 4). The values for each frequency were then averaged across
54
55 24 time to yield a phase locking index for each estimated frequency. Again statistical analyses
56
57 25 were run on a Region of Interest (ROI: Fz, Cz, F1, F2, FC1, FC2, C1, C2, FC3) best
58
59
60

1 describing the 5 Hz ITC activity, and selected in the same way as it was done for the PSD
2 analyses. Non-parametric statistics were used to verify whether the difference of Phase
3 locking index across all estimated frequencies in the regular and irregular condition was
4 different from zero (Wilcoxon signed-rank, paired, FDR corrected).

5 **Results**

6 **Behavioral task**

7 Overall, participants' detection scores were high ($M = 79\%$, $SD = 10.4\%$), indicating that the
8 task was rather easy. No differences between detections in speech stimuli following regular or
9 irregular cues were found (regular cue: mean score = 78% , $SD = 9.1\%$; irregular cue: mean
10 score = 80% , $SD = 12.2\%$; $p > .60$), nor was there a difference in detection times (regular
11 cue: $M = 957$ ms, $SD = 235.1$ ms; irregular cue: $M = 955$ ms, $SD = 244.2$ ms, $p > .80$).

12 **Entrainment to speech after regular / irregular cue presentation**

13 First, we examined whether the power spectral density estimates (PSD) varied as a function of
14 the preceding musical cues when participants listened to speech (recall that the speech stimuli
15 were always metrical sentences and that the same stimuli were presented repeatedly after
16 different cues varied in regularity). Overall, averaged over cue condition, several peaks were
17 visible in the PSD of speech at expected frequencies of accent and syllable recurrence (5 Hz,
18 1.65 Hz and 0.8 Hz, see Fig. 3B), reflecting the three largest peaks visible on the *acoustic*
19 PSD of the speech stimuli (5 Hz, 1.65 Hz, 0.8 Hz; Fig. 2, right panel).

20 -----Fig.3 ABOUT HERE-----

21 Figure 4 shows that the normalized amplitude ratios at the expected frequencies for speech
22 stimuli preceded by regular cues were always significantly above 1 (5 Hz: $z = 2.1$, $p = 0.02$;

1
2
3 1 1.65 Hz: $z = 1.6$, $p = 0.05$; 0.8 Hz: $z = 2.1$, $p = 0.02$). This means that the peak value was
4
5 2 larger than the immediately surrounding values. When speech was preceded by irregular cues,
6
7 3 only the peak at 0.8 Hz was larger than the neighboring values ($z = 2.1$, $p = 0.02$).
8
9

10
11 4 -----Fig.4 ABOUT HERE-----
12

13
14 5 Direct comparisons of spectral peaks during speech perception in the regular and irregular
15
16 6 conditions showed a trend for larger peaks in the regular condition at 5 Hz and 1.65 Hz
17
18 7 although these differences did not reach significance ($z = 1.8$, $p = 0.07$; $z = 1.7$, $p = 0.08$). No
19
20 8 differences were present for the 0.8 Hz ($z = 0.8$, $p = 0.4$). However, the same contrast using
21
22 9 nonparametric bootstrapping analyses on current source densities as estimated by sLoreta did
23
24 10 show significantly larger responses at both 1.65 and 5 Hz for speech stimuli preceded by a
25
26 11 regular cue, compared to those preceded by an irregular cue (Fig. 5). These differences
27
28 12 reached significance, at the source level, in the left hemisphere over the middle and superior
29
30 13 temporal gyri and middle frontal and precentral gyri.
31
32
33
34

35
36 14 -----Fig.5 ABOUT HERE-----
37

38
39 15 Results of the inter-trial coherence (ITC) analyses for the speech stimuli also revealed a clear
40
41 16 peak at 5 Hz for speech stimuli preceded by regular cues (Fig. 6B, top right). This peak
42
43 17 corresponds to the 5 Hz peak visible on the envelope's PSD of the speech stimuli (Fig. 2B).
44
45 18 Statistical comparisons of ITC for sentences preceded by regular and irregular cues showed
46
47 19 significantly larger coherence in the regular condition around 5 Hz with p -values < 0.05
48
49 20 between 5 and 5.8 Hz (Fig. 6D).
50
51

52
53 21 -----Fig.6 ABOUT HERE-----
54
55
56
57 22
58
59
60

1 **Entrainment during cue presentation**

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

10 As expected from the *acoustic* PSD analyses (Fig. 2, left panel), results of the PSD for
11 musical cues showed clear peaks at 5 Hz for the regular and at 4.3 Hz for the irregular cues
12 (Fig. 3A). As to the rhythmical target frequencies, statistical analyses revealed significant
13 peaks at 5 Hz, 1.65 and 0.8 Hz for the regular cues ($z = 4.03, p < 0.0001$; $z = 2.9, p = 0.02, z$
14 $= 3.8, p = 0.0001$, respectively). For the irregular cues, the peaks at these frequencies were
15 smaller and mostly not significant (5 Hz: $z = 1.7, p = 0.05$; 1.65 Hz: $z = 1.6, p = 0.06$; 0.8 Hz:
16 $z = 1.3, p = 0.09$). Direct comparisons of spectral peaks during perception of regular and
17 irregular cues showed larger peaks during the regular condition at 5 Hz and 0.8 Hz ($z = 3.7, p$
18 < 0.001 ; $z = 2.5, p = 0.01$), while the difference at 1.65 Hz did not reach significance
19 ($p=0.15$).

20 Paralleling PSD analyses, ITC analyses also showed a clear peak at 5 Hz for the regular cue
21 and at 4.4 Hz for the irregular cue (Fig. 6A). Statistical comparison of ITC for regular and
22 irregular cues showed significantly larger coherence for regular cues between 4.4 and 6 Hz
23 with p -values < 0.001 between 4.8 and 5.6 Hz (Fig. 6C).

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

3 **Behavioral responses and neural entrainment**

4 Finally, despite the fact that the behavioral task did not in itself yield differences between
5 regularity conditions, we examined whether behavioral responses were associated with neural
6 entrainment in the cue conditions. Consequently, ITC measures for the cues and for speech
7 were correlated with the behavioral response accuracy of participants. Importantly, the ITC at
8 5 Hz measured during regular cue presentation was a good predictor of the overall accuracy in
9 the behavioral task (regular: $r = 0.45$, $p < 0.05$, irregular: $r = 0.3$, n.s., Spearman non-
10 parametric correlation, see Fig. 7). By contrast the ITC measured during speech presentation
11 did not correlate with behavioral responses ($r < 0.1$).

12 -----Fig.7 ABOUT HERE-----

13 **Discussion**

14 We investigated whether neural phase entrainment to the rhythmic structure of a non-verbal
15 musical cue enhances neural entrainment to a subsequent speech stimulus. Participants
16 listened to rhythmically regular or irregular non-verbal cues preceding the presentation of a
17 spoken utterance that matched the rhythm of the regular cue. Our results show that listening to
18 the regular cue modifies neural response at critical frequencies in the speech stimulus
19 compared to the irregular condition, especially at the syllable level. Analyses on the surface
20 (PSD) and source level jointly indicated stronger phase coupling at 5 Hz and, to a smaller
21 extent, at 1.65 Hz after regular vs. irregular cue presentation. This finding was corroborated
22 by analyses of inter-trial phase coherence (ITC), which index phase similarities across trials,
23 and revealed stronger phase coupling of neural responses around 5 Hz during regular vs.

1 irregular cue processing. Finally, the behavioral task we chose for the study turned out to be
2 very easy for the participants, and therefore, may have failed to show differences in memory
3 for words between the regular and irregular condition. This lack in sensitivity may also be at
4 the origin of the lack of correlation between the behavioural measure and neural activity
5 during speech perception. Yet, individual ITCs for regular cues were indicative of the
6 participants' success in memorizing the exact wording of the following speech stimuli. This
7 suggests that the state of the network at the beginning of the sentence (induced by the regular
8 cues) seems to play an important role in speech perception. Overall, the result integrates well
9 with findings from other tasks that showed behavioral benefits of non-verbal cueing for
10 speech processing (Cason, Hidalgo et al., 2015; Cason & Schön, 2012; Falk & Dalla Bella,
11 2016; Kotz & Gunter, 2015).

12 These findings support the hypothesis that the processing of metrical levels in speech can be
13 influenced by metrical levels heard before in a nonverbal (musical) sequence via neural
14 entrainment. Neural oscillations 1) phase-lock to and 2) are amplified at event periodicities
15 defined by the metrical levels that are present in both stimulus streams (Lakatos et al., 2008).
16 Thus, entraining neural oscillations with a temporally clearly structured auditory rhythm
17 results in a facilitated entrainment to speech when the temporal structure of speech and the
18 musical rhythm coincide. The observed carry-over effect from the musical cue to speech
19 raises the question if processes underlying neural entrainment for (rhythmically regular)
20 verbal and non-verbal stimuli follow similar principles. Some theories describe the function of
21 neural entrainment to beat and meter as enhancing "predictive timing" (Fujioka, Ross, &
22 Trainor, 2015) and thereby, attending towards salient events in an auditory stimulus (Kotz
23 & Schwartze, 2010; Large & Jones, 1999). In line with the oscillatory selection hypothesis
24 (Schroeder & Lakatos, 2009; see Frey, Ruhnau, & Weisz, 2015, for an overview), it has been
25 proposed that neural phase entrainment is a consequence of a rhythmic processing mode (in

1
2
3 1 contrast to “continuous processing mode”) that the brain adopts by aligning phases of high
4
5 2 neural excitability with salient (“attended”) events in a rhythmic signal (e.g., Peelle, Gross, &
6
7 3 Davis, 2013). In contrast to music which typically features musical beat structure, such a
8
9 4 process is less obvious for spontaneous speech that varies substantially in its temporal
10
11 5 regularity and predictability of prominent prosodic events (e.g., accents or syllables; see
12
13 6 Arvaniti, 2009; Cummins, 2015). However, intense behavioral and neuroimaging research of
14
15 7 the past years has convincingly demonstrated that rhythmically predictable verbal contexts
16
17 8 (e.g., through regular metrical patterns such as found in poetry, nursery rhymes, etc.) benefit
18
19 9 spoken language processing through enhanced predictive timing and higher attending towards
20
21 10 relevant events in speech (Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013; Rothermich,
22
23 11 Schmidt-Kassow, & Kotz, 2012; Schmidt-Kassow & Kotz, 2009; Schwartze, Rothermich,
24
25 12 Schmidt-Kassow, & Kotz, 2011). Furthermore, beneficial carry-over effects between non-
26
27 13 verbal and verbal rhythmic contexts have been observed. A non-verbal but temporally
28
29 14 predictable auditory cue mapping onto metrical structure in speech can induce temporal
30
31 15 predictions that direct attending towards upcoming events in speech at predicted times (Falk
32
33 16 & Dalla Bella, 2016; Falk, Volpi-Moncorger, & Dalla Bella, 2017; Kotz & Gunter, 2015). As
34
35 17 a consequence, facilitation of speech perception and production processes have been observed
36
37 18 (Cason, Astésano et al., 2015; Cason, Hidalgo et al., 2015; Cason & Schön, 2012). However,
38
39 19 to our knowledge, our study provides first evidence of how these carry-over effects are
40
41 20 implemented in terms of brain dynamics, by increasing both the amplitude and the phase
42
43 21 coupling strength at linguistically salient periodicities.

44
45
46
47
48
49
50 22 As for music, our results replicated previous findings showing phase-locking in the theta and
51
52 23 delta range to metrical levels of musical signals (e.g., Brochard, Abecasis, Potter, Ragot, &
53
54 24 Drake, 2003; Fujioka et al., 2010; Nozaradan et al., 2012; Schaefer, Vlek, & Desain, 2011).
55
56
57 25 For example, Nozaradan et al. (2011) found that a periodic stream of pure tones with a 2.4 Hz
58
59
60

1 frequency elicited steady-state EPs at this frequency and at subharmonic resonances at 1.2 or
2 0.8 Hz. The phase-locking could be altered depending on imagined binary or ternary
3 groupings of the sound. Thus, beat perception was largely based on predictions generated
4 through the participants' imaginary groupings. In the present study, we also found enhanced
5 spectral peaks in neural response (at 5 Hz and 0.8 Hz) as well as substantially higher ITC (at 5
6 Hz), when participants listened to the regular musical cue compared to the irregular cue.
7 Interestingly, while our stimuli were much shorter (i.e., 5 s) than in the aforementioned study
8 (i.e., 33 s), their metrical levels (0.8, 1.65 and 5 Hz) were still clearly visible in the neural
9 responses. This was possibly due to the fact that they exhibited acoustic information pointing
10 to metrical (i.e., enhanced intensity of tones in a ternary meter) and musical structure (i.e.,
11 pitch variations, phrase-final lengthening, timbre of a musical instrument). At least when
12 using an ecological cue, even a short exposure may be sufficient to entrain neural response to
13 metrical levels. Note that the presence of spectral peaks in the EEG response to the musical
14 cue could also be accounted for by a regular succession of evoked responses. Interestingly,
15 the oscillatory activity did not wear off immediately when the musical stimulus stopped.
16 Neural response to the following speech stimulus was also enhanced at points in time that
17 were predictable from the rhythmically regular cue. This maintenance of regular neural
18 activity across time and stimuli would be unusual for evoked-activity, but could be explained
19 in terms of entrainment of oscillatory activity. This finding indicates that neural oscillatory
20 dynamics seem to be maintained, even when the stimulus quality changes from non-verbal to
21 verbal.

22 In the present study, we used highly metrical and temporally predictable structures in both
23 verbal and musical stimuli. Despite the close metrical matching of both types of stimuli, the
24 carry-over effect from the musical cue to the speech stimulus via sustained neural entrainment
25 is not trivial, as speech was still substantially more variable than the musical cue in several

1 respects. First, the speech stimuli varied considerably in semantic and intonational structure.
2
3 Second, the speech stimuli displayed higher temporal variability at the basic metrical level of
4
5 syllable durations than the tones in the cue, as no temporal adjustments were done on syllabic
6
7 intervals. This may explain why we found enhanced ITC at precisely 5 Hz in the regular cue,
8
9 but an enhancement in a broader range around 5 Hz in speech. Third, while the musical beat
10
11 level was marked by higher intensity in the cue (at 1.65 Hz), this acoustic marker was not
12
13 present in the French speech stimuli in which accents are marked by pitch and duration
14
15 variations rather than intensity (Welby, 2006). Nevertheless, a peak is visible in the EEG
16
17 power spectrum (Fig. 3B & Fig.4) at 1.65 Hz in speech (i.e., the accentual level) only after
18
19 regular cue presentation. These are important findings that highlight the adaptive mechanisms
20
21 of neural oscillatory behavior (e.g. Gross et al., 2013). Flexible adaptation requires at least
22
23 two properties of the system. First, the neural entrainment shown at different frequencies
24
25 during non-verbal stimuli is robust to small frequency detuning during speech and can be
26
27 quickly retuned to new frequencies. Second, the phase of the oscillatory system keeps track of
28
29 previous oscillatory states (cue) when adapting to new oscillatory dynamics (speech), thus
30
31 explaining the greater entrainment to speech when preceded by regular cues.
32
33
34
35
36
37
38
39 Given two non identical oscillators, frequency entrainment will take place depending on
40
41 frequency detuning, namely the difference between two oscillators. The greater the difference,
42
43 the smaller should be the coupling (although the relation is not linear). For example, the 4.3
44
45 Hz generated by the irregular cues in our study did not seem to enhance the 5 Hz of
46
47 subsequent speech (Fig. 3A). By contrast, if the mismatch is not very large, the frequencies of
48
49 two systems become equal, or entrained, i.e., synchronization takes place. Generally, the
50
51 width of the synchronization region will increase with coupling strength (Pikowsky,
52
53 Rosenblum, & Kurths, 2001). Future research should further examine the possible regions of
54
55 synchronization in the case of auditory cueing.
56
57
58
59
60

1
2
3 1 It is also a remarkable finding, that a PSD peak was present at 1.65 Hz in the neural response
4
5 2 to speech following the regular cue, but not in the response to the regular cue itself. This
6
7 3 result may point to a process of oscillatory adaptation to metrical levels that increases in
8
9 4 strength over time provided that the metrical level is sufficiently salient and interpretable in
10
11 5 the signal (such as the regular accent structure in speech). It is indeed an open question how
12
13 6 stimulus *interpretation* may affect the process of neural entrainment to speech and music (see
14
15 7 Zoefel & VanRullen, 2015a, for a discussion). There is an ongoing debate whether the brain
16
17 8 merely tunes in to salient *acoustic* events (i.e., constituted by the peaks in the spectral
18
19 9 amplitude envelope) or if oscillatory activity is modulated by higher-order *linguistic* structure
20
21 10 in speech (i.e., sustained by linguistic predictions from the phonological, syntactic, semantic
22
23 11 etc. domain; (Zoefel & VanRullen, 2015b, 2016). Evidence for a “more-than-acoustics”
24
25 12 account in speech comes from studies showing that neural phase entrainment with speech in
26
27 13 the 5 Hz frequency range is influenced by speech intelligibility (Ahissar et al., 2001; Doelling
28
29 14 et al., 2014; Gross et al., 2013; Luo & Poeppel, 2007; Peelle & Davis, 2012; Peelle et al.,
30
31 15 2013). In music research, there is also evidence that neural phase entrainment is modulated by
32
33 16 higher-order hierarchical structure and its interpretation (e.g., Fujioka et al., 2015; Nozaradan
34
35 17 et al., 2011; Nozaradan et al., 2012). Future studies should further investigate the conditions
36
37 18 under which neural entrainment is altered by higher-order processes and predictions specific
38
39 19 to language and music.

40
41 20 In sum, cross-domain effects of neural phase entrainment from music to speech such as
42
43 21 observed in the present study require a highly flexible mechanism driving oscillatory
44
45 22 adaptation. The boundaries of flexibility of the mechanism await further clarification. For
46
47 23 example, it is an open issue whether neural phase entrainment will be observed when the cue
48
49 24 and subsequent speech stimuli are less regular and predictable. Note that the regular cues in
50
51 25 our study provided optimal matching conditions with the speech as they closely followed the
52
53
54
55
56
57
58
59
60

1
2
3 1 melodic contour and tempo of the sentences. They were exactly timed with the speech stimuli
4
5 2 such that the metrical pattern of speech directly resumed at expected times derived from the
6
7 3 metrical pattern of the cue. Future investigations should clarify whether different versions of
8
9 4 the cue (e.g., larger differences in pitch, tempo or timing) would yield similar or attenuated
10
11 5 neural entrainment. So far, behavioral results using auditory cues matching conversational
12
13 6 (i.e., rhythmically less regular) speech excerpts point towards benefits for verbal processing
14
15 7 (Cason, Astésano et al., 2015). On the other hand, these and other results (e.g., Falk
16
17 8 & Dalla Bella, 2016) revealed that benefits for speech processing can be boosted when motor
18
19 9 activity is combined with the auditory cue. These findings point towards accounts of neural
20
21 10 entrainment to speech and music that underscore the role of the motor system in generating
22
23 11 temporal predictions (Chemin, Mouraux, & Nozaradan, 2014; Chen, Penhune, & Zatorre,
24
25 12 2008; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; see Morillon, Hackett,
26
27 13 Kajikawa, & Schroeder, 2015, for a review). The theta-range in neural phase entrainment was
28
29 14 often interpreted as relating to sensorimotor cycles in syllable production, such as jaw
30
31 15 kinematics which, by close connections between auditory and motor systems, would become
32
33 16 relevant for speech perception (Fujioka et al., 2015; Kelso, Vatikiotis-Bateson, Saltzman, &
34
35 17 Kay, 1985; Kotz & Schwartz, 2010; Peelle & Davis, 2012). Involvement of the motor cortex
36
37 18 is also suggested in our study by higher activations found in the pre-central and middle frontal
38
39 19 gyri during speech processing (both at 1.65 and 5 Hz) after regular cue presentation.
40
41
42
43
44

45 46 **Conclusion**

47
48
49 21 The results of the present study provide fresh insights into the mechanisms underlying neural
50
51 22 entrainment and the link between music and speech rhythm processing. They offer an
52
53 23 explanation of how benefits of rhythmic non-verbal cueing on subsequent speech processing
54
55 24 can be generated via neural phase entrainment. These issues may be of particular relevance
56
57
58
59
60

1 for future research into therapeutical approaches using musical stimulation in speech
2 rehabilitation such as musical training in dyslexia (Flaugnacco et al., 2015; Overy, 2003) or
3 rhythmic cueing in hearing impairment (Thaut, McIntosh, & Hoemberg, 2014; Cason,
4 Hidalgo et al., 2015).

5 **Funding**

6 This work was supported by the European Union Seventh Framework Program (FP7-
7 PEOPLE-2012-IEF, n° 327586 to S.F.), the Brain and Language Research Institute (BLRI,
8 ANR-11-LABX-0036 to S.F. and D.S.) and the LMU*excellent* program within the framework
9 of the German excellence initiative (to S.F.).

11 **Acknowledgments**

12 We thank Deirdre Bolger, Virginie Epting, Patrick Marquis and Chloé Volpi, for help with
13 stimulus recording and data analyses. Corresponding author: simone.falk@germanistik.uni-
14 muenchen.de

15 **References**

- 16 Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M.
17 (2001). Speech comprehension is correlated with temporal response patterns recorded from
18 auditory cortex. *Proceedings of the National Academy of Sciences of the United States of*
19 *America*, 98(23), 13367–13372. doi:10.1073/pnas.201400998
- 20 Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in*
21 *Cognitive Sciences*, 16(7), 390–398. doi:10.1016/j.tics.2012.05.003
- 22 Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1-2), 46–63.
23 doi:10.1159/000208930

- 1
2
3 1 Bedard, C., Kroeger, H., & Destexhe, A. (2006). Does the 1/f frequency scaling of brain
4 signals reflect self-organized critical states? *Physical review letters*, *97*(11), 118102.
5
6 2
7 3 Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*(9-
8 10), 341–345.
9 4
10
11 5 Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our
12 internal clock: direct brain evidence of subjective accents in isochronous sequences.
13
14 6
15 7 *Psychological Science*, *14*(4), 362–366.
16
17 8 Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of
18 neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*,
19
20 9
21 10 *18*(6), 300–309. doi:10.1016/j.tics.2014.02.005
22
23 11 Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: rhythmic
24 priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50.
25
26 12
27 13 doi:10.1016/j.actpsy.2014.12.002
28
29 14 Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming
30 enhances speech production abilities: evidence from prelingually deaf children.
31
32 15
33 16 *Neuropsychology*, *29*(1), 102–107. doi:10.1037/neu0000115
34
35 17 Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of
36 speech. *Neuropsychologia*, *50*(11), 2652–2658.
37
38 18
39 19 doi:10.1016/j.neuropsychologia.2012.07.018
40
41 20 Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009).
42 The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436.
43
44 21
45 22 doi:10.1371/journal.pcbi.1000436
46
47 23 Chemin, B., Mouraux, A., & Nozaradan, S. (2014). Body movement selectively shapes the
48 neural representation of musical rhythms. *Psychological Science*, *25*(12), 2147–2159.
49
50 24
51 25 doi:10.1177/0956797614551161
52
53
54
55
56
57
58
59
60

- 1
2
3 1 Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits
4
5 2 motor regions of the brain. *Cerebral Cortex*, *18*(12), 2844–2854.
6
7 3 doi:10.1093/cercor/bhn042
8
9 4 Cravo, A.M., Rohenkohl, G., Wyart, V., & Nobre, A.C. (2013). Temporal expectation
10
11 5 enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual
12
13 6 cortex. *Journal of Neuroscience*, *33*(9), 4002–4010. doi: 10.1523/JNEUROSCI.4675-
14
15 7 12.2013
16
17 8 Cummins, F. (2015). Rhythm and Speech. In M. A. Redford (Ed.), *Blackwell Handbooks in*
18
19 9 *Linguistics. The Handbook of Speech Production* (1st ed., pp. 158–177). Hoboken NY:
20
21 10 John Wiley & Sons.
22
23 11 Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of*
24
25 12 *Phonetics*, *26*(3), 145–171.
26
27 13 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-
28
29 14 trial EEG dynamics including independent component analysis. *Journal of Neuroscience*
30
31 15 *methods*, *134*(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009
32
33 16 Ding, N., & Simon, J.Z. (2013). Power and phase properties of oscillatory neural responses in
34
35 17 the presence of background activity. *Journal of Computational Neuroscience*, *34*(2), 337-
36
37 18 343.
38
39 19 Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles
40
41 20 and interpretations. *Frontiers in Human Neuroscience*, *8*, 311.
42
43 21 doi:10.3389/fnhum.2014.00311
44
45 22 Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of
46
47 23 hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158-
48
49 24 164. doi: 10.1038/nn.4186
50
51 25 Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive
52
53 26 delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing.
54
55 27 *NeuroImage*, *85 Pt 2*, 761–768. doi:10.1016/j.neuroimage.2013.06.035
56
57
58
59
60

- 1
2
3 1 Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., & Jäncke, L. (2014). Music and
4 language expertise influence the categorization of speech and musical sounds: behavioral
5 2 language expertise influence the categorization of speech and musical sounds: behavioral
6 and electrophysiological measurements. *Journal of Cognitive Neuroscience*, *26*(10), 2356–
7 3 and electrophysiological measurements. *Journal of Cognitive Neuroscience*, *26*(10), 2356–
8 4 2369. doi:10.1162/jocn_a_00632
9
10
11 5 Falk, S., & Dalla Bella, S. (2016). It is better when expected: Aligning verbal and motor
12 6 rhythms enhances verbal processing. *Language, Cognition, and Neuroscience*, *31*(5), 699–
13 7 708.
14
15
16
17 8 Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music
18 9 Training Increases Phonological Awareness and Reading Skills in Developmental
19 10 Dyslexia: A Randomized Control Trial. *PloS One*, *10*(9), e0138715.
20 11 doi:10.1371/journal.pone.0138715
21
22
23
24 12 François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the
25 13 development of speech segmentation. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(9),
26 14 2038–2043. doi:10.1093/cercor/bhs180
27
28
29
30 15 François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental
31 16 biological process that underlies auditory learning: the role of musical practice. *Hearing*
32 17 *Research*, *308*, 122–128. doi:10.1016/j.heares.2013.08.018
33
34
35
36 18 Frey, J. N., Ruhnau, P., & Weisz, N. (2015). Not so different after all: The same oscillatory
37 19 processes support different types of attention. *Brain Research*, *1626*, 183–197.
38 20 doi:10.1016/j.brainres.2015.02.017
39
40
41
42 21 Fujioka, T., Ross, B., & Trainor, L. J. (2015). Beta-Band Oscillations Represent Auditory
43 22 Beat and Its Metrical Hierarchy in Perception and Imagery. *Journal of Neuroscience*,
44 23 35(45), 15187–15198. doi:10.1523/JNEUROSCI.2397-15.2015
45
46
47
48 24 Fujioka, T., Zendel, B. R., & Ross, B. (2010). Endogenous neuromagnetic activity for mental
49 25 hierarchy of timing. *Journal of Neuroscience*, *30*(9), 3458–3466.
50 26 doi:10.1523/JNEUROSCI.3086-09.2010
51
52
53
54
55
56
57
58
59
60

- 1
2
3 1 Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, Richard S J, & Laufs,
4
5 2 H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech
6
7 3 perception and production. *Neuron*, *56*(6), 1127–1134. doi:10.1016/j.neuron.2007.09.038
8
9 4 Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging
10
11 5 computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.
12
13 6 doi:10.1038/nn.3063
14
15 7 Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013).
16
17 8 Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS*
18
19 9 *Biology*, *11*(12), e1001752. doi:10.1371/journal.pbio.1001752
20
21 10 Henry, M.J., Herrmann, B., & Obleser, J. (2014). Entrained neural oscillations in multiple
22
23 11 frequency bands comodulate behavior. *Proceedings of the National Academy of Sciences*,
24
25 12 *111*(41), 14935-14940. doi: 10.1073/pnas.1408741111
26
27 13 Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity:
28
29 14 behavior, function, and structure. *Neuron*, *76*(3), 486–502.
30
31 15 doi:10.1016/j.neuron.2012.10.011
32
33 16 Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G.
34
35 17 (2009). Musical training shapes structural brain development. *Journal of Neuroscience*,
36
37 18 *29*(10), 3019–3025. doi:10.1523/JNEUROSCI.5118-08.2009
38
39 19 Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience*, *27*(5),
40
41 20 521–538. doi:10.3233/RNN-2009-0519
42
43 21 Jones, M. R. (2009). Musical time. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford Library*
44
45 22 *of Psychology. The Oxford Handbook of Music Psychology* (pp. 81–92). Oxford, New
46
47 23 York: Oxford University Press.
48
49 24 Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis
50
51 25 (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 209–242). Boston: Kluwer.
52
53
54
55
56
57
58
59
60

- 1
2
3 1 Kelso, J. A., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic
4 analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling.
5 2
6 *The Journal of the Acoustical Society of America*, 77(1), 266–280. doi:10.1121/1.392268
7 3
8
9 4 Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related
10 deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, 1337, 62–68.
11 5
12 doi:10.1111/nyas.12657
13 6
14
15 7 Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely
16 subcortico-cortical framework. *Trends in Cognitive Sciences*, 14(9), 392–399.
17 8
18 doi:10.1016/j.tics.2010.06.005
19 9
20
21 10 Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory
22 skills. *Nature Reviews. Neuroscience*, 11(8), 599–605. doi:10.1038/nrn2882
23 11
24
25 12 Kraus, N., Hornickel, J., Strait, D. L., Slater, J., & Thompson, E. (2014). Engagement in
26 community music classes sparks neuroplasticity and language development in children
27 13
28 from disadvantaged backgrounds. *Frontiers in Psychology*, 5, 1403.
29 14
30 doi:10.3389/fpsyg.2014.01403
31 15
32
33 16 Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of
34 neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872), 110–
35 17
36 113. doi:10.1126/science.1154735
37 18
38
39 19 Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-
40 varying events. *Psychological Review*, 106(1), 119–159. doi:10.1037/0033-
41 20
42 295X.106.1.119
43 21
44
45 22 Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate
46 speech in human auditory cortex. *Neuron*, 54(6), 1001–1010.
47 23
48 doi:10.1016/j.neuron.2007.06.004
49 24
50
51 25 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.
52 *Journal of Neuroscience methods*, 164(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024
53 26
54
55
56
57
58
59
60

- 1
2
3 1 Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical
4 training influences linguistic abilities in 8-year-old children: more evidence for brain
5 2 plasticity. *Cerebral Cortex*, *19*(3), 712–723. doi:10.1093/cercor/bhn120
6
7
8
9 4 Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor
10 control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology*,
11 5 *31*, 230–238. doi:10.1016/j.conb.2014.12.005
12
13
14
15 7 Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal
16 8 entrainment to beat and meter. *Journal of Neuroscience*, *31*(28), 10234–10240.
17 9 doi:10.1523/JNEUROSCI.0411-11.2011
18
19
20 10 Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat
21 11 and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*(49), 17572–17581.
22 12 doi:10.1523/JNEUROSCI.3203-12.2012
23
24
25
26 13 Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of*
27 14 *the New York Academy of Sciences*, *999*, 497–505.
28
29
30
31
32 15 Park, H., Ince, Robin A A, Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down
33 16 signals increase coupling of auditory low-frequency oscillations to continuous speech in
34 17 human listeners. *Current biology*, *25*(12), 1649–1653. doi:10.1016/j.cub.2015.04.049
35
36
37
38 18 Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography
39 19 (sLORETA): technical details. *Methods and Findings in Experimental and Clinical*
40 20 *Pharmacology*, *24 Suppl D*, 5–12.
41
42
43
44 21 Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to
45 22 Comprehension. *Frontiers in Psychology*, *3*, 320. doi:10.3389/fpsyg.2012.00320
46
47
48 23 Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human
49 24 auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387.
50 25 doi:10.1093/cercor/bhs118
51
52
53
54
55 26 Pikowsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization. A universal concept in*
56 27 *nonlinear sciences*. New York: Cambridge University Press.
57
58
59
60

- 1
2
3 1 Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., . . .
4
5 2 Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in
6
7 3 children with developmental language disorders. *Neuropsychology*, *27*(1), 121–131.
8
9 4 doi:10.1037/a0031277
10
11 5 Riecke, L., Sack, A.T., & Schroeder, C.E. (2015). Endogenous Delta/Theta sound-brain phase
12
13 6 entrainment accelerates the buildup of auditory streaming. *Current Biology*, *25*(24), 3196-
14
15 7 3201. doi: 10.1016/j.cub.2015.10.045.
16
17 8 Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm
18
19 9 facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One*, *8*(2), e56000.
20
21 10 doi:10.1371/journal.pone.0056000
22
23 11 Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: regular
24
25 12 meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244.
26
27 13 doi:10.1016/j.neuropsychologia.2011.10.025
28
29 14 Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Decomposing rhythm processing:
30
31 15 electroencephalography of perceived and self-imposed rhythmic patterns. *Psychological*
32
33 16 *Research*, *75*(2), 95–106. doi:10.1007/s00426-010-0293-4
34
35 17 Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity.
36
37 18 *Progress in Brain Research*, *217*, 37–55. doi:10.1016/bs.pbr.2014.11.020
38
39 19 Schmidt-Kassow, M., & Kotz, S. A. (2009). Attention and perceptual regularity in speech.
40
41 20 *Neuroreport*, *20*(18), 1643–1647. doi:10.1097/WNR.0b013e328333b0c6
42
43 21 Schön, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: perspectives
44
45 22 for language rehabilitation. *Annals of the New York Academy of Sciences*, *1337*, 32–39.
46
47 23 doi:10.1111/nyas.12635
48
49 24 Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of
50
51 25 sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. doi:10.1016/j.tins.2008.09.012
52
53
54
55
56
57
58
59
60

- 1
2
3 1 Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics
4 of Active Sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–
5 176. doi:10.1016/j.conb.2010.02.010
6
7
8
9 4 Schwartze, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Temporal
10 regularity effects on pre-attentive and attentive processing of deviance. *Biological*
11 *Psychology*, *87*(1), 146–151. doi:10.1016/j.biopsycho.2011.02.021
12
13
14
15 7 Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase
16 entrainment of human delta oscillations can mediate the effects of expectation on reaction
17 speed. *Journal of Neuroscience*, *30*(41), 13578–13585. doi: 10.1523/JNEUROSCI.0703-
18 10.2010
19
20
21
22 11 Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early
23 childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3),
24 191–201. doi:10.1016/j.bandl.2012.09.001
25
26
27
28 14 Thaut, M. H., McIntosh, G. C., & Hoemberg, V. (2014). Neurobiological foundations of
29 neurologic music therapy: rhythmic entrainment and the motor system. *Frontiers in*
30 *Psychology*, *5*, 1185. doi:10.3389/fpsyg.2014.01185
31
32
33
34 17 Tierney, A., & Kraus, N. (2015). Neural entrainment to the rhythmic structure of music.
35 *Journal of Cognitive Neuroscience*, *27*(2), 400–408. doi:10.1162/jocn_a_00704
36
37
38
39 19 Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of*
40 *Phonetics*, *34*(3), 343–371.
41
42
43
44 21 Zoefel, B., & VanRullen, R. (2015a). The Role of High-Level Processes for Oscillatory Phase
45 Entrainment to Speech Sound. *Frontiers in Human Neuroscience*, *9*, 651.
46 doi:10.3389/fnhum.2015.00651
47
48
49
50 24 Zoefel, B., & VanRullen, R. (2015b). Selective perceptual phase entrainment to speech
51 rhythm in the absence of spectral energy fluctuations. *Journal of Neuroscience*, *35*(5),
52 1954–1964. doi:10.1523/JNEUROSCI.3484-14.2015
53
54
55
56
57
58
59
60

1
2
3 1 Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level
4
5 2 features of speech sound. *NeuroImage*, 124(Pt A), 16–23.
6
7 3 doi:10.1016/j.neuroimage.2015.08.054
8
9 4
10 5
11 6
12 7
13 8
14 9
15 10
16 11
17 12
18 13
19 14
20 15
21 16
22 17
23 18
24 19
25 20
26 21
27 22
28 23
29 24
30 25
31 26
32 27
33 28
34 29
35 30
36 31
37 32
38 33
39 34
40 35
41 36
42 37
43 38
44 39
45 40
46 41
47 42
48 43
49 44
50 45
51 46
52 47
53 48
54 49
55 50
56 51
57 52
58 53
59 54
60 55

7 Captions to Figures

8
9 **Figure 1.** Oscillograms of the non-verbal cues (upper panel) and an example of a verbal
10 stimulus (‘the son of the neighbor blocks the road, he retards the bank director’). Big triangles
11 mark primary accents, medium triangles mark secondary accents and small triangles mark
12 non-accented syllables.
13

14 **Figure 2.** Audio Signal. A: Mean envelope power spectral density (PSD) for regular and
15 irregular cues computed after band-pass filtering (300-2500 Hz) and Hilbert transform. Note
16 that the peak at 3.3 Hz for the regular cue is the first harmonic of the 1.65Hz peak, although
17 this frequency was not present in the metrical structure. B: Mean envelope power spectral
18 density for target sentences. The arrows point to the frequencies of interest.
19

20 **Figure 3.** EEG signal. Top row. A: 1/f rectified Power spectral density (PSD) for Regular
21 and Irregular Cues. B: 1/f rectified Power spectral density for sentences preceded by regular
22 and irregular cues. Bottom row. C: Topographic maps of the raw PSD at 5Hz for regular cues.
23 D: Topographic maps of the raw PSD at 5Hz for all sentences (regular and irregular
24 conditions).

1

2
3
4
5
6 **Figure 4.** EEG signal. Normalized amplitude ratios at the expected frequencies for speech
7 stimuli preceded by regular and irregular cues. Values larger than 1 indicate a peak in the
8 spectrum. Whiskers represent the within-subject standard error from the mean.
9
10
11
12

13

14
15
16 **Figure 5.** Statistical non-parametric maps (t-values) of the difference between speech stimuli
17 preceded by regular and irregular cues (after bootstrapping) for the two most important target
18 frequencies (1.65 Hz, top and 5 Hz, bottom) over left and right hemispheres.
19
20
21
22
23

24

25
26
27 **Figure 6.** Top row. A: Intertrial coherence for regular and irregular cues. B: Intertrial
28 coherence for sentences preceded by regular and irregular cues. Bottom row. C: *p*-values
29 illustrating the differences between regular and irregular cues as a function of frequency. D:
30 *p*-values illustrating the differences between sentences preceded by regular and irregular cues
31 as a function of frequency.
32
33
34
35
36
37
38
39
40
41

42

43 **Figure 7.** Scatter plot of the performance in the behavioral task (% of Hits minus % of False
44 Alarms) versus inter-trial coherence during regular cue processing (preceding the speech
45 stimuli) for each participant. The dark line is the linear fit and the dotted lines indicate 95%
46 confidence intervals.
47
48
49
50

51

52

53

54

55

56

57

58

59

60

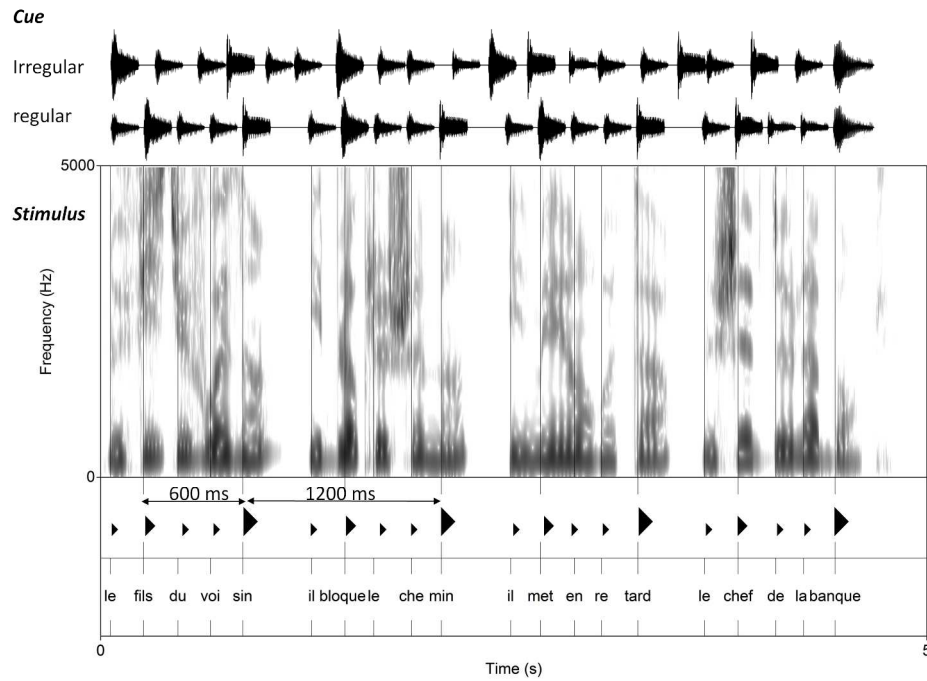


Figure 1: Oscillograms of the non-verbal cues (upper panel) and an example of a verbal stimulus ('the son of the neighbor blocks the road, he retards the bank director'). Big triangles mark primary accents, medium triangles mark secondary accents and small triangles mark non-accented syllables.

308x210mm (253 x 253 DPI)

Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

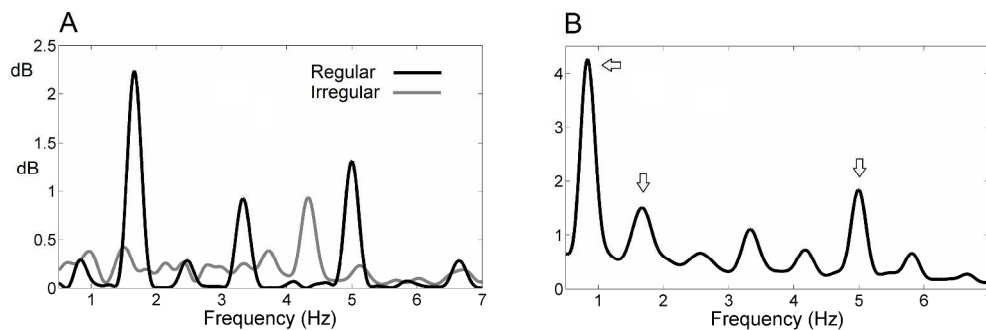


Figure 2. Audio Signal. A: Mean envelope power spectral density (PSD) for regular and irregular cues computed after band-pass filtering (300-2500 Hz) and Hilbert transform. Note that the peak at 3.3 Hz for the regular cue is the first harmonic of the 1.65Hz peak, although this frequency was not present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

1363x467mm (96 x 96 DPI)

Review Only

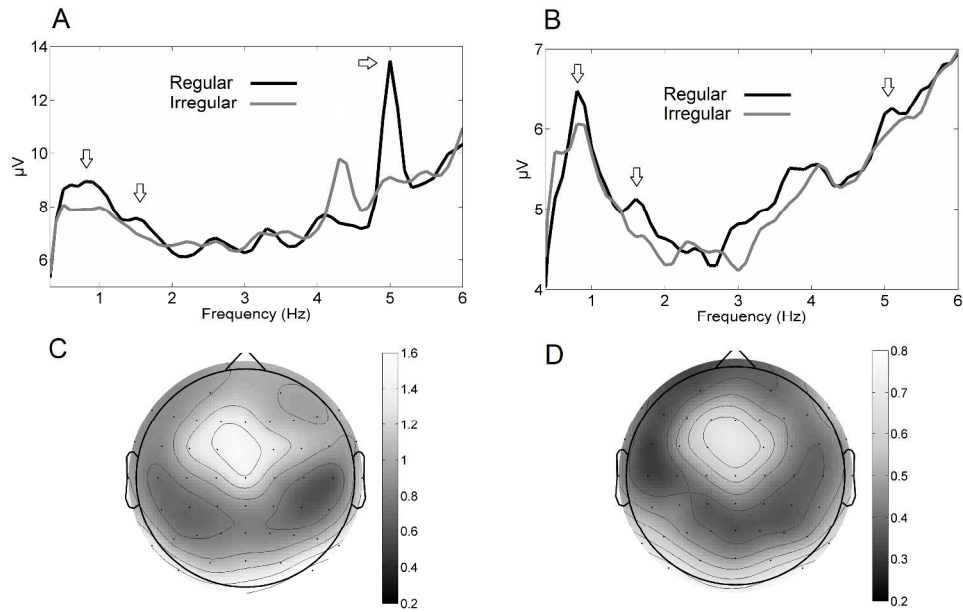


Figure 3. EEG signal. Top row. A: 1/f rectified Power spectral density (PSD) for Regular and Irregular Cues. B: 1/f rectified Power spectral density for sentences preceded by regular and irregular cues. Bottom row. C: Topographic maps of the raw PSD at 5Hz for regular cues. D: Topographic maps of the raw PSD at 5Hz for all sentences (regular and irregular conditions).

1400x880mm (96 x 96 DPI)

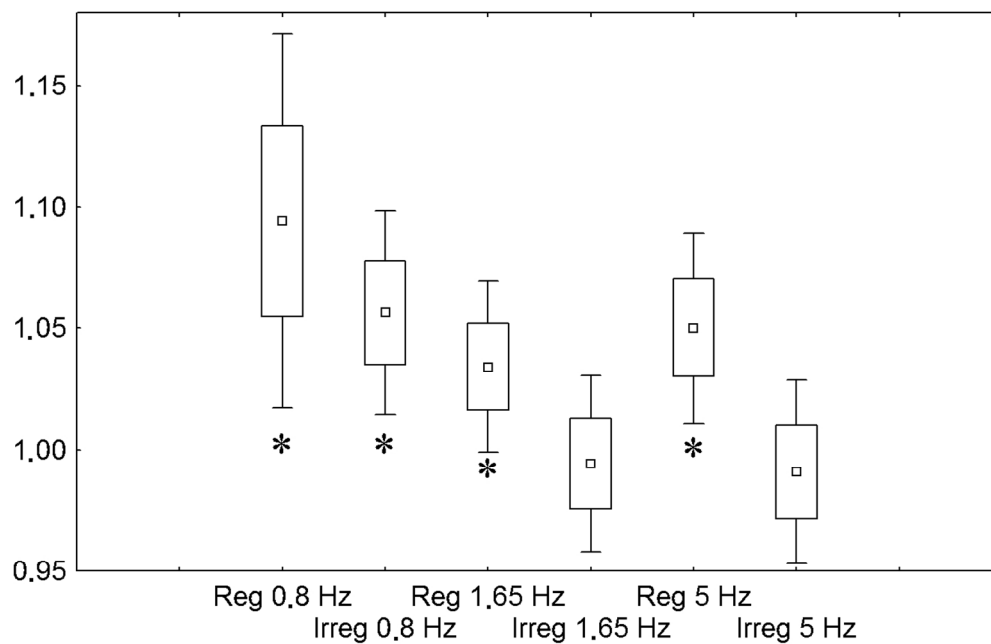


Figure 4. EEG signal. Normalized amplitude ratios at the expected frequencies for speech stimuli preceded by regular and irregular cues. Values larger than 1 indicate a peak in the spectrum. Whiskers represent the within-subject standard error from the mean.

354x234mm (96 x 96 DPI)

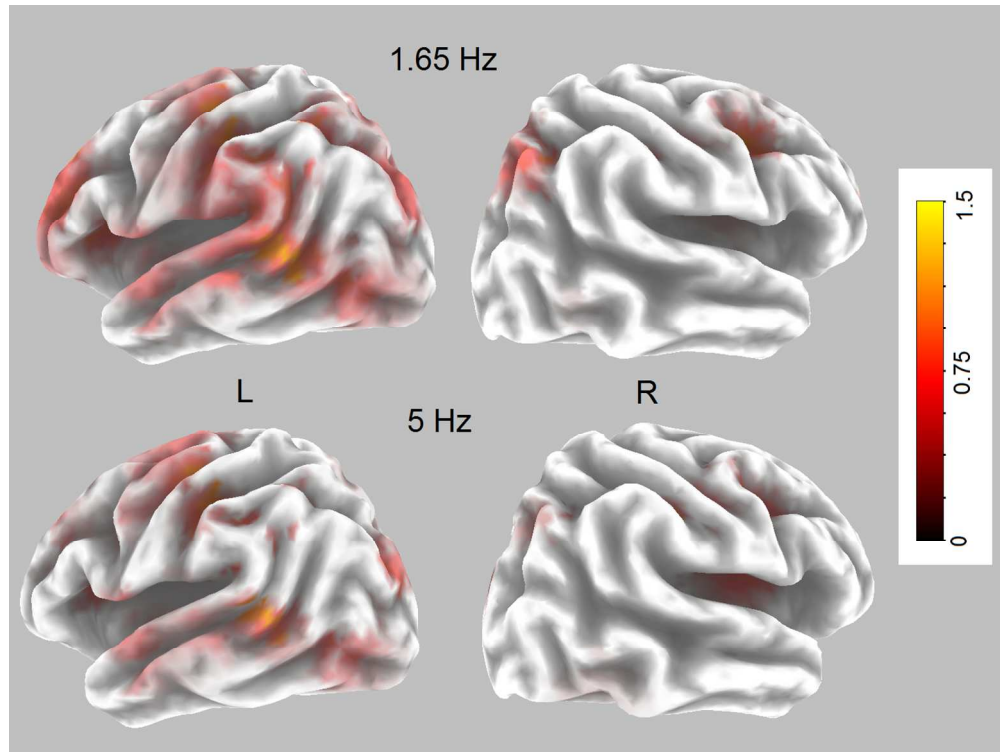


Figure 5. Statistical non-parametric maps (t-values) of the difference between speech stimuli preceded by regular and irregular cues (after bootstrapping) for the two most important target frequencies (1.65 Hz, top and 5 Hz, bottom) over left and right hemispheres

404x302mm (96 x 96 DPI)

Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

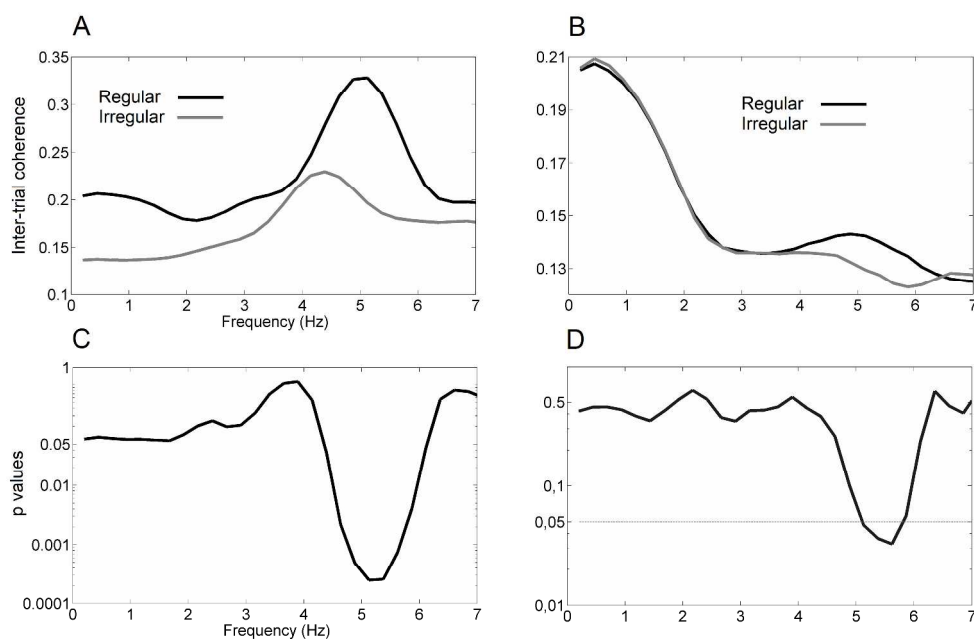


Figure 6. Top row. A: Intertrial coherence for regular and irregular cues. B: Intertrial coherence for sentences preceded by regular and irregular cues. Bottom row. C: p-values illustrating the differences between regular and irregular cues as a function of frequency. D: p-values illustrating the differences between sentences preceded by regular and irregular cues as a function of frequency.

1427x945mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

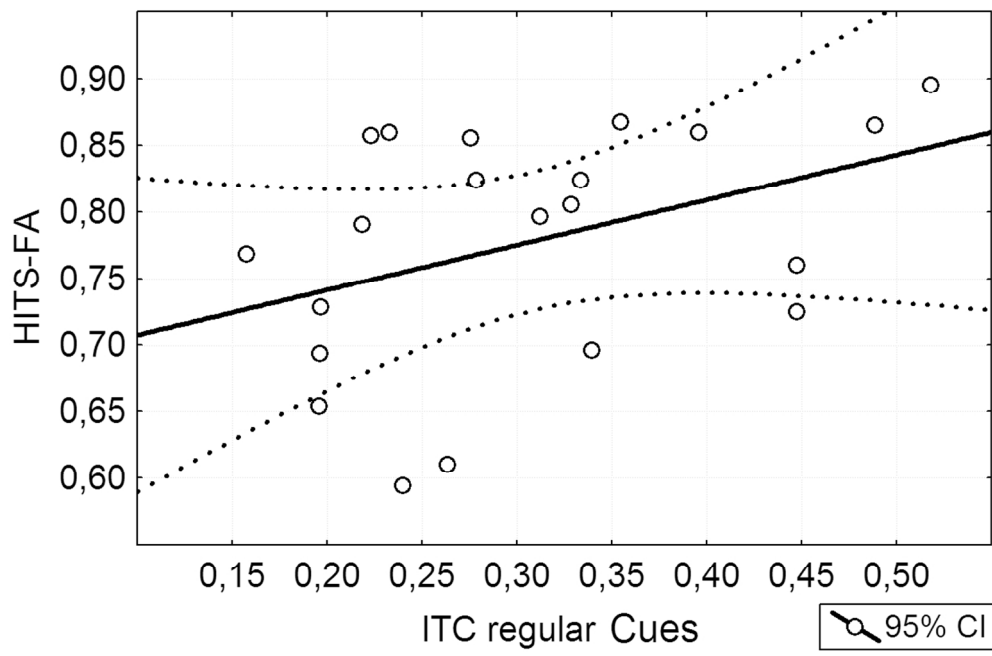


Figure 7. Scatter plot of the performance in the behavioral task (% of Hits minus % of False Alarms) versus inter-trial coherence at 5 Hz during regular cue processing (preceding the speech stimuli) for each participant. The dark line is the linear fit and the dotted lines indicate 95% confidence intervals.

106x70mm (300 x 300 DPI)

Preprint Only