Auditory and Motor Priming of Metric Structure Improves Understanding of Degraded Speech

Emma Berthault, Sophie Chen, Simone Falk, Benjamin Morillon, Daniele Schön

Title: Auditory and motor priming of metric structure improves understanding of degraded speech

Emma Berthault[1], Sophie Chen[1], Simone Falk[2], Benjamin Morillon[1], Daniele Schön[1]

[1.] Aix Marseille Université, INSERM, INS, Institut de Neurosciences des Systèmes, Marseille, France

[2.] Université de Montréal, Linguistique et Traduction, International Laboratory for Brain, Music and Sound Research, Montréal, Canada

Acknowledgments: This research was supported by grants ANR-21-CE28-0010 (DS), ANR-20-CE28-0007 (BM), ERC-CoG-101043344 (BM), ANR-16-CONV-0002 (ILCB), ANR-17-EURE-0029 (NeuroMarseille) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

Abstract

Speech comprehension is enhanced when preceded (or accompanied) by a congruent rhythmic prime reflecting the metrical sentence structure. Although these phenomena have been described for auditory and motor primes separately, their respective and synergistic contribution has not been addressed. In this experiment, participants performed a speech comprehension task on degraded speech signals that were preceded by a rhythmic prime that could be auditory, motor or audiomotor. Both auditory and audiomotor rhythmic primes facilitated speech comprehension speed. While the presence of a purely motor prime (unpaced tapping) did not globally benefit speech comprehension, comprehension accuracy scaled with the regularity of motor tapping. In order to investigate inter-individual variability, participants also performed a Spontaneous Speech Synchronization test. The strength of the estimated perception-production coupling correlated positively with overall speech comprehension scores. These findings are discussed in the framework of the dynamic attending and active sensing theories.

Keywords: human, behavior, speech, audio-motor, synchronization, coupling

1. Introduction

Both speech and music unfold in time and contain temporal patterns, although to a different degree of regularity. Thus, perceiving and making sense of these communicative and often multimodal signals requires to parse the temporal dimension in an appropriate manner. It has been proposed that a possible mechanism allowing optimal temporal parsing is a dynamic fluctuation of attention. The dynamic attending theory (DAT) suggests that attention can synchronize in time to external rhythmic events, which would allow to optimize perception by directing attention towards relevant points in time (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999).

While the DAT was developed in the domain of music perception, neurophysiological and psycholinguistic analyzes demonstrated that similar organizational principles exist for speech as well (Poeppel, 2003). The temporal constrains are less stringent in speech compared to music, nonetheless the speech signal is sufficiently rhythmic to elicit robust temporal regularities (Arnal et al., 2015; Fiveash et al., 2021). For instance, speech contains a rhythmicity between 3 and 8 Hz which is consistent across languages (Ding et al., 2017; Varnet et al., 2017).

Considering the DAT in both music and speech brought several researchers to investigate the possible interactions across domains. This resulted in several findings showing that a periodic musical prime can facilitate language grammatical processing in children with specific language impairment, dyslexia as well as typically developing children and adults (Canette et al., 2019). Other studies showed an effect of an informative rhythmic prime (that matched the prosodic structure of subsequent sentences) on speech processing, in healthy participants (Cason & Schön, 2012) but also in children with hearing impairment (Cason, Hidalgo, et al., 2015). These effects seem to be mediated by increased stimulus-brain coupling at periodicities that are present in both the rhythmic cue and speech (Falk, Lanzilotti, et al., 2017).

In the DAT, the peaks and troughs of temporal allocation are usually directly indexed to the temporal regularities of the auditory input. A complementary explanation is that temporal attention involves the motor system. This framework, known as active sensing (Kleinfeld et al., 2006; Schroeder et al., 2010), is originally based on the fact that during the perception of a sensory flow, the motor system controls the orienting of sensing organs (e.g., ocular saccades,

sniffing, whisking). In hearing research, auditory processing is generally considered as disconnected from movement, but the motor system is seen as playing a major role in determining the temporal priors necessary for auditory processing. More precisely, rhythmic movements seem to engage a top-down modulation that sharpens auditory representations. In other words, the motor system would play a role in determining the temporal predictability of the auditory sequence by modulating temporal attention and improving the segregation between relevant and distracting information (Morillon et al., 2015; Zalta et al., 2024).

Cyclic fluctuations of attention induced by overt rhythmic motor activity improve the segmentation of auditory information (Morillon et al., 2014; Morillon & Baillet, 2017) and this effect scales with motor rhythmic precision (Zalta et al., 2020). Importantly, the motor cortex is not only involved in auditory perception but in speech perception as well (Du et al., 2014; Keitel et al., 2018; Morillon et al., 2019; Wilson et al., 2004) and a facilitatory effect of rhythmic movements on speech perception has been reported (Cason, Astésano, et al., 2015; Falk, Volpi-Moncorger, et al., 2017; Falk & Dalla Bella, 2016).

Here, we investigate the tenets of dynamic attending and active sensing in speech comprehension. To this aim, participants performed a speech comprehension task preceded by the absence or presence of a rhythm that could be auditory, motor or audiomotor. Specifically, we presented participants with spectrally degraded speech stimuli containing a strong metrical regularity at the prosodic level. Speech stimuli were preceded or not by an informative rhythmic prime that matched the metrical structure of the subsequent sentence. When the rhythmic prime was present, it could be auditory, motor or audiomotor. If the premises of the DAT hold, we expect to observe a better performance following an auditory (or audiomotor) prime compared to the silent condition. If the premises of active sensing hold, we expect to observe a better performance with a motor (or audiomotor) prime compared to the silent condition.

Moreover, we also evaluated the relation between degraded-speech comprehension and the strength of speech perception-production coupling. At this aim, participants performed a Spontaneous Speech Synchronization test (SSS-test -Assaneo et al., 2019; Lizcano-Cortés et al., 2022). This test assesses the ability of participants to synchronize the repetition of a syllable [ta] with a heard syllable train. Results of this test show a bimodal distribution within the general population, with the presence of high and low synchronizers, and these differences have been linked to neurophysiological and anatomical differences (Assaneo et al., 2019;

Lizcano-Cortés et al., 2022). In addition, high synchronizers have shown a significant learning advantage in a phonological word-learning task and on statistical learning (Assaneo et al., 2019; Orpella et al., 2022). The use of this test thus allows to assess here whether inter-individual variability in sensorimotor coupling is related to differences in degraded-speech comprehension as well as whether this relation depends upon (interacts with) the different priming modalities (auditory, motor and audiomotor).

2. Methods

2.1 Participants

Twenty-two French participants (11 females, mean age = 23, SD = 2, 6 left-handed), mainly university students, took part in this study. All had normal hearing and normal or corrected vision. All gave informed consent to participate in the study. Participants received an adequate remuneration at the end of the experiment. All participants performed two tasks: the speech comprehension task and the Spontaneous Speech Synchronization test (SSS-test).

2.2 Speech comprehension task

2.2.1 Stimuli

The linguistic material used was the same as the one used by Falk and collaborators (Falk, Lanzilotti, et al., 2017). This linguistic material consisted of 63 spoken French utterances sharing the same syntactic structure, that is, two short main uncoordinated phrases featuring a simple subject-predicate-object structure (three of them were dedicated to the training session). Every utterance comprised 20 syllables subdivided in four accentual phrases of five syllables each (marked by vertical bars in the example below):

le fils du marchand | il crie dans la rue | il veut écouler | les fruits aux clientes

the merchant's son | he shouts in the street | he wants to sell | his fruit to the clients

The utterances were recorded (44.1 kHz, 16 bit) by a native French female speaker. The utterances were cued before each recording by a metronome to be read at a regular pace (i.e., 600 msec inter-onset intervals between accented syllables). When necessary, manual corrections were done (using PRAAT; (Boersma & Weenink, 2001) to obtain a highly regular

meter with an average inter-onset interval of 600 msec (±20 msec) between accented syllables (measured between the "p-centers" of inter-accent-intervals, using the algorithm of Cummins & Port; Cummins & Port, 1998). This metric regularity creates quasi regular recurrences at the syllabic level (~5 Hz), very regular recurrences at the accentual level (1.66 Hz) as well as at the level of the accentual phrase (i.e., phrase-final accents, 0.8 Hz). The average utterance duration was 4.8 sec (range = 4.7–5 sec). These utterances were then spectrally degraded with a 1.6 kHz low-pass filter to make participants' auditory perception and comprehension more difficult, while preserving the temporal dynamics of the sentences. The specific degradation parameters were based on the literature (Avilala et al., 2010) as well as on a pilot study showing a comprehension performance around 70-80% correct.

2.2.2 Procedure

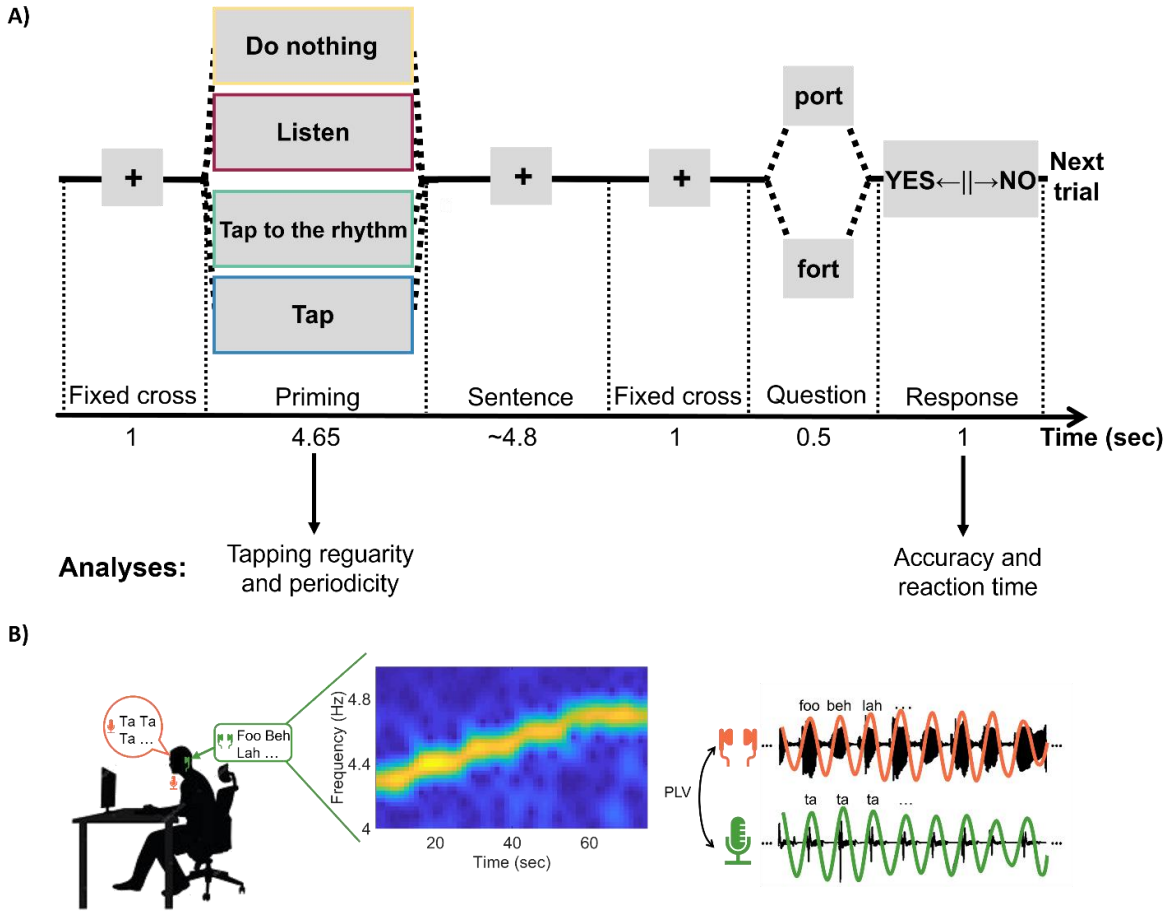
Participants were comfortably seated at 80 cm from a screen. They were instructed to listen to degraded verbal sentences and to decide whether a target word presented subsequently was present or not in the sentence (Fig. 1A). Before each sentence, they would look at the screen and either stay still, listen to a sound, tap with a sound, or tap at their own tempo on the keyboard. When asked to tap, participants were instructed to stop tapping when the instruction was replaced by a fixation cross (before the sentence started).

After a short training session (3 stimuli, 12 trials), the experiment began. During the experiment, 60 speech stimuli were presented four times in four sessions separated by a short break. In each trial, a prime with visual instruction was presented (auditory, motor, audiomotor or silent), immediately followed by a speech utterance (Fig. 1A). Then, one second sec after the end of the speech stimulus, a target word (noun or verb) was presented on the screen for 0.5 s. Participants were asked to press as quickly as possible one of two buttons (right or left arrow of the keyboard) to decide whether the word was present in the previously heard utterance or not. To avoid repetition across sessions, four different words were used as target words for each utterance. Half of the words were present in the speech utterance, whereas the other half were not. In the latter case, target words had a high phonetic similarity with a word that was present in the utterance (e.g., fort/port).

Each utterance was preceded by a different sensorimotor prime lasting 4.65 seconds: auditory, motor, audiomotor, silent. In the motor prime condition, a single sound (pizzicato of a cello followed by silence) signalled participants that they had to start tapping at their preferred pace (index on the upper arrow of the keyboard). In the audiomotor prime condition, participants

were asked to tap on the keyboard in rhythm with a 1.66 Hz isochronous sequence, composed of eight repetitions of a same sound (pizzicato of a cello, IOI = 600 ms). The same sequence was used as an auditory prime to which participants would only listen without any motor response. In the silent prime condition, participants were instructed to look at the instruction on the screen for the priming duration (Fig. 1A). The type of prime changed every four trials (mini-block). For the auditory and audiomotor primes, the interval between the rhythmic sound and the speech stimulus was manipulated in such a way that the stimulus onset asynchrony between the last note of the prime and the first accent in the utterance was always 600 msec (hence, the beat was kept constant and uninterrupted between primes and speech stimuli).

Along the four sessions, each specific speech stimulus was heard in the four conditions (i.e. preceded by the four primes). The distribution of speech stimuli across mini-blocks and the order of conditions (auditory, motor, audiomotor, silent) were pseudo-randomized and counterbalanced across participants. All answers were given with the right hand. The task was programmed on Presentation software (Neurobehavioral Systems, Berkeley, CA). Two loudspeakers (Q Acoustics Q3050) were used for sound presentation. For all participants, the volume was set at the same levels (~ 70 dB).



172

FIGURE 1. Experimental design. (A) Trial time course. A fixation cross is presented; the instruction is displayed on the screen. When a rhythm is used as prime, it can be auditory, motor or audiomotor. A degraded sentence immediately follows the prime. After the sentence a 1 second silent is followed by a word that is visually presented. This word could be present or absent in the preceding sentence. Participants finally perform a two-alternative force-choice (yes/no) task. (B) Spontaneous speech synchronization test adapted from Assaneo et al, 2019. On the left: experimental set-up. On the center: spectrogram of the stimulus speech envelope showing the increasing syllabic rate. On the right: example of the perceived (upper panel) and produced (lower panel) signals. Orange line: Envelope of the perceived signal bandpass filtered between 3.5–5.5 Hz. Green line: Envelope of the produced signal bandpass filtered between 3.5–5.5 Hz. PLV means Phase Locking Value.

173

174

175

2.2.3 Data analysis of the speech comprehension task

2.2.3.1 Tapping

The aim of the analysis of the tapping data (motor and audiomotor conditions) was to assess whether language task performance was influenced by the preceding tapping behavior. More precisely we estimated 1) the regularity of the tapping (low vs high), 2) the distance of the last tap relatively to the expected tap and 3) the tapping frequency (inter-tapping interval close or far from the metrical structure of the sentences, that is 600 msec). In the prime period (4.65 sec) participants produced ~8 inter-tapping intervals (ITI). We computed their standard deviation (SD), assessing the regularity (SD of ITI) of the tapping behavior. Finally, for each participant, we transformed these continuous variables into categorical variables. For regularity we used a median split yielding trials with high and low regularity. For the position of the last tap of each trial, we used a median split yielding taps close or further apart from the expected tap (i.e., 600 msec before the first primary stress of the subsequent sentence, for the last tap). For the tapping frequency, after removing a few outlier values (3 inter-quartile ranges below or above the median) we used a median-split yielding frequencies close or further apart from the expected tapping frequency (i.e., 1.66 Hz or 600 msec).

Logistic regression and linear mixed models were used to model the relation between comprehension performance and tapping behaviour for the motor and audiomotor conditions: glmer(accuracy ~ tapping_regularity +(1 |subject); lmer(RT ~ tapping_regularity + (1|subject); and similarly for the position of the last tap and the tapping frequency. We also assessed the possible effect of learning throughout the experiment by modelling the four experimental blocks. Since no learning was visible at this level, we did not consider further this variable.

2.2.3.2 Accuracy and reaction times

Participants' response accuracy was scored 1 for hits and 0 for miss and incorrect responses. Reaction times (RTs) corresponded to the duration in milliseconds between the onset of the (visual presented) target word and participants' response. RTs were only analyzed for correct responses. Trials with values lower or greater than two and a half standard deviations from the mean were excluded from the RT analysis (~4.2%).

We computed all statistical analyses using R (R Core Team, 2021) with lme4 (Bates et al., 2015) and lsmeans (Lenth, 2016) packages. For accuracy, we computed a logistic regression

to explain accuracy as a function of conditions, with subject as the random effect: glmer(accuracy ~ conditions + 1|subject, family=binomial). This model was compared to the null model glmer(accuracy ~ 1 + 1|subject). Statistical significance of the fixed effect was assessed by model comparison using the Akaike Information Criterion, thus arbitrating between complexity and explanatory power of the models. For RTs, we repeated the same steps using linear mixed models with subject as the random effect: lmer(RTs ~ conditions + 1|subject). Normality and homoscedasticity of the residuals of all the models were systematically visually inspected. Post-hoc comparisons were corrected for multiple comparisons using the Tukey test (lsmeans package; Lenth, 2016). Post-hoc power estimates were carried using the simr R package (Green & MacLeod, 2016) to ensure reasonable statistical power (greater than 80%).

2.3 Spontaneous Speech Synchronization (SSS) test

2.3.1 Stimuli

The audio material used was the same as the one used for the Accelerated Explicit Version of the SSS-test by Assaneo and collaborators (Assaneo et al., 2019; Lizcano-Cortés et al., 2022). This linguistic material consisted of three audio files. Each audio was used in a different part of the SSS-test. One was used for volume adjustment and was composed of a train of 16 synthesized syllables randomly concatenated but reversed in time. One was used for training (~ 10 sec) and was composed of a train of synthesized syllables « ta ». Frequency of occurrence of « ta » was 4.3 Hz. For the main task, a train of 16 synthesized syllables was randomly concatenated with an increasing syllabic rate, starting at 4.3 Hz and increasing in steps of 0.1 Hz every 10 sec until it reached 4.7 Hz, for a total duration of 80 sec (Fig. 1B).

2.3.2 Procedure

Participants seated in front of a computer and wore a lapel microphone. In-ear-speakers were used for sound presentations. First, participants adjusted the volume while simultaneously listening to an audio sequence and whispering the syllable « ta ». Participants gradually increased the volume until they could not hear their own whisper while still being at a comfortable level. This volume was then applied during the rest of the SSS-test. In a second part, participants did a short training wherein they first passively listened to the audio sequence and then whispered the syllable « ta » at the same rate for 10 sec. Finally, the main SSS task consisted of listening to the audio sequence of syllables while whispering the

syllable « ta ». Participants were explicitly instructed to synchronize the « ta » whisper with the audio stimulus. The training and the main task were repeated twice.

2.3.3 Data Analysis

We analyzed only the data of the main task. First, to improve signal to noise ratio, we applied a noise reduction to the audio recordings using Audacity software (version 3.1.3, Audacity Team,2021). Then, for each run, we computed the phase locking value (PLV) between the envelope of the produced speech and the envelope of the syllabic stream (range: 0-1). Following cochlear filtering, the envelope was estimated as the absolute value of the Hilbert transform, then averaged across bands. Envelopes were resampled at 100 Hz, filtered between 3.5 and 5.5 Hz and their phases were extracted by means of the Hilbert transform. The PLV was computed for windows of 5 seconds length with an overlap of 2 seconds. The results for all time windows were averaged providing one PLV value for each of the two blocks. Then the PLV was averaged across blocks, yielding a global perception-production coupling strength (see Assaneo et al., 2019 for more details). One participant was excluded due to very inconsistent PLV in the two runs.

Because we were interested in a potential link between PLV of participants and their performances during the speech comprehension task, we computed a linear model to explain accuracy as a function of PLV: lm(accuracy ~ PLV). We used a similar approach for RTs, and mean frequency and regularity of the tapping. Normality and homoscedasticity of the residuals of all the models were systematically visually inspected.

3. Results

3.1 Auditory primes facilitate speech comprehension speed

The overall accuracy was 74.5% (range = 58-89; Fig. 2). Participants were able to perform the task above chance while not being at ceiling (M= 74.5, SD = 8.97, chance: $t(21) = 12.8$; ceiling $t(21) = -13.3$; all $p < .001$). The prime type did not significantly affect accuracy (all: $|\beta| < 0.1$, $|SE| = 0.1$, $p > .6$; Fig. 2). By contrast, the prime type did affect reaction times: RTs were faster in the auditory and audiomotor prime conditions compared to both the silent prime condition (auditory vs. silent: $\beta = 47.1$, $SE = 11.5$, $t = 4.1$, $p < .001$; audiomotor vs. silent: $\beta = 47.4$, $SE = 11.5$, $t = 4.1$, $p < .001$) and the motor prime condition (auditory vs. motor: β = -

38.3, SE = 11.6, t = -3.3, p = .005; audiomotor vs. motor: β = 38.6, SE = 11.6, t = 3.3, p = .005). RTs in the silent and motor prime conditions did not differ (β = 8.8, SE = 11.7, t = 0.7, p = .87) and RTs in the auditory and audiomotor prime conditions did not differ (β = 0.3, SE = 11.4, t = 0.02, p = 1).

3.2 Motor tapping precision modulates speech comprehension accuracy

Concerning the tapping behaviour, as expected, participants tapped with a frequency closer to the metronome (1.66 Hz; 600ms) during the audiomotor prime compared to the motor one (audiomotor vs. motor: β = 35.8, SE = 3.23, t = 11.08, p < .001; motor: M = 564 msec, SD = 76 msec; audiomotor: M = 587 msec, SD = 51 msec). By contrast, participants tapped more regularly during the motor prime compared to the audiomotor one (audiomotor vs. motor: β = -12, SE = 5.83, t = 11.08, p = .04; SD of ITI: motor: M = 69.9 msec, SD = 29.2 msec; audiomotor: M = 81.9 msec, SD = 29.2 msec).

The regularity of the tapping behavior in the audiomotor condition did not affect performance

281 in the language task (accuracy: $\beta = 0.05$, SE = 0.1, $z = 0.4$, $p = .69$ and RT: $\beta = 17.2$, SE =

16.3, t = 1.1, p = .29). However, in the motor condition more regular tapping (i.e., smaller SD

of ITI) was associated with better accuracy (β = -0.3, SE = 0.12, z = -2.5, p = .014; Fig. 3) compared to less regular tapping. Of note, the mean accuracy in the regular tapping trials in

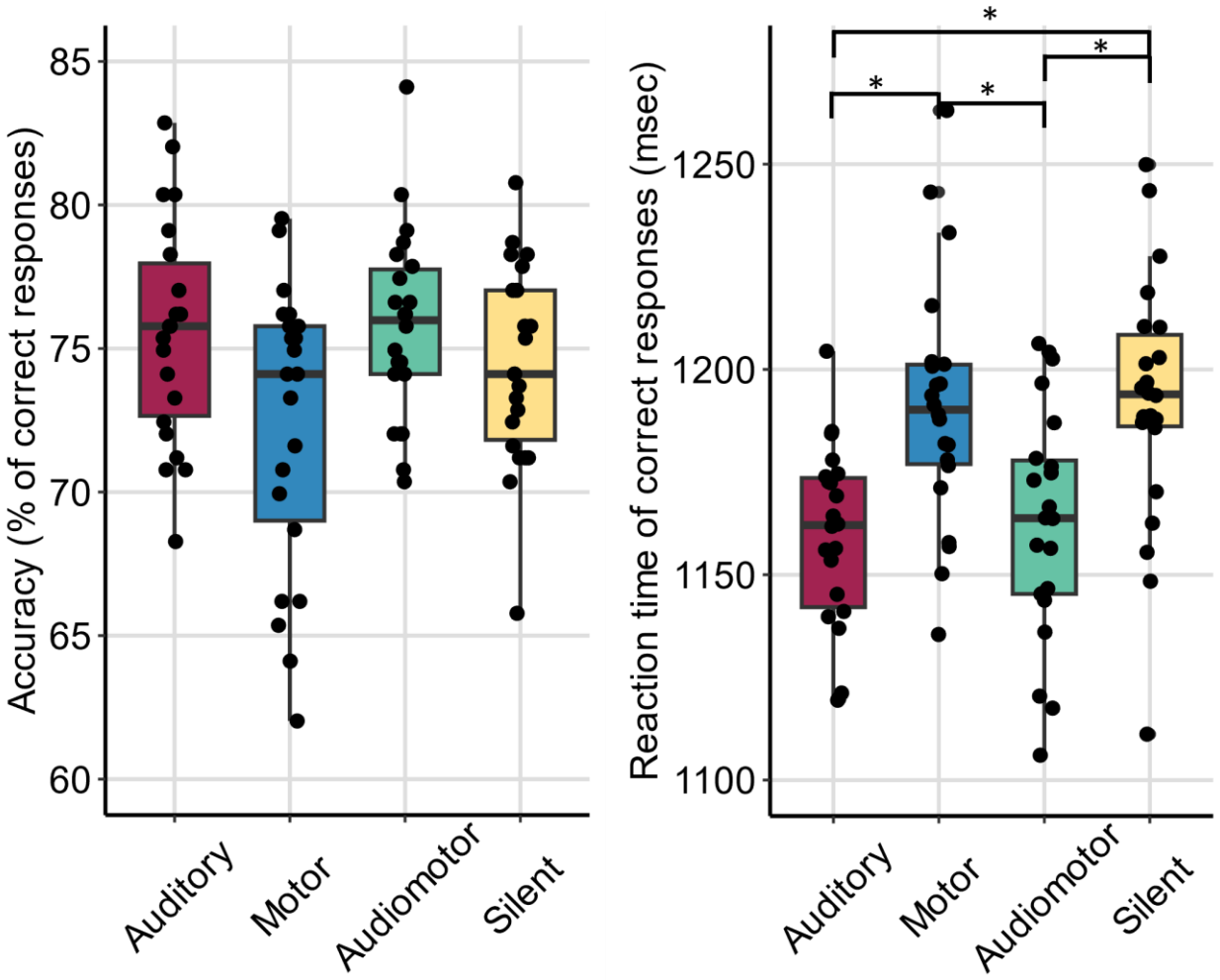


FIGURE 2. Boxplots of accuracy and correct reaction times (in millisecond) in the speech comprehension task for sentences preceded by auditory, motor, audiomotor or silent primes. Accuracy indicates the percentage of correct responses to the target words. Dots represent individual participants. Stars indicate significant effects ($p < 0.05$; $n = 22$).

the motor condition (75%) is not higher than the accuracy of the other conditions (76%, 76% and 74% for auditory, audiomotor and silent conditions). By contrast, the accuracy in the irregular tapping trials (in the motor condition, 69%) is lower than the overall accuracy of the auditory and audiomotor conditions ($p = 0.02$ for both comparisons). Finally, during motor prime and audiomotor prime no effects on performance were observed for the distance of the last tap relative to the expected tap (RTs both conditions: $p > .25$; accuracies both conditions: $p > .57$), nor for the distance of the tapping frequency relative to the metrical structure of speech (RTs both conditions: $p > .18$; accuracies both conditions: $p > .07$).

3.3 Spontaneous speech synchronization strength correlates with speech comprehension accuracy

As expected, the degree of synchronization in the SSS-test (PLV) varied across participants (range = 0.19-0.84). Estimating the relation between this perception-production coupling strength and the performance accuracy in the speech comprehension task revealed a positive relation: participants with higher PLV in the synchronization task performed better in the speech comprehension task ($R^2 = .25$, $p = .022$; Fig. 4). In order to assess whether this effect differed across the four prime conditions (auditory: $R^2 = 0.23$; motor: $R^2 = 0.24$; audiomotor: $R^2 = 0.18$; silent: $R^2 = 0.19$) we modelled the interaction between PLV and conditions (accuracy ~ PLV * conditions). This was not significant, meaning that the effect was stable across the four conditions ($df = 3$, $F = .07$, $p = .97$). In contrast, the PLV was not significantly correlated to correct reaction times (condition average: $R^2 = .06$, $p = 0.27$, all individual conditions $R^2 < 0.08$, $p > .22$) nor to the nor to the mean frequency or regularity of the tapping (for both audiomotor and motor $p > .2$, for both frequency and regularity).”

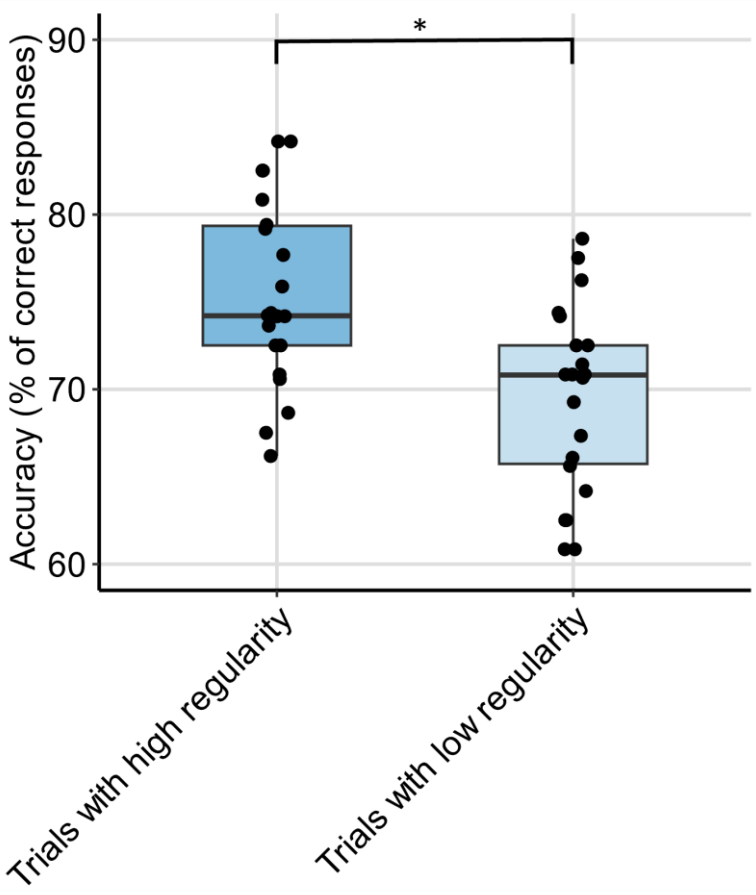


FIGURE 3. Boxplots of accuracy in the motor prime condition of the speech comprehension task, as a function of the tapping regularity. Same conventions as in Fig. 2.

309

310

311

312

313

314

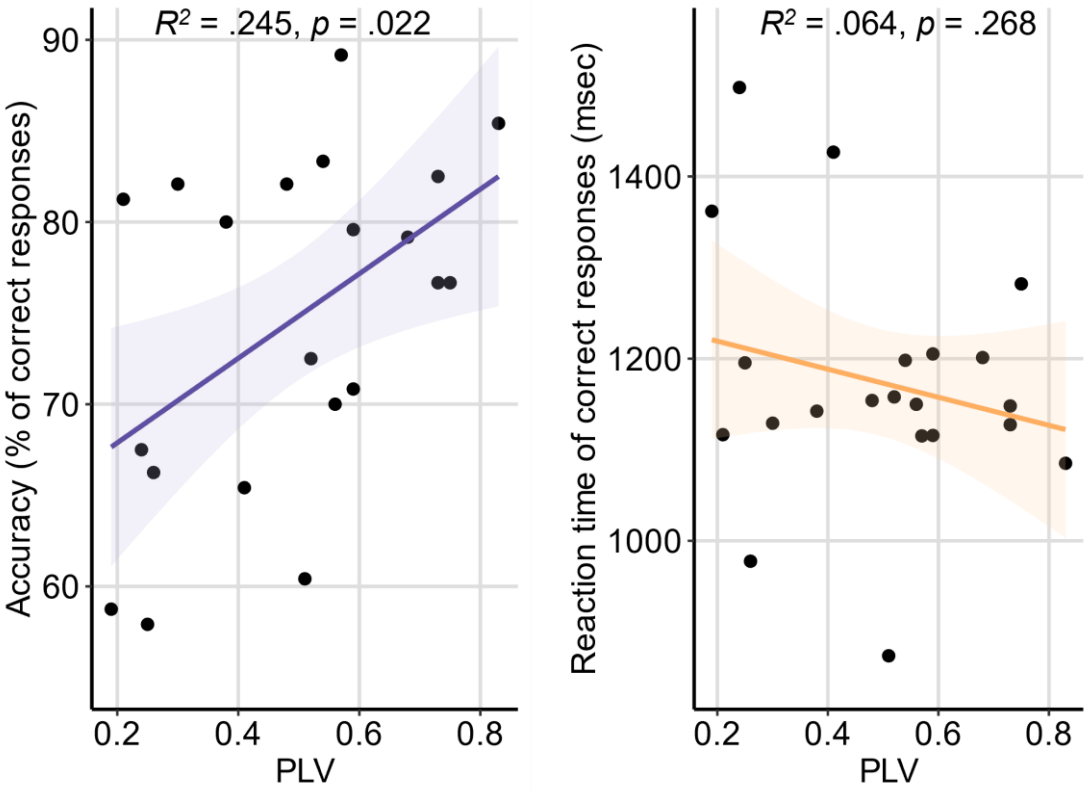


FIGURE 4. Scatter plots of speech comprehension scores as a function of speech perception-production coupling strength (PLV). Scores corresponds to (left) accuracy, i.e., the percentage of correct responses to target words in the speech comprehension task or (right) correct reaction times (in milliseconds). PLV corresponds to the coupling strength between speech production and perception in the SSS-test. Dots represent individual participants.

4. Discussion

The aim of this study was to investigate the predictive power of dynamic attending (DAT) and active sensing theories in a speech comprehension experiment. We tested whether the presence of an informative temporal prime, either auditory and/or motor, facilitates speech comprehension. Participants listened to spectrally degraded rhythmic speech that could be preceded by auditory, motor or audiomotor rhythms. A target word was then presented on the screen and participants had to decide whether it was present in the previously heard utterance. Compared to a silent control condition, auditory and audiomotor rhythmic primes matching

the prosodic rhythm of the utterances reduced reaction times to correctly detect target words (Fig. 2). Furthermore, in the motor prime condition, accuracy scores were affected by movement quality, with greater tapping regularity resulting in improved speech comprehension (Fig. 3).

Overall, the benefit of perceptual rhythmic primes (auditory and audiomotor, Fig. 2) fits well with previous findings showing the behavioral benefit of a musical prime on speech and language processing in the general adult population (Cason, Astésano, et al., 2015; Cason & Schön, 2012; Falk & Dalla Bella, 2016) as well as in participants with hearing or language disorders (Bedoin et al., 2016; Canette et al., 2019; Cason, Hidalgo, et al., 2015; Przybylski et al., 2013). This result is in line with the predictions of the DAT (Large & Jones, 1999). The regularity in the auditory prime informs about the strong metrical structure of the subsequent sentence, which may provide predictable temporal cues allowing to direct attention at salient moments of the speech stream (Arnal & Giraud, 2012; Pitt & Samuel, 1990; Port, 2003; Rimmele et al., 2018; Zion Golumbic et al., 2012).

In a previous work using a similar design, the authors were not able to observe a clear behavioral advantage of auditory rhythmic priming, on accuracy or RT (Falk, Lanzilotti, et al., 2017). In the present study, the effect on RT is now visible thanks to the speech degradation procedure that renders the task more challenging and prevents ceiling effects. Thus, while participants may have accumulated more evidence on the target word leading to shorter RTs in the primed conditions (Ratcliff & McKoon, 1997), it is a possibility that noise at the perceptual level may still not be high enough to show effects on accuracy, which would require to work at overall more difficult perceptual conditions than what was chosen here (75% accuracy), by increasing speech degradation to operate closer to chance level (50 %).

Although, at first sight, a motor prime does not lead to speech comprehension improvement, a difference was observed between trials with more regular and less regular motor tapping (Fig. 3). Previous studies using the same metrically regular stimuli reported a benefit of motor alignment to a rhythmic prime on verbal processing (Falk, Volpi-Moncorger, et al., 2017; Falk & Dalla Bella, 2016). Here, since no external cue informing about the metrical structure of the subsequent sentence was given to guide the tapping behavior in the motor prime condition, this entails that the quality of the self-generated rhythmic movement modulates the processing quality of the subsequent degraded sentence. A key dimension of these unpaced taps is their regularity, with speech comprehension being significantly improved when

movements are more regular compared to less regular tapping (Fig. 3). During perception of melodies, the quality of motor tapping (its precision) is also correlated with performance accuracy (Morillon et al., 2014). These results show that rhythmic motor activity improves the segmentation of (subsequent) auditory information through top-down influences that sharpen sensory representations, enacting auditory active sensing (Morillon et al., 2015). However, these findings should be considered in the context of the usage of highly regular (speech or melodic) stimuli and their generalization should be assessed by further experiments using less regular natural speech streams (Ding et al., 2017; Varnet et al., 2017).

Interestingly, the regularity of the tapping has an effect only in the motor and not in the audiomotor condition. In the latter, the sound is consistently presented in a perfectly rhythmic manner. Thus, it seems plausible that the introduction of accompanying movements does not exert any discernible impact. The auditory prime remains consistently highly precise in terms of temporal predictions. In contrast, in the motor condition, the temporal predictions are bound to the tapping regularity which fluctuates across trials, influencing task outcomes (Fig.3). This suggests a potential link between enhanced comprehension and the activation of the auditory dorsal pathway through periodic (i.e. temporally regular) primes (Rimmele et al., 2018), independently of the end-effector (auditory or motor). This is also in line with the main result, namely, auditory stimuli consistently presented with perfect rhythmicity induce a greater speech comprehension advantage compared to the motor condition, which prompts less accurate rhythmicity.

Of note, in our study, the prime type changes every four trials (mini bloc design), which may also explain the absence of a significant effect on behavior at the condition level (motor vs. silent; Fig. 2) or the absence of an additive effect on behavior in the audiomotor condition (audio-motor vs. auditory; Fig. 2). Implementing efficiently the strategy that consists of involving the motor system to increase the temporal precision of auditory attention may require a longer series of trials. Indeed, previous studies showing a significant benefit of tapping on auditory perception used a long bloc design (20 or 40 trials per condition; Morillon et al., 2014; Zalta et al., 2020).

Another aim of this study was to assess whether the synchronization strength in a task requiring speech perception-production coupling would be related to performance in degraded speech comprehension. Anatomical differences in the arcuate fasciculus are indeed observed between participants with high compared to low speech perception-production coupling

(Assaneo et al., 2019). Our results show a positive correlation between the performance of these two tasks (independently of the type of prime; Fig. 4). Hence, degraded-speech comprehension abilities can be partly predicted based on perception-production coupling abilities (and vice-versa). This result is compatible with the fact that the processing of degraded speech implicates the motor cortex (Du et al., 2014; Hickok et al., 2011). This also extends to speech comprehension previous findings showing a positive association between speech perception-production coupling skills and speech rate discrimination (Kern et al., 2021), statistical learning (Assaneo et al., 2019; Orpella et al., 2022) and syllable discrimination (Assaneo et al., 2021). Moreover, this result complements a finding showing a positive correlation between speech perception-production coupling strength and accelerated speech comprehension accuracy (Lubinus et al., 2023). Interestingly, neural tracking of the speech envelope is related to rhythmic priming of speech (Falk, Lanzilotti, et al., 2017). That general perception-production synchronization abilities can predict accuracy in the speech comprehension task (Fig. 4) suggests that further synchronization measures, as for example the amount of neural tracking or of motor entrainment to the auditory prime, could help to establish a clearer link between the prime and the parsing of the heard utterances in future studies.

The relation between speech perception-production coupling and speech comprehension was equivalent across the three prime conditions (auditory, motor and audiomotor). One may expect that, because the audiomotor prime requires perception-production coupling, it could benefit more to participants with low speech perception-production coupling. Alternatively, participants with high speech perception-production coupling may be the ones who benefit from top-down effects of the motor system to enhance perception (Assaneo et al., 2021). The absence of such an effect may be due to the fact that a short prime or a mini-block design are not sufficient to change perception-production coupling. Consistent with the findings reporting anatomical differences between good and poor speech synchronizers (Assaneo et al., 2019), our finding that poor synchronizers have poorer degraded speech comprehension point to the fact that degraded speech more heavily relies on perception-production coupling, similarly to what has been hypothesized for some language developmental disorders (Fiveash et al., 2021; Goswami, 2011). In this perspective, our design may not be enough to change neural dynamics that are subtended by anatomical differences, and longer stimulation protocols may be necessary to improve speech-in-noise processing in poor synchronizers, such as those used in clinical settings (Flaugnacco et al., 2015; Hidalgo et al., 2019).

To conclude, our study provides new evidence for a facilitatory effect of auditory (and audiomotor) rhythmic priming of speech, at least in presence of a strong metrical structure. It also shows that motor priming (tapping) can affect speech comprehension and that this depends upon the regularity of the tapping dynamics. Finally, perception-production coupling abilities are related to overall performance accuracy in degraded speech comprehension.

DECLARATIONS OF CONFLICT OF INTEREST

None.

References

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. Trends in Cognitive Sciences, 16(7), 390-398. https://doi.org/10.1016/j.tics.2012.05.003

Arnal, L. H., Poeppel, D., & Giraud, A.-L. (2015). Temporal coding in the auditory cortex. Handbook of Clinical Neurology, 129, 85-98. https://doi.org/10.1016/B978-0-444-62630-1.00005-6

Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. Nature Human Behaviour, 5(1), 71-82. https://doi.org/10.1038/s41562-020-00962-0

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nature Neuroscience, 22(4), 627-632. https://doi.org/10.1038/s41593-019-0353-z

Audacity Team (2021) Audacity(R) : Free Audio Editor and Recorder [Computer application]. (3.1.3). (2022). [Logiciel]. https://audacityteam.org/

Avilala, V. K., Prabhu P, P., & Barman, A. (2010). The Effect of Filtered Speech on Speech Identification Scores of Young Normal Hearing Adults. Journal of All India Institute of Speech and Hearing, Vol 29, 115-119.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67, 1-48. https://doi.org/10.18637/jss.v067.i01

Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally Regular Musical Primes Facilitate Subsequent Syntax Processing in Children with Specific Language Impairment. Frontiers in Neuroscience, 10. https://www.frontiersin.org/articles/10.3389/fnins.2016.00245

Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. Glot international, 5, 341-345.

Canette, L.-H., Bedoin, N., Lalitte, P., Bigand, E., & Tillmann, B. (2019). The Regularity of Rhythmic Primes Influences Syntax Processing in Adults. Auditory Perception & Cognition, 2(3), 163-179. https://doi.org/10.1080/25742442.2020.1752080

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm : Rhythmic priming and audio-motor training affect speech perception. Acta Psychologica, 155, 43-50. https://doi.org/10.1016/j.actpsy.2014.12.002

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech production abilities : Evidence from prelingually deaf children. Neuropsychology, 29(1), 102-107. https://doi.org/10.1037/neu0000115

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. Neuropsychologia, 50(11), 2652-2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. Journal of Phonetics, 26(2), 145-171. https://doi.org/10.1006/jpho.1998.0070

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. Neuroscience & Biobehavioral Reviews, 81, 181-187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proceedings of the National Academy of Sciences, 111(19), 7126-7131. https://doi.org/10.1073/pnas.1318738111

Falk, S., & Dalla Bella, S. (2016). It is better when expected : Aligning speech and motor rhythms enhances verbal processing. Language, Cognition and Neuroscience, 31(5), 699-708. https://doi.org/10.1080/23273798.2016.1144892

Falk, S., Lanzilotti, C., & Schön, D. (2017). Tuning Neural Phase Entrainment to Speech. Journal of Cognitive Neuroscience, 29(8), 1378-1389. https://doi.org/10.1162/jocn_a_01136

Falk, S., Volpi-Moncorger, C., & Dalla Bella, S. (2017). Auditory-Motor Rhythms and Speech Processing in French and German Listeners. Frontiers in Psychology, 8, 395. https://doi.org/10.3389/fpsyg.2017.00395

Fiveash, A., Bedoin, N., Gordon, R. L., & Tillmann, B. (2021). Processing rhythm in speech and music : Shared mechanisms and implications for developmental speech and language disorders. Neuropsychology, 35(8), 771-791. https://doi.org/10.1037/neu0000766

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music Training Increases Phonological Awareness and Reading Skills in Developmental Dyslexia : A Randomized Control Trial. PLOS ONE, 10(9), e0138715. https://doi.org/10.1371/journal.pone.0138715

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. Trends in Cognitive Sciences, 15(1), 3-10. https://doi.org/10.1016/j.tics.2010.10.001

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution, 7(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing : Computational Basis and Neural Organization. Neuron, 69(3), 407-422. https://doi.org/10.1016/j.neuron.2011.01.019

Hidalgo, C., Pesnot, -Lerousseau Jacques, Marquis, P., Roman, S., & Sch, ön D. (2019). Rhythmic Training Improves Temporal Anticipation and Adaptation Abilities in Children

With Hearing Loss During Verbal Interaction. Journal of Speech, Language, and Hearing Research, 62(9), 3234-3247. https://doi.org/10.1044/2019_JSLHR-S-18-0349

Jones, M. R. (1976). Time, our lost dimension : Toward a new theory of perception, attention, and memory. Psychological Review, 83(5), 323-355. https://doi.org/10.1037/0033-295X.83.5.323

Jones, M. R., & Boltz, M. (1989). Dynamic Attending and Responses to Time.

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biology, 16(3), e2004473. https://doi.org/10.1371/journal.pbio.2004473

Kern, P., Assaneo, M. F., Endres, D., Poeppel, D., & Rimmele, J. M. (2021). Preferred auditory temporal processing regimes and auditory-motor synchronization. Psychonomic Bulletin & Review, 28(6), 1860-1873. https://doi.org/10.3758/s13423-021-01933-w

Kleinfeld, D., Ahissar, E., & Diamond, M. E. (2006). Active sensation : Insights from the rodent vibrissa sensorimotor system. Current Opinion in Neurobiology, 16(4), 435-444. https://doi.org/10.1016/j.conb.2006.06.009

Large, E. W., & Jones, M. R. (1999). The dynamics of attending : How people track time-varying events. Psychological Review, 106(1), 119-159. https://doi.org/10.1037/0033-295X.106.1.119

Lenth, R. V. (2016). Least-Squares Means : The R Package lsmeans. Journal of Statistical Software, 69, 1-33. https://doi.org/10.18637/jss.v069.i01

Lizcano-Cortés, F., Gómez-Varela, I., Mares, C., Wallisch, P., Orpella, J., Poeppel, D., Ripollés, P., & Assaneo, M. F. (2022). Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers. STAR Protocols, 3(2), 101248. https://doi.org/10.1016/j.xpro.2022.101248

Lubinus, C., Keitel, A., Obleser, J., Poeppel, D., & Rimmele, J. M. (2023). Explaining flexible continuous speech comprehension from individual motor rhythms. Proceedings of the Royal Society B: Biological Sciences, 290(1994), 20222410. https://doi.org/10.1098/rspb.2022.2410

Morillon, B., Arnal, L. H., Schroeder, C. E., & Keitel, A. (2019). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neuroscience and Biobehavioral Reviews, 107, 136-142. https://doi.org/10.1016/j.neubiorev.2019.09.012

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences of the United States of America, 114(42), E8913-E8921. https://doi.org/10.1073/pnas.1705373114

Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in Auditory Active Sensing. Current opinion in neurobiology, 31, 230-238. https://doi.org/10.1016/j.conb.2014.12.005

Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. Nature Communications, 5, 5255. https://doi.org/10.1038/ncomms6255

Orpella, J., Assaneo, M. F., Ripollés, P., Noejovich, L., López-Barroso, D., Diego-Balaguer, R. de, & Poeppel, D. (2022). Differential activation of a frontoparietal network explains population-level differences in statistical learning from speech. PLOS Biology, 20(7), e3001712. https://doi.org/10.1371/journal.pbio.3001712

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. Journal of Experimental Psychology: Human Perception and Performance, 16, 564-573. https://doi.org/10.1037/0096-1523.16.3.564

Poeppel, D. (2003). The analysis of speech in different temporal integration windows : Cerebral lateralization as 'asymmetric sampling in time'. Speech Communication, 41(1), 245-255. https://doi.org/10.1016/S0167-6393(02)00107-3

Port, R. F. (2003). Meter and speech. Journal of Phonetics, 31(3), 599-611. https://doi.org/10.1016/j.wocn.2003.08.001

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., Kotz, S. A., & Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. Neuropsychology, 27(1), 121-131. https://doi.org/10.1037/a0031277

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. Psychological Review, 104, 319-343. https://doi.org/10.1037/0033-295X.104.2.319

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends in Cognitive Sciences, 22(10), 870-882. https://doi.org/10.1016/j.tics.2018.08.003

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of Active Sensing and Perceptual Selection. Current opinion in neurobiology, 20(2), 172-176. https://doi.org/10.1016/j.conb.2010.02.010

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. The Journal of the Acoustical Society of America, 142(4), 1976-1989. https://doi.org/10.1121/1.5006179

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nature Neuroscience, 7(7), 701-702. https://doi.org/10.1038/nn1263

Zalta, A., Petkoski, S., & Morillon, B. (2020). Natural rhythms of periodic temporal attention. Nature Communications, 11(1), 1051. https://doi.org/10.1038/s41467-020-14888-8

Zalta, A., Large, E.W., Schön, D., Morillon, B. (2024). Neural dynamics of predictive timing and motor engagement in music listening. Science Advances, in press

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. Brain and Language, 122(3), 151-161. https://doi.org/10.1016/j.bandl.2011.12.010