



HAL
open science

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

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

► **To cite this version:**

Emma Berthault, Sophie Chen, Simone Falk, Benjamin Morillon, Daniele Schön. Auditory and Motor Priming of Metric Structure Improves Understanding of Degraded Speech. *Cognition*, In press, 16, 10.2139/ssrn.4676099 . hal-04546828

HAL Id: hal-04546828

<https://hal.science/hal-04546828>

Submitted on 15 Apr 2024

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.

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

3

4 **Emma Berthault¹, Sophie Chen¹, Simone Falk², Benjamin Morillon¹, Daniele Schön¹**

5 ¹ **Aix Marseille Université, INSERM, INS, Institut de Neurosciences des Systèmes, Marseille,**
6 **France**

7 ² **Université de Montréal, Linguistique et Traduction, International Laboratory for Brain,**
8 **Music and Sound Research, Montréal, Canada**

9

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

13

14

15 **Abstract**

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

29

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

31

33 1. Introduction

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

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

49 Considering the DAT in both music and speech brought several researchers to investigate the
50 possible interactions across domains. This resulted in several findings showing that a periodic
51 musical prime can facilitate language grammatical processing in children with specific
52 language impairment, dyslexia as well as typically developing children and adults (Canette et
53 al., 2019). Other studies showed an effect of an informative rhythmic prime (that matched the
54 prosodic structure of subsequent sentences) on speech processing, in healthy participants
55 (Cason & Schön, 2012) but also in children with hearing impairment (Cason, Hidalgo, et al.,
56 2015). These effects seem to be mediated by increased stimulus-brain coupling at periodicities
57 that are present in both the rhythmic cue and speech (Falk, Lanzilotti, et al., 2017).

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

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

70 Cyclic fluctuations of attention induced by overt rhythmic motor activity improve the
71 segmentation of auditory information (Morillon et al., 2014; Morillon & Baillet, 2017) and
72 this effect scales with motor rhythmic precision (Zalta et al., 2020). Importantly, the motor
73 cortex is not only involved in auditory perception but in speech perception as well (Du et al.,
74 2014; Keitel et al., 2018; Morillon et al., 2019; Wilson et al., 2004) and a facilitatory effect of
75 rhythmic movements on speech perception has been reported (Cason, Astésano, et al., 2015;
76 Falk, Volpi-Moncorger, et al., 2017; Falk & Dalla Bella, 2016).

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

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

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

101

102 2. Methods

103 2.1 Participants

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

109 2.2 Speech comprehension task

110 2.2.1 Stimuli

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

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

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

119 The utterances were recorded (44.1 kHz, 16 bit) by a native French female speaker. The
120 utterances were cued before each recording by a metronome to be read at a regular pace (i.e.,
121 600 msec inter-onset intervals between accented syllables). When necessary, manual
122 corrections were done (using PRAAT; (Boersma & Weenink, 2001) to obtain a highly regular

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

133 2.2.2 Procedure

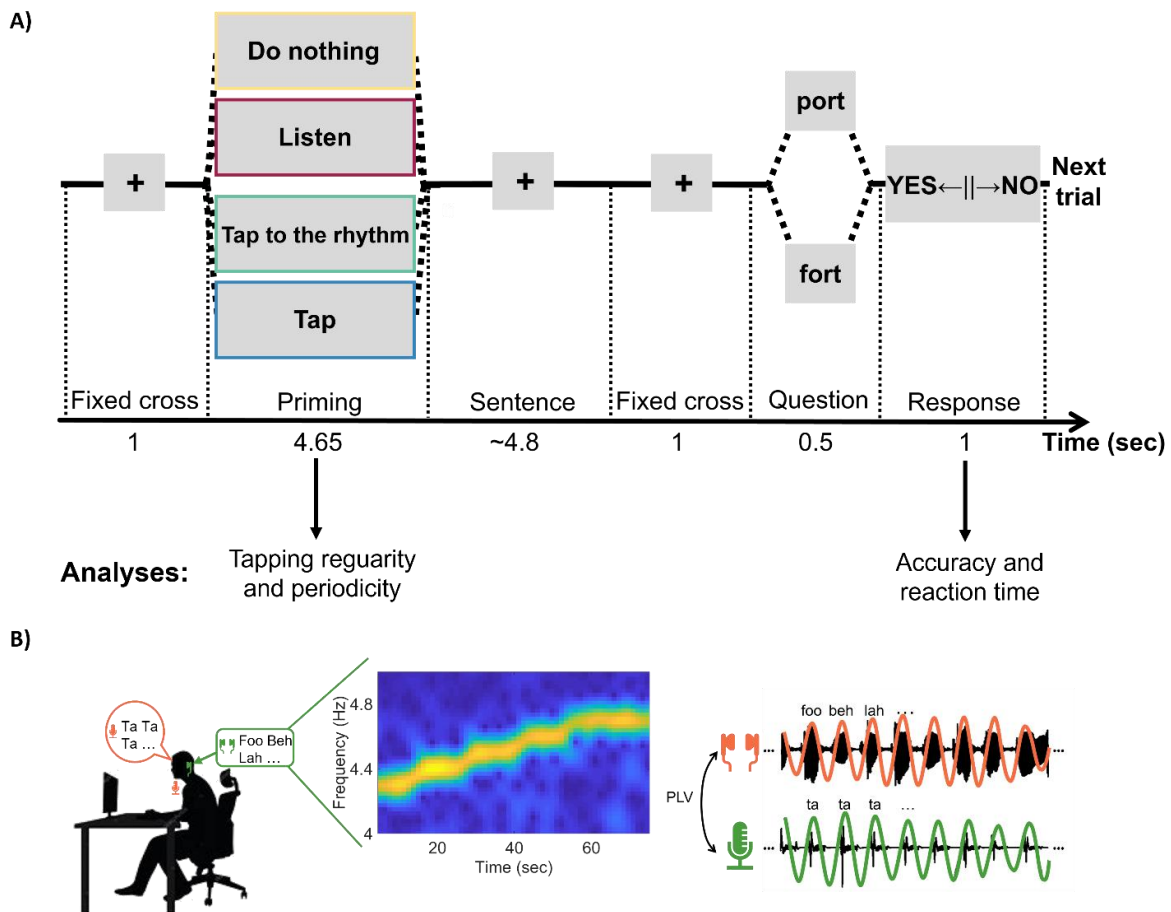
134 Participants were comfortably seated at 80 cm from a screen. They were instructed to listen
135 to degraded verbal sentences and to decide whether a target word presented subsequently was
136 present or not in the sentence (Fig. 1A). Before each sentence, they would look at the screen
137 and either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

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

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

155 were asked to tap on the keyboard in rhythm with a 1.66 Hz isochronous sequence, composed
156 of eight repetitions of a same sound (pizzicato of a cello, IOI = 600 ms). The same sequence
157 was used as an auditory prime to which participants would only listen without any motor
158 response. In the silent prime condition, participants were instructed to look at the instruction
159 on the screen for the priming duration (Fig. 1A). The type of prime changed every four trials
160 (mini-block). For the auditory and audiomotor primes, the interval between the rhythmic
161 sound and the speech stimulus was manipulated in such a way that the stimulus onset
162 asynchrony between the last note of the prime and the first accent in the utterance was always
163 600 msec (hence, the beat was kept constant and uninterrupted between primes and speech
164 stimuli).

165 Along the four sessions, each specific speech stimulus was heard in the four conditions (i.e.
166 preceded by the four primes). The distribution of speech stimuli across mini-blocks and the
167 order of conditions (auditory, motor, audiomotor, silent) were pseudo-randomized and
168 counterbalanced across participants. All answers were given with the right hand. The task was
169 programmed on Presentation software (Neurobehavioral Systems, Berkeley, CA). Two
170 loudspeakers (Q Acoustics Q3050) were used for sound presentation. For all participants, the
171 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

176 2.2.3 Data analysis of the speech comprehension task

177 2.2.3.1 Tapping

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

192 Logistic regression and linear mixed models were used to model the relation between
193 comprehension performance and tapping behaviour for the motor and audiomotor conditions:
194 `glmer(accuracy ~ tapping_regularity +(1 |subject); lmer(RT ~ tapping_regularity +`
195 `(1|subject);` and similarly for the position of the last tap and the tapping frequency. We also
196 assessed the possible effect of learning throughout the experiment by modelling the four
197 experimental blocks. Since no learning was visible at this level, we did not consider further
198 this variable.

199 2.2.3.2 Accuracy and reaction times

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

205 We computed all statistical analyses using R (R Core Team, 2021) with lme4 (Bates et al.,
206 2015) and lsmeans (Lenth, 2016) packages. For accuracy, we computed a logistic regression

207 to explain accuracy as a function of conditions, with subject as the random effect:
208 `glmer(accuracy ~ conditions + 1|subject, family=binomial)`. This model was compared to the
209 null model `glmer(accuracy ~ 1 + 1|subject)`. Statistical significance of the fixed effect was
210 assessed by model comparison using the Akaike Information Criterion, thus arbitrating
211 between complexity and explanatory power of the models. For RTs, we repeated the same
212 steps using linear mixed models with subject as the random effect: `lmer(RTs ~ conditions +`
213 `1|subject)`. Normality and homoscedasticity of the residuals of all the models were
214 systematically visually inspected. Post-hoc comparisons were corrected for multiple
215 comparisons using the Tukey test (`lsmeans` package; Lenth, 2016). Post-hoc power estimates
216 were carried using the `simr` R package (Green & MacLeod, 2016) to ensure reasonable
217 statistical power (greater than 80%).

218 2.3 Spontaneous Speech Synchronization (SSS) test

219 2.3.1 Stimuli

220 The audio material used was the same as the one used for the Accelerated Explicit Version of
221 the SSS-test by Assaneo and collaborators (Assaneo et al., 2019; Lizcano-Cortés et al., 2022).
222 This linguistic material consisted of three audio files. Each audio was used in a different part
223 of the SSS-test. One was used for volume adjustment and was composed of a train of 16
224 synthesized syllables randomly concatenated but reversed in time. One was used for training
225 (~ 10 sec) and was composed of a train of synthesized syllables « ta ». Frequency of
226 occurrence of « ta » was 4.3 Hz. For the main task, a train of 16 synthesized syllables was
227 randomly concatenated with an increasing syllabic rate, starting at 4.3 Hz and increasing in
228 steps of 0.1 Hz every 10 sec until it reached 4.7 Hz, for a total duration of 80 sec (Fig. 1B).

229 2.3.2 Procedure

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

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

240 2.3.3 Data Analysis

241 We analyzed only the data of the main task. First, to improve signal to noise ratio, we applied
242 a noise reduction to the audio recordings using Audacity software (version 3.1.3, Audacity
243 Team,2021). Then, for each run, we computed the phase locking value (PLV) between the
244 envelope of the produced speech and the envelope of the syllabic stream (range: 0-1).
245 Following cochlear filtering, the envelope was estimated as the absolute value of the Hilbert
246 transform, then averaged across bands. Envelopes were resampled at 100 Hz, filtered between
247 3.5 and 5.5 Hz and their phases were extracted by means of the Hilbert transform. The PLV
248 was computed for windows of 5 seconds length with an overlap of 2 seconds. The results for
249 all time windows were averaged providing one PLV value for each of the two blocks. Then
250 the PLV was averaged across blocks, yielding a global perception-production coupling
251 strength (see Assaneo et al., 2019 for more details). One participant was excluded due to very
252 inconsistent PLV in the two runs.

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

258

259 3. Results

260 3.1 Auditory primes facilitate speech comprehension speed

261 The overall accuracy was 74.5% (range = 58-89; Fig. 2). Participants were able to perform the
262 task above chance while not being at ceiling ($M = 74.5$, $SD = 8.97$, chance: $t(21) = 12.8$;
263 ceiling $t(21) = -13.3$; all $p < .001$). The prime type did not significantly affect accuracy (all: $|\beta|$
264 < 0.1 , $|\text{SE}| = 0.1$, $p > .6$; Fig. 2). By contrast, the prime type did affect reaction times: RTs
265 were faster in the auditory and audiomotor prime conditions compared to both the silent prime
266 condition (auditory vs. silent: $\beta = 47.1$, $\text{SE} = 11.5$, $t = 4.1$, $p < .001$; audiomotor vs. silent: $\beta =$
267 47.4 , $\text{SE} = 11.5$, $t = 4.1$, $p < .001$) and the motor prime condition (auditory vs. motor: $\beta = -$

268 38.3, SE = 11.6, $t = -3.3$, $p = .005$; audiomotor vs. motor: $\beta = 38.6$, SE = 11.6, $t = 3.3$, $p =$
269 .005). RTs in the silent and motor prime conditions did not differ ($\beta = 8.8$, SE = 11.7, $t = 0.7$,
270 $p = .87$) and RTs in the auditory and audiomotor prime conditions did not differ ($\beta = 0.3$, SE
271 = 11.4, $t = 0.02$, $p = 1$).

272 3.2 Motor tapping precision modulates speech comprehension accuracy

273 Concerning the tapping behaviour, as expected, participants tapped with a frequency closer to
274 the metronome (1.66 Hz; 600ms) during the audiomotor prime compared to the motor one
275 (audiomotor vs. motor: $\beta = 35.8$, SE = 3.23, $t = 11.08$, $p < .001$; motor: M = 564 msec, SD =
276 76 msec; audiomotor: M = 587 msec, SD = 51 msec). By contrast, participants tapped more
277 regularly during the motor prime compared to the audiomotor one (audiomotor vs. motor: $\beta =$
278 -12, SE = 5.83, $t = 11.08$, $p = .04$; SD of ITI: motor: M = 69.9 msec, SD = 29.2 msec;
279 audiomotor: M = 81.9 msec, SD = 29.2 msec).

280 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: $\beta = 17.2$, $SE =$

282 16.3, $t = 1.1$, $p = .29$). However, in the motor condition more regular tapping (i.e., smaller SD

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

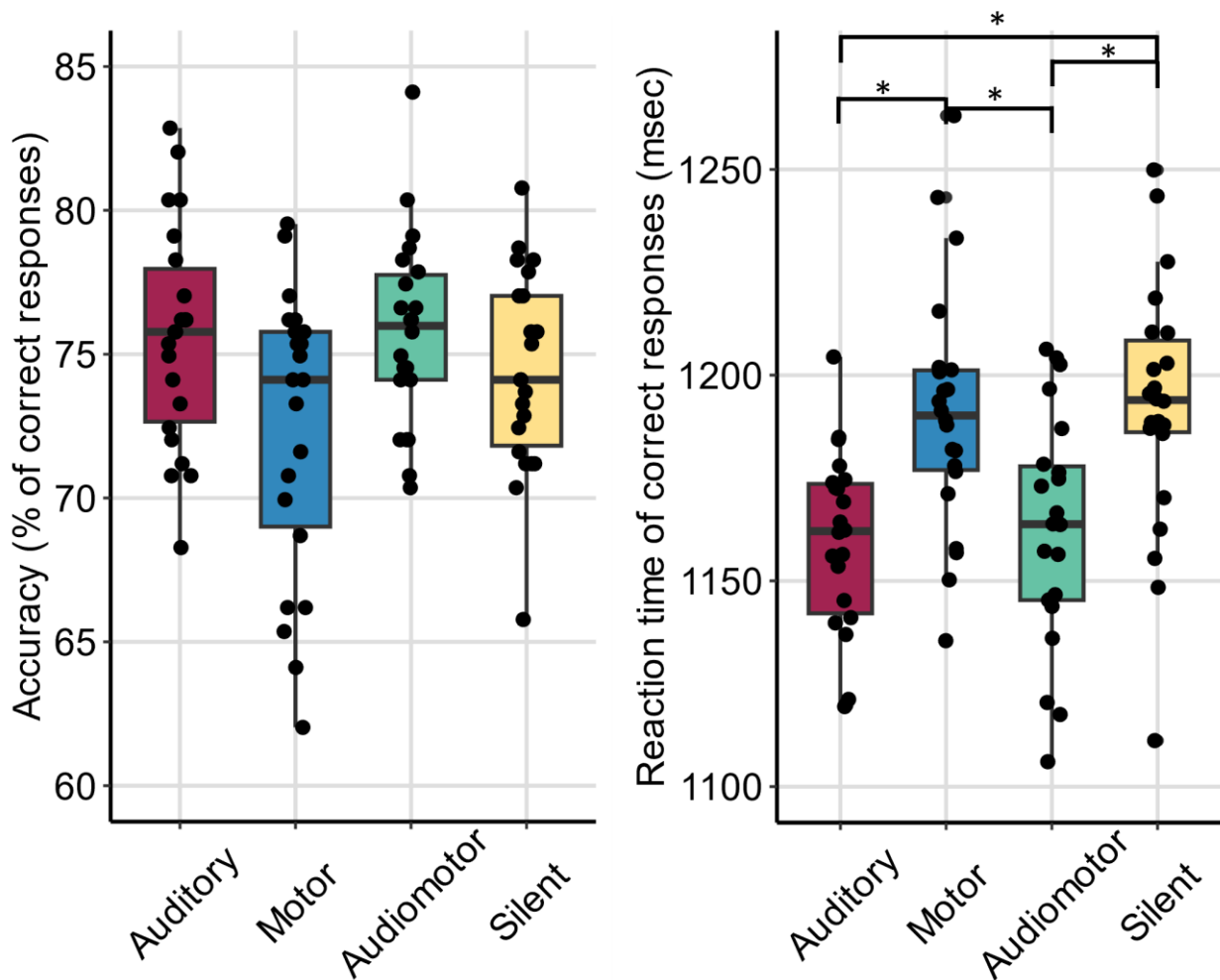


FIGURE 2. Boxplots of accuracy and correct reaction times (in millisecond) in the speech comprehension task for sentences preceded by auditory, motor, audiomotor or silent primes. Accuracy indicates the percentage of correct responses to the target words. Dots represent individual participants. Stars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

285 the motor condition (75%) is not higher than the accuracy of the other conditions (76%, 76%
 286 and 74% for auditory, audiomotor and silent conditions). By contrast, the accuracy in the
 287 irregular tapping trials (in the motor condition, 69%) is lower than the overall accuracy of the
 288 auditory and audiomotor conditions ($p = 0.02$ for both comparisons). Finally, during motor
 289 prime and audiomotor prime no effects on performance were observed for the distance of the
 290 last tap relative to the expected tap (RTs both conditions: $p > .25$; accuracies both conditions:
 291 $p > .57$), nor for the distance of the tapping frequency relative to the metrical structure of
 292 speech (RTs both conditions: $p > .18$; accuracies both conditions: $p > .07$).

293

294 3.3 Spontaneous speech synchronization strength correlates with speech comprehension
295 accuracy

296 As expected, the degree of synchronization in the SSS-test (PLV) varied across participants
297 (range = 0.19-0.84). Estimating the relation between this perception-production coupling
298 strength and the performance accuracy in the speech comprehension task revealed a positive
299 relation: participants with higher PLV in the synchronization task performed better in the
300 speech comprehension task ($R^2 = .25$, $p = .022$; Fig. 4). In order to assess whether this effect
301 differed across the four prime conditions (auditory: $R^2 = 0.23$; motor: $R^2 = 0.24$; audiomotor:
302 $R^2 = 0.18$; silent: $R^2 = 0.19$) we modelled the interaction between PLV and conditions
303 (accuracy ~ PLV * conditions). This was not significant, meaning that the effect was stable
304 across the four conditions ($df = 3$, $F = .07$, $p = .97$). In contrast, the PLV was not significantly
305 correlated to correct reaction times (condition average: $R^2 = .06$, $p = 0.27$, all individual
306 conditions $R^2 < 0.08$, $p > .22$) nor to the nor to the mean frequency or regularity of the
307 tapping (for both audiomotor and motor $p > .2$, for both frequency and regularity).”

308

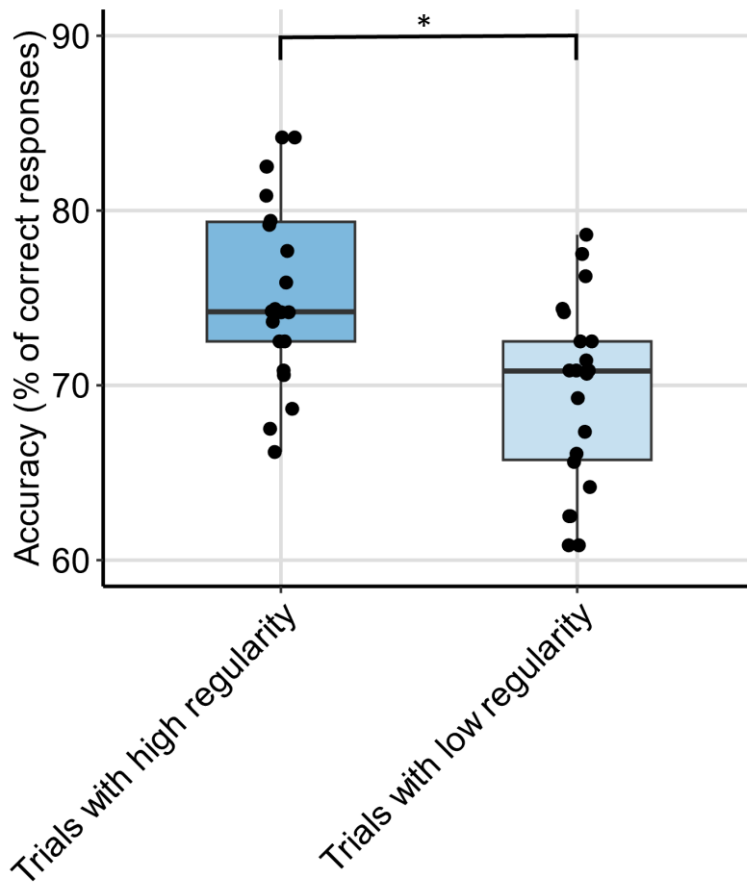


FIGURE 3. Boxplots of accuracy in the motor prime condition of the speech comprehension task, as a function of the tapping regularity. Same conventions as in Fig. 2.

309

310

311

312

313

314

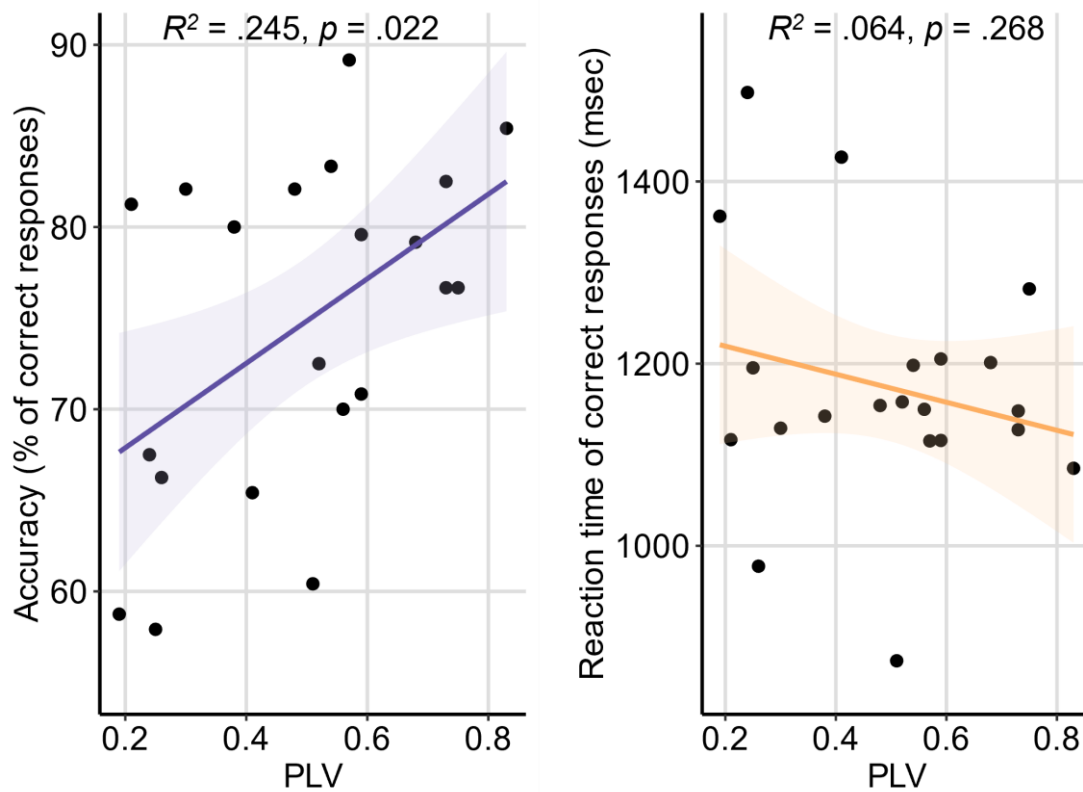


FIGURE 4. Scatter plots of speech comprehension scores as a function of speech perception-production coupling strength (PLV). Scores corresponds to (left) accuracy, i.e., the percentage of correct responses to target words in the speech comprehension task or (right) correct reaction times (in milliseconds).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

315

316

317 4. Discussion

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

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

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

339 In a previous work using a similar design, the authors were not able to observe a clear
340 behavioral advantage of auditory rhythmic priming, on accuracy or RT (Falk, Lanzilotti, et
341 al., 2017). In the present study, the effect on RT is now visible thanks to the speech
342 degradation procedure that renders the task more challenging and prevents ceiling effects.
343 Thus, while participants may have accumulated more evidence on the target word leading to
344 shorter RTs in the primed conditions (Ratcliff & McKoon, 1997), it is a possibility that noise
345 at the perceptual level may still not be high enough to show effects on accuracy, which would
346 require to work at overall more difficult perceptual conditions than what was chosen here
347 (75% accuracy), by increasing speech degradation to operate closer to chance level (50 %).

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

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

365 Interestingly, the regularity of the tapping has an effect only in the motor and not in the
366 audiomotor condition. In the latter, the sound is consistently presented in a perfectly rhythmic
367 manner. Thus, it seems plausible that the introduction of accompanying movements does not
368 exert any discernible impact. The auditory prime remains consistently highly precise in terms
369 of temporal predictions. In contrast, in the motor condition, the temporal predictions are
370 bound to the tapping regularity which fluctuates across trials, influencing task outcomes
371 (Fig.3). This suggests a potential link between enhanced comprehension and the activation of
372 the auditory dorsal pathway through periodic (i.e. temporally regular) primes (Rimmele et al.,
373 2018), independently of the end-effector (auditory or motor). This is also in line with the main
374 result, namely, auditory stimuli consistently presented with perfect rhythmicity induce a
375 greater speech comprehension advantage compared to the motor condition, which prompts
376 less accurate rhythmicity.

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

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

389 (Assaneo et al., 2019). Our results show a positive correlation between the performance of
390 these two tasks (independently of the type of prime; Fig. 4). Hence, degraded-speech
391 comprehension abilities can be partly predicted based on perception-production coupling
392 abilities (and vice-versa). This result is compatible with the fact that the processing of
393 degraded speech implicates the motor cortex (Du et al., 2014; Hickok et al., 2011). This also
394 extends to speech comprehension previous findings showing a positive association between
395 speech perception-production coupling skills and speech rate discrimination (Kern et al.,
396 2021), statistical learning (Assaneo et al., 2019; Orpella et al., 2022) and syllable
397 discrimination (Assaneo et al., 2021). Moreover, this result complements a finding showing a
398 positive correlation between speech perception-production coupling strength and accelerated
399 speech comprehension accuracy (Lubinus et al., 2023). Interestingly, neural tracking of the
400 speech envelope is related to rhythmic priming of speech (Falk, Lanzilotti, et al., 2017). That
401 general perception-production synchronization abilities can predict accuracy in the speech
402 comprehension task (Fig. 4) suggests that further synchronization measures, as for example
403 the amount of neural tracking or of motor entrainment to the auditory prime, could help to
404 establish a clearer link between the prime and the parsing of the heard utterances in future
405 studies.

406 The relation between speech perception-production coupling and speech comprehension was
407 equivalent across the three prime conditions (auditory, motor and audiomotor). One may
408 expect that, because the audiomotor prime requires perception-production coupling, it could
409 benefit more to participants with low speech perception-production coupling. Alternatively,
410 participants with high speech perception-production coupling may be the ones who benefit
411 from top-down effects of the motor system to enhance perception (Assaneo et al., 2021). The
412 absence of such an effect may be due to the fact that a short prime or a mini-block design are
413 not sufficient to change perception-production coupling. Consistent with the findings
414 reporting anatomical differences between good and poor speech synchronizers (Assaneo et al.,
415 2019), our finding that poor synchronizers have poorer degraded speech comprehension point
416 to the fact that degraded speech more heavily relies on perception-production coupling,
417 similarly to what has been hypothesized for some language developmental disorders (Fiveash
418 et al., 2021; Goswami, 2011). In this perspective, our design may not be enough to change
419 neural dynamics that are subtended by anatomical differences, and longer stimulation
420 protocols may be necessary to improve speech-in-noise processing in poor synchronizers,
421 such as those used in clinical settings (Flaugnacco et al., 2015; Hidalgo et al., 2019).

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

427

428

429 DECLARATIONS OF CONFLICT OF INTEREST

430 None.

431

432

433 **References**

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

436 Arnal, L. H., Poeppel, D., & Giraud, A.-L. (2015). Temporal coding in the auditory cortex.
437 *Handbook of Clinical Neurology*, 129, 85-98. [https://doi.org/10.1016/B978-0-444-62630-](https://doi.org/10.1016/B978-0-444-62630-1.00005-6)
438 [1.00005-6](https://doi.org/10.1016/B978-0-444-62630-1.00005-6)

439 Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically
440 can shape hearing. *Nature Human Behaviour*, 5(1), 71-82. [https://doi.org/10.1038/s41562-](https://doi.org/10.1038/s41562-020-00962-0)
441 [020-00962-0](https://doi.org/10.1038/s41562-020-00962-0)

442 Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D.
443 (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating
444 language learning. *Nature Neuroscience*, 22(4), 627-632. [https://doi.org/10.1038/s41593-019-](https://doi.org/10.1038/s41593-019-0353-z)
445 [0353-z](https://doi.org/10.1038/s41593-019-0353-z)

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

- 448 Avilala, V. K., Prabhu P, P., & Barman, A. (2010). The Effect of Filtered Speech on Speech
449 Identification Scores of Young Normal Hearing Adults. *Journal of All India Institute of*
450 *Speech and Hearing*, Vol 29, 115-119.
- 451 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
452 Using lme4. *Journal of Statistical Software*, 67, 1-48. <https://doi.org/10.18637/jss.v067.i01>
- 453 Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally Regular
454 Musical Primes Facilitate Subsequent Syntax Processing in Children with Specific Language
455 Impairment. *Frontiers in Neuroscience*, 10.
456 <https://www.frontiersin.org/articles/10.3389/fnins.2016.00245>
- 457 Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. *Glott*
458 *international*, 5, 341-345.
- 459 Canette, L.-H., Bedoin, N., Lalitte, P., Bigand, E., & Tillmann, B. (2019). The Regularity of
460 Rhythmic Primes Influences Syntax Processing in Adults. *Auditory Perception & Cognition*,
461 2(3), 163-179. <https://doi.org/10.1080/25742442.2020.1752080>
- 462 Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm : Rhythmic
463 priming and audio-motor training affect speech perception. *Acta Psychologica*, 155, 43-50.
464 <https://doi.org/10.1016/j.actpsy.2014.12.002>
- 465 Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming
466 enhances speech production abilities: Evidence from prelingually deaf children.
467 *Neuropsychology*, 29(1), 102-107. <https://doi.org/10.1037/neu0000115>
- 468 Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of
469 speech. *Neuropsychologia*, 50(11), 2652-2658.
470 <https://doi.org/10.1016/j.neuropsychologia.2012.07.018>
- 471 Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of*
472 *Phonetics*, 26(2), 145-171. <https://doi.org/10.1006/jpho.1998.0070>
- 473 Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal
474 modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81, 181-187.
475 <https://doi.org/10.1016/j.neubiorev.2017.02.011>

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

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

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

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

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

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

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

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

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

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

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

507 Jones, M. R. (1976). Time, our lost dimension : Toward a new theory of perception, attention,
508 and memory. *Psychological Review*, 83(5), 323-355. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.83.5.323)
509 [295X.83.5.323](https://doi.org/10.1037/0033-295X.83.5.323)

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

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

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

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

520 Large, E. W., & Jones, M. R. (1999). The dynamics of attending : How people track time-
521 varying events. *Psychological Review*, 106(1), 119-159. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.106.1.119)
522 [295X.106.1.119](https://doi.org/10.1037/0033-295X.106.1.119)

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

525 Lizcano-Cortés, F., Gómez-Varela, I., Mares, C., Wallisch, P., Orpella, J., Poeppel, D.,
526 Ripollés, P., & Assaneo, M. F. (2022). Speech-to-Speech Synchronization protocol to classify
527 human participants as high or low auditory-motor synchronizers. *STAR Protocols*, 3(2),
528 101248. <https://doi.org/10.1016/j.xpro.2022.101248>

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

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

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

540 Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor
541 control of sensory dynamics in Auditory Active Sensing. *Current opinion in neurobiology*, 31,
542 230-238. <https://doi.org/10.1016/j.conb.2014.12.005>

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

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

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

553 Poeppel, D. (2003). The analysis of speech in different temporal integration windows :
554 Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1),
555 245-255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)

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

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

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

564 Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word
565 identification. *Psychological Review*, 104, 319-343. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.104.2.319)
566 295X.104.2.319

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

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

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

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

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

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

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