Explaining flexible continuous speech comprehension from individual motor rhythms

Christina Lubinus[1], Anne Keitel[2], Jonas Obleser[3,4], David Poeppel[1,5,6,7,] and Johanna M. Rimmele[1,6]

[1]Departments of Neuroscience and Cognitive Neuropsychology
Max Planck Institute for Empirical Aesthetics
60322 Frankfurt am Main, Germany,

[2]Psychology,
University of Dundee,
Dundee DD1 4HN, UK,

[3]Department of Psychology,
University of Lübeck,
Lübeck, Germany,

[4]Center for Brain, Behavior, and Metabolism,
University of Lübeck,
Lübeck, Germany,

[5]Department of Psychology,
New York University,
New York, NY, USA

[6]Max Planck NYU Center for Language, Music, and Emotion,
Frankfurt am Main, Germany, New York, NY, USA

[7]Ernst Strüngmann Institute for Neuroscience (in Cooperation with Max Planck Society),
Frankfurt am Main, Germany

Corresponding author:
Christina Lubinus
Phone: +49 (69) 8300479-348
Email: christina.lubinus@ae.mpg.de

Abstract

When speech is too fast, the tracking of the acoustic signal along the auditory pathway deteriorates, leading to suboptimal speech segmentation and decoding of speech information. Thus, speech comprehension is limited by the temporal constraints of the auditory system. Previous research suggests that individual differences in auditory-motor coupling strength might shape such temporal constraints. In two behavioral experiments, we characterize individual differences in the comprehension of naturalistic speech as function of the individual coupling strength between the auditory and motor systems, and the preferred frequencies of the systems. As expected, speech comprehension declined at higher rates. Importantly, both, higher auditory-motor coupling and higher spontaneous speech motor production rates were related to better speech-comprehension performance. Furthermore, performance increased with higher linguistic sentence predictability (using a recurrent neural network) -particularly at higher speech rates and for individuals with high auditory-motor coupling strength- and with participants' higher working memory capacity (Digit Span). These findings support the notion of an individual preferred auditory regime that allows for optimal auditory (speech) processing. Crucially, we provide evidence for a model that assigns a central role to motor-system-dependent individual flexibility in continuous speech comprehension.

1. Introduction

Speech comprehension crucially relies on temporal processing, as human speech and other naturalistic signals have a complex temporal structure with information at different time-scales(1). The temporal constraints of the auditory system limit our ability to understand speech at fast rates(2,3). Interestingly, the motor system seems to underwrite temporal prediction in auditory perception(4), whereas how strongly the auditory and motor system interact varies widely across individuals(5). Accordingly, current oscillatory models of speech comprehension propose that properties of the auditory but also the motor system affect the quality of auditory processing(6,7). In two behavioral experiments, we investigate how the auditory, the motor system and their coupling strength shape the individual flexibility of comprehending fast continuous speech.

Auditory temporal constraints have been observed as preferred rate or preferred regimes of auditory processing (for speech(8,9), but also tones(10,11), and amplitude modulated sound(11–14)) and explained in the context of neurocognitive models of speech perception: Auditory cortex capitalizes on temporal information by dynamically aligning ongoing brain activity to the temporal patterns inherent to the acoustic speech signal(15–18). By hypothesis, by tracking (or entraining to) quasi-rhythmic temporal fluctuations in the speech envelope, endogenous theta brain rhythms in auditory cortex partition the continuous auditory stream into smaller chunks (at roughly the syllabic scale), allowing for decoding of segmental phonology – and ultimately linguistic meaning(15,18–20). Decoding of the speech signal is accomplished seemingly effortlessly within an optimal range centered in the traditional theta band(18), whereas comprehension deteriorates strongly for speech presented beyond ~9 Hz(2,3). While much research has focused on the apparent *stability* of the average acoustic modulation rate at the syllabic scale(8,9), our *flexibility* in speech comprehension (9,21), that is, what constitutes individual differences in understanding fast speech rates, is poorly understood.

The motor system, and auditory-motor coupling in particular, are attractive candidates to facilitate individual differences in auditory speech processing abilities due to the motor systems' modulatory effect on speech perception and its susceptibility to training(22,23). In speech comprehension, the auditory and speech motor brain areas are tightly intertwined(24–28), although the extent to which speech motor processing modulates auditory processing is debated. Specifically, endogenous brain rhythms in both auditory(20,29) and motor(29,30) cortex have been observed to track the acoustic speech signal, and are characterized by preferred frequencies (19,31,32). In contrast to neural measures of preferred frequencies (31–33), here we used a

behavioral estimate (termed “preferred” or “spontaneous rate”). Furthermore, coupling between auditory and motor brain areas during speech processing(5,29,30,34) has been hypothesized to provide temporal predictions about upcoming sensory events to the auditory cortex(4,34–36). Hence, the precision of these predictions may be proportional to the strength of auditory-motor cortex coupling. Interestingly, recent work showed varying auditory-motor cortex coupling strength across the human population(5,6,10,37,38). Assaneo et al.(5) developed a behavioral protocol (spontaneous speech synchronization (SSS) test) which quantifies the strength of auditory-to-motor synchronization during speech production in individuals. Based on the bimodal distribution of auditory-to-motor synchronization in the population, individuals can be classified into high versus low synchronizers. Importantly, beyond superior behavioral synchronization, high synchronizers have stronger structural and functional connectivity between auditory and speech motor cortices. Thus, the SSS-test provides a behavioral measure to approximate individual differences in neuronal auditory-motor coupling strength. We propose that the individual variability in auditory-motor synchronization, previously observed to predict differences in word learning(5), syllable detection(6), and rate discrimination(10), as well as the individual variability in preferred auditory and motor rate predict differences in an individuals’ ability to comprehend continuous speech at fast syllabic rates.

The influence of individual auditory-motor coupling strength on behavioral performance has so far been established for behavioral paradigms using rather basic auditory and speech stimuli (e.g., tones or syllables)(5,6,10). The current study assesses its importance -in a more naturalistic context- during the comprehension of continuous speech. This adds several layers of complexity. First, speech comprehension is improved when linguistic predictability of a sentence is high(39), even in adverse listening situations(40,41). Presumably, the rich linguistic context is used in the brain to derive linguistic predictions about upcoming words and sentences(42–45). Second, as speech unfolds over time, processing of continuous (i.e., longer and more complex) speech naturally demands more working memory capacity for maintenance and access to linguistic and context information(46). To what extent such higher-level processes interact with other predictions (e.g., temporal predictions from the auditory or motor system) is not understood and possible neuronal mechanisms are currently debated(47).

In summary, we investigate the role of auditory-motor coupling strength (SSS-test) and the role of preferred rhythms of the auditory and motor systems for the individual flexibility of the comprehension of continuous speech. First, based on a rich literature(3,18,48–50), we expected a decline in comprehension performance at higher syllabic rates beyond the theta range. Second, as a faciliatory effect of auditory-motor coupling on auditory processing has

been observed(5,6,10), we hypothesized that individual differences in comprehension performance could be predicted by individual auditory-motor coupling strength, with superior speech comprehension for high synchronizers. Such a faciliatory effect might be strongest in demanding listening situations such as at fast syllabic rates(10). Third, while the consequences of potential individual variation in the preferred rates of the motor and auditory systems are not clearly understood, based on previous findings(29), we expected a systematic relation of both preferred (auditory and motor) rates with individual speech comprehension performance. Finally, we hypothesized that both linguistic predictability and working memory span should positively affect speech comprehension.

2. Methods

Two behavioral experiments and a control experiment were conducted: Experiment 1 was performed in the laboratory, while Experiment 2 and the control were online studies. Participants were English native speakers with normal hearing and no neurological or psychological disorders (Exp 1: N = 34, Exp 2: N = 82, Control: N = 39). Participation was voluntary and approved by the local ethics committees. For a detailed description of participants, stimuli, exclusion criteria, and tasks please refer to Supplementary Methods, Fig. 1-3, and Tables 1-2.

2.1.1. Design and materials

Speech comprehension task

In two speech comprehension tasks, we measured participants ability to comprehend sentences at various syllabic rates. Sentences were presented at 7 (Exp 1: [8.2, 9.0, 9.8, 11.0, 12.1, 14.0, 16.4]) or 6 (Exp 2: [5.00, 10.69, 12.48, 13.58, 14.38, 15.00]) rates. In Experiment 1, participants performed a classic intelligibility task. On each trial (N = 70), a sentence was presented through headphones and participants verbally repeated all perceived words. Responses were recorded.

In Experiment 2, speech comprehension was measured by a word-order task. Participants listened to one sentence per trial (N = 240), followed by the presentation of two words from the sentence on screen. Participants indicated via button press which word they heard first.

Speech production task

In the speech production tasks we estimated participants individual spontaneous speech motor (production) rate. In Experiment 1, the speech production task was operationalized by participants reading a text excerpt (216 words). Participants were instructed to read the text excerpt out loud (presented on a printout) at a comfortable and natural pace while their speech was recorded.

In Experiment 2, participants were asked to produce continuous, “natural” speech. To facilitate fluent production, they were prompted by a question/statement belonging to six thematic categories (6 trials; own life, preferences, people, culture/traditions, society/politics, general knowledge, see Supplementary Table 2). Each response period lasted 30 seconds and trials were separated by self-paced breaks. While speaking, participants simultaneously listened to white noise to minimize auditory feedback.

Auditory rate task (only Exp 2)

To measure participants preferred auditory rate, we implemented a two-interval forced choice (2IFC) task, presenting a reference and a comparison stimulus at each trial. Participants indicated via button press which stimulus they preferred. Stimuli were presented at syllabic rates from 3.00 to 8.50 syllables/s (3.00, 3.92, 4.83, 5.75, 6.67, 7.58, 8.50). A reference rate, e.g. 3.00 syllables/s, was compared to all syllabic rates, including itself. For each reference/comparison pair the same sentence was presented – the stimuli only differed in their syllabic rates.

Spontaneous speech synchronization (SSS) test (only Exp 2)

We measured participant’s auditory-motor synchronization using the SSS-test (for details: Assaneo et al.(5)). In the main task, participants listened to a random syllable train and whispered along for a duration of 80s. They were instructed to synchronize their own syllable production to the stimulus presented through their headphones. The syllable rate in the auditory stimulus progressively increased in frequency from 4.3 to 4.7 syllables/s in increments of 0.1 syllables/s, every 60 syllables. Participants completed two trials, while the whispering was recorded.

Digit span test (only Exp 2)

Working memory capacity was quantified using the digit span test (forward and backward (51)). As for the backward test data is missing for N = 21 participants, only the forward span is reported. Digit spans were presented auditorily and participants typed in their responses (52).

2.1.2. Analysis

Spontaneous speech motor production rate (Exp 1 + 2)

The individual *spontaneous speech motor production rate* (i.e., articulation rate(53)) was computed using Praat software(54) by automatically detecting syllable nuclei. The number of syllables (syllable nuclei) was divided by the duration of the utterance (disregarding silent pauses). For Experiment 1, the production rate was computed across the entire reading paragraph. For Experiment 2, it was first calculated for each trial (30 s) separately. Rate estimates were then transformed into z-scores and trials were excluded if z-scores were larger than 2. The production rate was averaged across the remaining trials.

Preferred auditory rate (Exp 2)

First, participants with low performance in the catch trials of the preferred auditory rate task (below 75% correct) were excluded; amongst the remaining participants ($N = 82$) performance was very high ($M = 98.48\%$, $SD = 3.71$). To compute the preferred auditory rate, a distribution of preferred frequencies was derived from all trials by aggregating the frequency of each trials' preferred item. Then a gaussian function was fitted to each participants' distribution and two parameters were extracted: the peak as index for the preferred frequency and the full-width-at-half-maximum (FWHM) as index for the specificity of the response (lower FWHM equals stronger preference for one frequency).

Auditory-motor synchronization (Exp 2)

From the SSS-test (5) we derived the participant's auditory-motor synchronization by calculating the phase-locking value (PLV, see formula 1)(55) between the (cochlea) envelopes of the auditory and the speech signals. To obtain the cochlea envelope of the syllable train (auditory channels: 180–7,246 Hz), we used the Neural System Laboratory (NSL) Auditory Model MATLAB toolbox(56). For the recorded speech signal the amplitude envelope was quantified

as the absolute value of the Hilbert transform. Both envelopes were downsampled to 100 Hz and bandpass filtered (3.5-5.5 Hz) before their phase was extracted by means of the Hilbert transform. The PLV was first estimated for each trial of the SSS-test (time windows 5s, overlap 2s) and then averaged across runs, resulting in a mean PLV. The distribution of mean PLV values was subjected to a k-means algorithm (k = 2) to split participants into a high- and a low-synchronizer group.

$$PLV = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i\left(\theta_1(t)-\theta_2(z)\right)}\right| \quad (1)$$

Linguistic predictability – Recurrent neural network (Exp 2)

Linguistic predictability was measured as perplexity (*M* = 10.79, *SD* = 0.97). Perplexity relates inversely to raw probability values, i.e. lower perplexity equals higher sentence predictability. For each sentence, perplexity was computed as the exponential of the log cross-entropy of word probabilities of all individual words. Prior to this, word probabilities were obtained using a recurrent neural network (RNN, see Supplementary Methods for full details on RNN and perplexity).

Mixed effects models

For both experiments we performed mixed effects analyses to quantify how speech comprehension was affected by all variables of interest, using the R packages *lmer* and *mgcv*. Mixed-effects, rather than fixed-effects models were chosen to account for idiosyncratic variation within variables, i.e. repeated measures and therefrom resulting interdependencies between data points (57,58). Thus, both models included random intercepts for *participant* and *items*.

In Experiment 1, we used a generalized additive mixed-effects model (GAMM). For the dependent variable *speech comprehension*, we calculated the percentage of correctly repeated words for each sentence and subject from the speech comprehension task. The number of correct words was counted manually and transformed into a percentage. Then the dependent variable (single-trial data) was modelled as a function of the fixed effects *syllabic rate* and *spontaneous speech motor production rate*. A random slope for *syllabic rate* could not be included because the model failed to converge, thus the model included only random intercepts. Overall, the model explained ~77% of the variance.

In Experiment 2, the dependent variable *speech comprehension* was binary (*correct* vs *incorrect* word order judgment). Thus, we employed a generalized linear mixed-effects model (GLMM) with a logistic link function. In terms of fixed effects, the model included all variables of interest (*syllabic rate*, *preferred motor rate*, *preferred auditory rate*, *auditory-motor synchronization*, *working memory, sentence predictability*). Additionally, we introduced several linguistic and other covariates (*predictability target 1*, *predictability target 2*, *sentence length* (# of words), *target distance* (i.e., distance in words between the target words), *compression/dilation* of audio file) for nuisance control(59). In addition to random intercepts, the model contained a by-participant random slope for *syllabic rate*, allowing the strength of the effect of the rate manipulation on the dependent variable to vary between participants(57,58). Continuous predictor variables were z-transformed to facilitate the interpretation and comparison of the strength of the different predictors(60). Thus, the coefficients of all continuous predictors reflect log changes in comprehension for each unit (*SD*) increase in a given predictor. We observed no problems with (multi-)collinearity, all variance inflation factors were < 1.2 (package car version 3.0-10(61)). Overall, the model explained ~38% of the variance.

3. Results

3.1. Experiment 1

In Experiment 1, we asked the question: to what extent is speech comprehension affected by one's spontaneous speech motor production rate? *Speech comprehension* was measured as the percentage of correctly repeated words in an intelligibility task (2.75% to 93.70% on average across participants, Fig 1A, right). We observed a mean *spontaneous speech motor production rate* of 4.11 syllables per second (*SD* = 0.35, min = 3.35, max = 4.85) across participants (Fig. 1A, left).

As expected, the GAMM revealed a main effect of *syllabic rate*: slower speech stimuli were associated with better speech comprehension (edf = 4.61, $F = 1260.90$, $p < .001$, Fig. 1B, left, see Supplementary Table 3). Importantly, we observed that the *spontaneous speech motor production rate* influenced speech comprehension: the higher the individual *spontaneous speech motor production rate*, the better the speech comprehension performance (edf = 1.00, $F = 4.37$, $p = .036$, Fig. 1B, right). This supports the hypothesis of the speech motor system playing a role in speech comprehension and invites the interpretation that this role may be linked to the endogenous rate of the speech motor system.

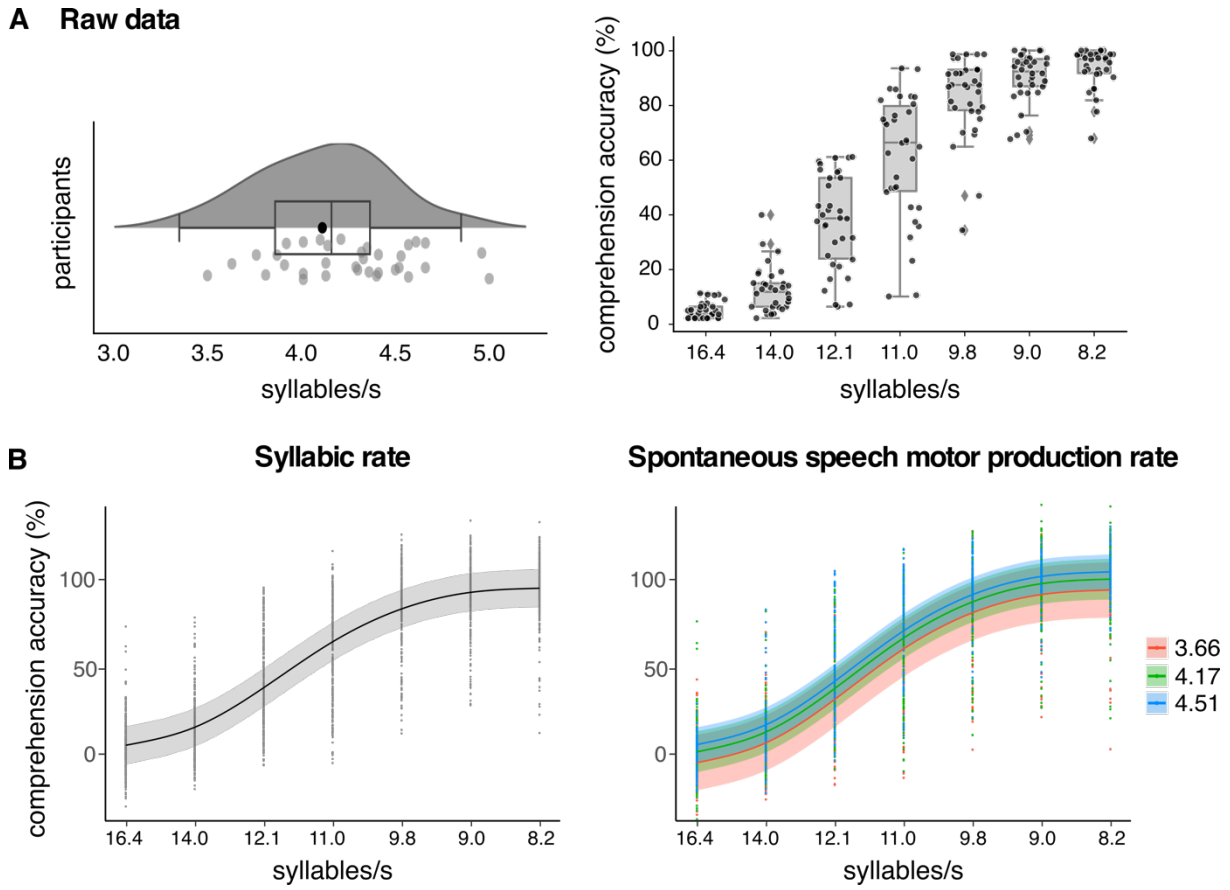

Figure 1. Results of Experiment 1. A. Raw data. We observed spontaneous speech motor production rates between 3.35 and 4.85 syllables/s (M = 4.11 syllables/s, left). Syllabic rate affected speech comprehension such that comprehension dropped dramatically for faster syllabic rates (right). **B. Predictions from a generalized additive mixed model (GAMM).** The predictions of the GAMM visualize the main effects of syllabic rate (left) and spontaneous speech motor production rate (right). The color code in the right panel reflects the spontaneous speech motor production rate (in syllables/s), summarized into 3 bins (3.66, 4.17, 4.51).

3.2. Experiment 2

In Experiment 1, using a well-established behavioral protocol, we observed that one's spontaneous speech motor production rate moderately but robustly affects individual speech comprehension performance under difficult listening conditions. In Experiment 2 we aimed, first, to replicate this observation in a larger sample and, second, to understand the complex interplay of multiple variables during speech comprehension beyond the spontaneous speech motor production rate. The advantage of this approach is that it enabled us to investigate the mechanistic links between multiple variables, while being feasible for the online setting due to its shorter duration.

First, in line with the first experiment, we observed a mean *spontaneous speech motor production rate* of 4.32 syllables per second across participants (*SD* = 0.45, Min = 3.36, Max = 5.38 syllables per second, Fig. 2B). Within-subject variance was low (Supplementary Fig. 4), suggesting that participants' articulation rate was stable across trials. Second, participants showed a *preferred auditory rate* of ~5.57 syllables per second (peak: *M* = 5.57, *SD* = 0.86, Min = 4.16, Max = 7.92; FWHM, *M* = 4.89, *SD* = 0.50, Min = 3.23, Max = 5.50; Fig. 2C). For most participants a preference for one frequency could be observed (Supplementary Fig. 5). Third, *auditory-to-motor speech synchronization* was quantified using the SSS-test(5), classifying participants as HIGH or LOW synchronizers (mean PLV HIGHs = 0.73, *SD* = 0.09, mean PLV LOWs = 0.36, *SD* = 0.09, Fig. 2A). Fourth, *working memory* was measured by means of the digit span test (51) which revealed a mean forward digit score of *M* = 8.46 (*SD* = 2.12, Min = 5.00, Max = 13.00, Fig. 2D).

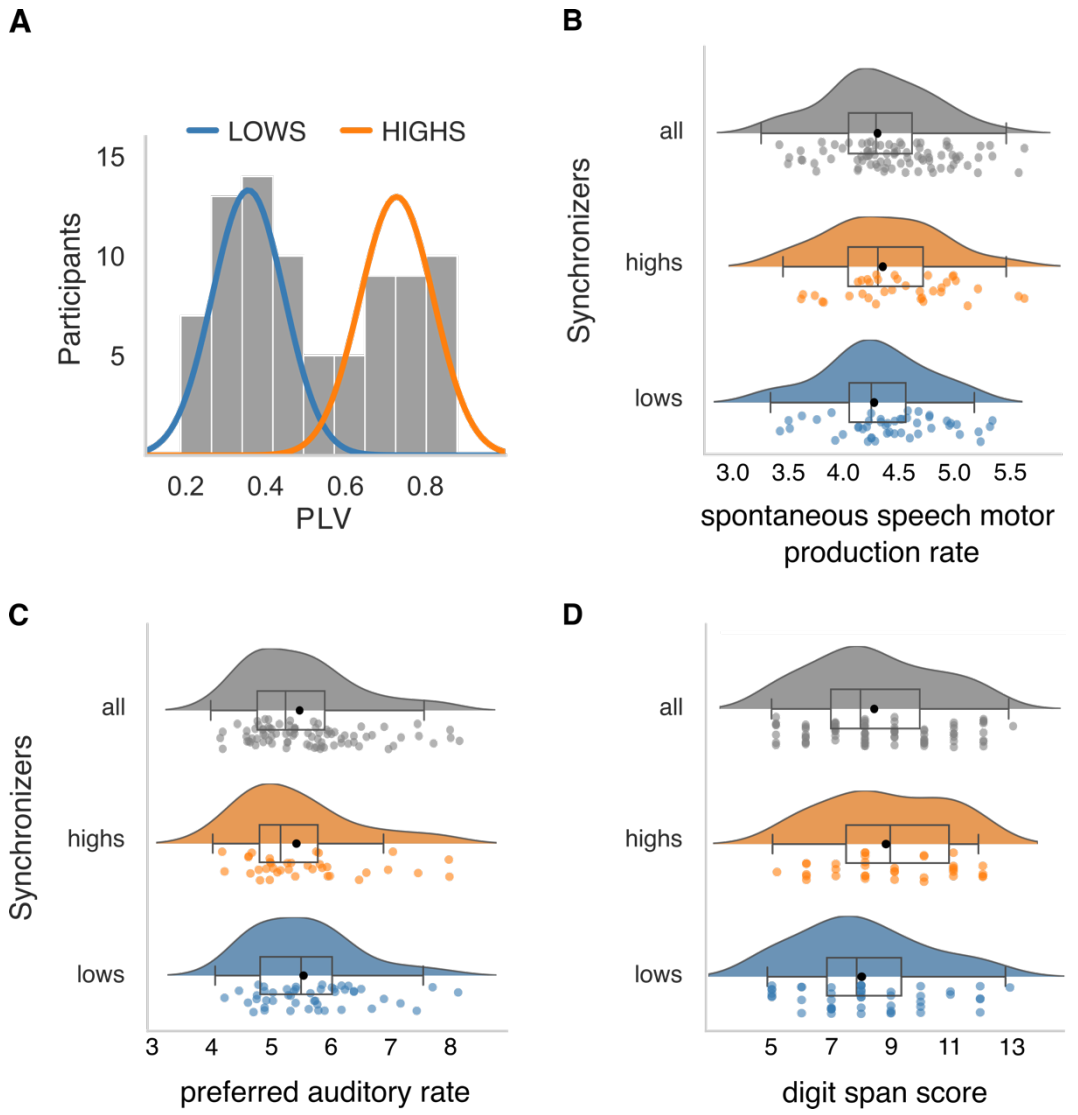

Figure 2. Behavioral results of experimental tasks. A. Histogram visualizing the auditory-motor synchronization strength, obtained with the SSS-test. Participants were classified into high and low synchronizers (highs, lows) based on their PLV using k-means clustering. Group affiliation is overlaid by colored lines representing fitted normal distributions. **B.** We observed a spontaneous speech motor production rate between 3.36 and 5.38 syllables/s (M = 4.32, SD = 0.45) and no group difference between high and low synchronizers (U = 767.0, p = .60). **C.** Participants showed a mean preferred auditory rate of 5.57 syllables/s (SD = 0.86), with no differences between high and low synchronizers (U = 897.5, p = .48). **D.** Working memory capacity was indicated by a mean digit-span forward score of 8.46 (SD = 2.12) and the score did not differ between high and low synchronizers (U = 666.5, p = .14).

To answer our main question, we modelled speech comprehension performance as a function of all variables of interest (*syllabic rate*, *preferred motor rate*, *preferred auditory rate*, *auditory-motor synchronization* (HIGH vs LOW), *working memory score)*, as well as an interaction term for *syllabic rate* and *auditory-motor synchronization* in a GLMM.

Syllabic rate significantly influenced participants' comprehension accuracy. For each increase of syllabic rate by one standard deviation, the odds of a correct word order judgment decreased

0.65 to 1 (log estimate: -0.44, *SE*: 0.06, *CI*: [-0.56 — -0.31], $z = -6.76$, $p < .001$, odds ratio: 0.65, Fig. 3A, Fig. 3B), consistent with a decline of speech comprehension performance at higher syllabic rates. In line with our hypothesis, *spontaneous speech motor production rate* and *auditory-motor synchronization* had a positive effect on speech comprehension. The odds of performing the word order task correctly increased as *spontaneous speech motor production rate* (log estimate: 0.17, *SE*: 0.07, *CI*: [0.04 — 0.31], $z = 2.45$, $p = .014$, odds ratio: 1.19, Fig. 3B) increased by a standard deviation, replicating our finding from the first experiment. For *auditory-motor synchronization*, being a dichotomous variable (i.e., HIGH vs. LOW), the odds of correct performance increased by 1.34 to 1 for a high vs. a low synchronizer (log estimate: 0.29, *SE*: 0.15, *CI*: [0.01 — 0.58], $z = 2.01$, $p = .045$, odds ratio: 1.34, Fig. 3A right, Fig. 3B). Thus, across all trials high synchronizers were more likely to correctly perform the word order judgment. Additionally, the model revealed a positive relation between *working memory score* and speech comprehension (log estimate: 0.19, *SE*: 0.07, *CI*: [0.04 — 0.33], $z = 2.54$, $p = .011$, odds ratio: 1.21, Fig. 3B), suggesting better working memory performance enabled participants to better perform on the speech comprehension task. We did not observe a reliable effect of *preferred auditory rate* on speech comprehension (log estimate: 0.13, *SE*: 0.07, *CI*: [-0.01 — 0.27], $z = 1.80$, $p = .072$, odds ratio: 1.14). In contrast to our hypothesis, we observed no interaction effect of *syllabic rate* and *auditory-motor synchronization* on speech comprehension (log estimate: -0.04, *SE*: 0.07, *CI*: [-0.17 — 0.10], $z = -0.52$, $p = .602$, odds ratio: 0.97).

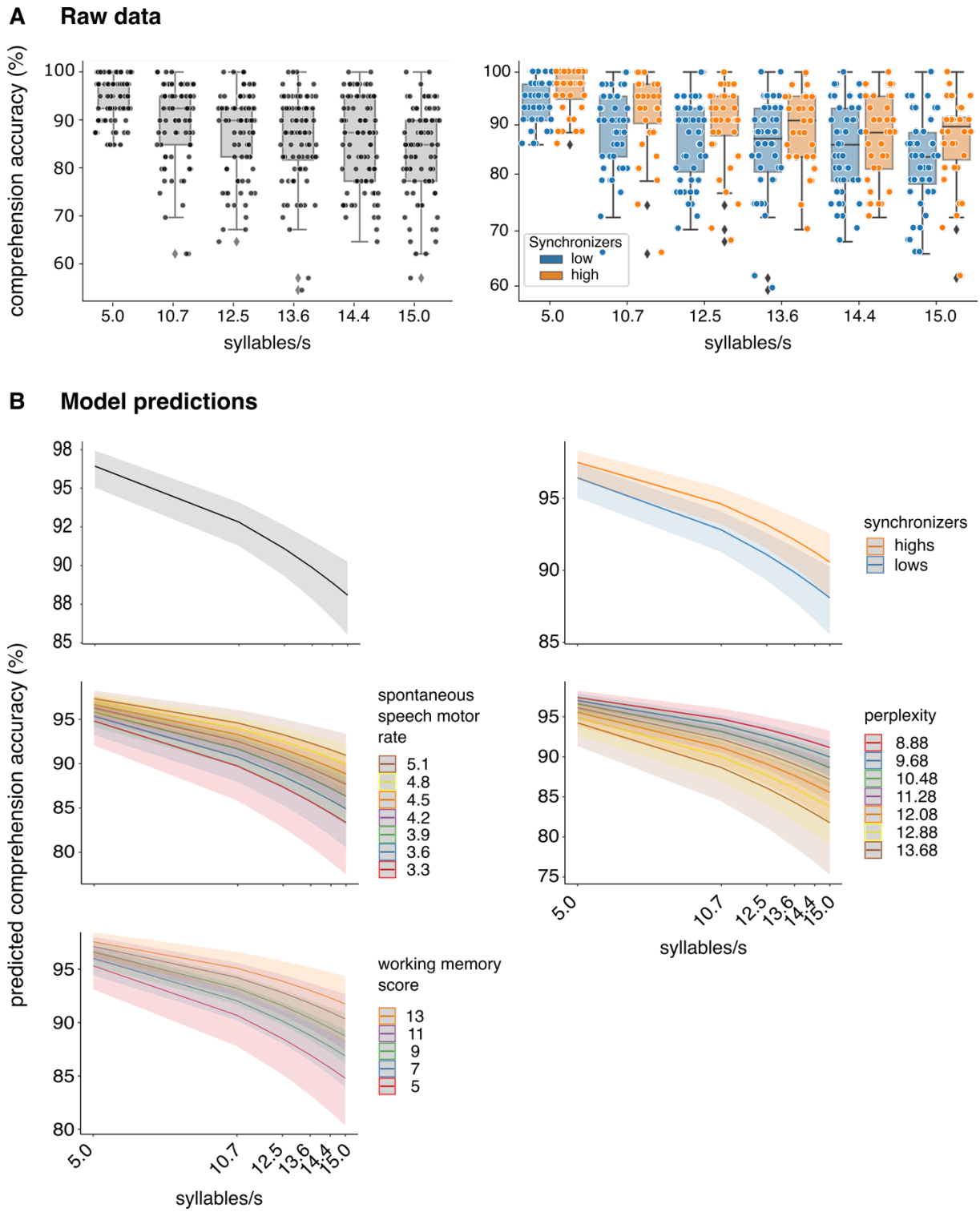

Figure 3. Speech comprehension as a function of predictors of interest. A. Boxplots visualizing raw data from the speech comprehension task as a function of syllabic rate for all participants (left) and for high and low synchronizers separately (right). **B.** Visualization of generalized linear mixed model predictions for all significant main effects. The model revealed a negative main effect of syllabic rate, a positive main effect of auditory-motor synchronization, and speech comprehension increased with a higher spontaneous speech motor production rate, lower perplexity (i.e., higher predictability), and better working memory capacity. Color coding reflects predicted values considered for each main effect.

Linguistic predictability and further linguistic variables

To account for linguistic influences, we expanded the GLMM by adding several linguistic variables: *perplexity*, *probability of target words*, *target distance*, and *stimulus length*. Adding these variables (with linguistic variables, AIC: 12675) improved model fit (without linguistic variables, AIC: 12848), as measured by a likelihood ratio test ($\chi2(6) = 184.24$, $p < .001$, see Supplementary Table 4).

The full GLMM revealed that *perplexity* had a statistically reliable, negative effect on speech comprehension (log estimate: -0.17, SE: 0.05, CI: [-0.27 — -0.07], z = -3.19, p = .001, odds ratio: 0.84, Fig. 3B) such that sentences with lower perplexity (which is equal to higher sentence predictability) lead to better speech comprehension performance. Additionally, we observed significant negative effects for *probability of target word 1* (log estimate: -0.07, SE: 0.03, CI: [-0.13 — -0.01], z = -2.28, p = .023, odds ratio: 0.93) and *target word 2* (log estimate: -0.09, SE: 0.04, CI: [-0.16 — -0.01], z = -2.27, p = .023, odds ratio: 0.92). Contrary to the perplexity effect, this suggests that task performance in the comprehension task was increased for unexpected target words.

Furthermore, the model revealed a positive slope for *target distance* (logit estimate:0.38, *SE*: 0.04, *CI*: [0.31—0.45], $z = 10.54$, $p < .001$, odds ratio: 1.47, Table 1), suggesting that each unit increase in distance was associated with an increase in speech comprehension performance. In contrast, suggesting the opposite relation, for *stimulus length* we observed a negative slope (logit estimate: -0.50, *SE*: 0.05, *CI*: [-0.61— -0.39], $z = -9.27$, $p < .001$, odds ratio: 0.61, Table 1), i.e., shorter sentences resulted in higher comprehension performance. Due to the large number of variables introduced for nuisance control, we applied a control for multiple comparisons (i.e. false discovery rate, see Supplementary Table 5). All effects remained robust; only auditory-motor synchronization changed from a significant effect to a trend ($p = 0.053$) (Note that this was a planned comparison and therefore is discussed).

Beyond the a priori-defined model, we performed an exploratory analysis which assessed the interaction between *syllabic rate*, *auditory-motor synchronization*, and *perplexity*. Particularly under demanding listening conditions (e.g. speech in noise, distorted speech), speech comprehension is facilitated by compensatory influences such as involvement of the motor system (62,63) and linguistic predictions (40,41). In our experiment, speech comprehension was demanding due to the manipulation of the syllabic rate. We expected both auditory-motor synchronization(10) and perplexity to interact with syllabic rate, such that both systems would become stronger predictors for speech comprehension, as syllabic rate increased.

Furthermore, also an interaction effect between auditory-motor synchronization and perplexity seems plausible. To this end, we included a 3-way interaction (*syllabic rate* * *auditory-motor synchronization* * *perplexity*) into the GLMM and additionally explored the three corresponding 2-way interactions.

Adding the interaction term improved model fit ($\chi^2(3) = 13.84$, $p = .004$ (AIC without interaction term: 12675, AIC with interaction term: 12668)). The model revealed two significant 2-way interaction effects: *syllabic rate* * *perplexity* (logit estimate: -0.13, *SE*: 0.05, *CI*: [-0.23—-0.03], $z = -2.52$, $p = .012$, odds ratio: 0.88) and *auditory-motor synchronization* * *perplexity* (logit estimate: -0.15, *SE*: 0.05, *CI*: [-0.25—-0.05], $z = -2.94$, $p = .003$, odds ratio: 0.86; see Supplementary Fig. 6 and Supplementary Table 6). The interaction effect between *syllabic rate* and *perplexity* indicates that particularly comprehension of sentences at fast syllabic rates improves when *perplexity* is low. Furthermore, the *auditory-motor synchronization* * *perplexity* interaction effect suggests that while having better overall speech comprehension, high synchronizers show a stronger effect of *perplexity* compared to low synchronizers, with even better speech comprehension for more predictable sentences. The *syllabic rate* * *auditory-motor synchronization* effect (logit estimate: -0.06, *SE*: 0.07, *CI*: [-0.19—0.08], $z = -0.86$, $p = .391$, odds ratio: 0.94), as tested before, and the three-way interaction effect of *syllabic rate* * *auditory-motor interaction* * *perplexity* (logit estimate: 0.09, *SE*: 0.05, *CI*: [-0.01—0.20], $z = 1.72$, $p = .085$, odds ratio: 1.10) did not show a statistically reliable effect on speech comprehension.

Control experiment

In Experiment 2, speech comprehension performance was exceptionally good, even at high syllabic rates. To ensure the high performance was not an artifact of the task or stimuli, we conducted a control experiment.

Using the materials from Experiment 2, two target words were presented on screen per trial. Participants were asked to judge which one was more likely to occur first in a *hypothetical* sentence. Importantly, participants did not listen to the sentences at any time. For each trial, we computed how many participants correctly guessed the word order (in percent, "*word order index*"). In a new GLMM analysis, this *word order index* was added as covariate into the model from the main analysis (all other parameters remained the same). The analysis revealed that *word order index* did not influence speech comprehension in a statistically meaningful way (log estimate: -0.01, SE: 0.04, CI: [-0.09 – -0.07], z = -0.28, p = .780, odds ratio: 0.99, see Supplementary Table 7).

4. Discussion

In two behavioral experiments, we show major effects of the *syllabic rate* on the comprehension of continuous speech. This finding is in line with proposals of speech comprehension being temporally constrained such that it is optimal for speech at lower syllabic rates. Crucially, in both protocols we observed that speech comprehension (across a wide range of frequencies, 5-15 syllables/s) was affected by participants' *spontaneous speech motor production rate*, with higher rates predicting better speech comprehension. In the second experiment we showed that, beyond the spontaneous rate of the speech motor system, the individual strength of speech auditory-motor synchronization also affected comprehension. In contrast, the preferred speech perception rate was not related to speech comprehension performance. Together, these findings suggest that while speech comprehension seems limited by general processing characteristics of the auditory system, interindividual differences in comprehension flexibility are moderated by the motor system and interactions between the auditory and motor systems (Fig. 5). Our findings furthermore allow us to generalize the effects of individual differences in the motor system on auditory perception, which have been previously shown for simpler stimuli (5,6,10,64), to more natural continuous speech.

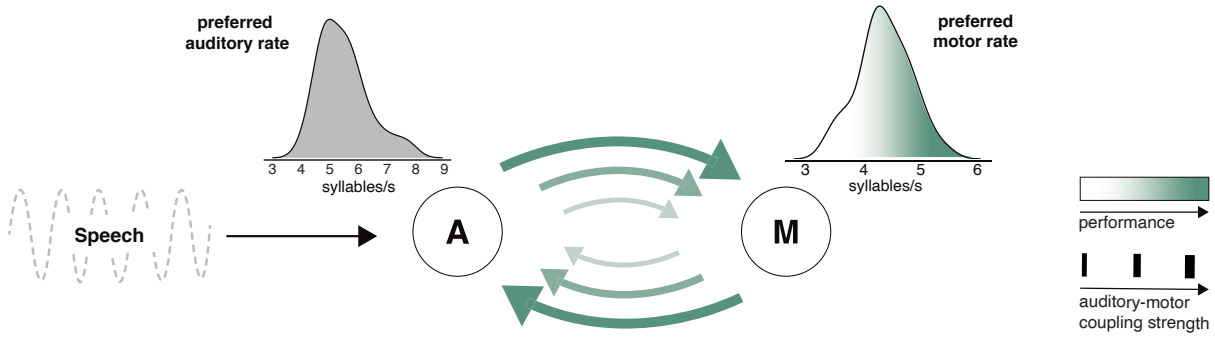

Figure 4. Schematic illustrating the relationship between speech comprehension performance, the preferred rates of the auditory (A) and motor systems (M), and auditory-motor coupling strength. The preferred motor and auditory rates are represented by the corresponding distribution generated from our experimental data. We propose that better speech comprehension at demanding rates, and by hypothesis auditory behavior more generally, is accompanied by a higher preferred rate of the motor system as well as stronger auditory-motor coupling strength. In contrast, the preferred rate of the auditory system seems not to determine auditory behavior.

As expected (2,18,48–50), we observed that speech comprehension accuracy declined as syllabic rate increased. Although speech comprehension dropped at higher rates in both

452 paradigms, the overall level of comprehension accuracy was much higher in Experiment 2,
453 with accuracy remaining very high (~85%), even for speech as fast as 15 syllables/s. In con-
454 trast, in Experiment 1 the increase in syllabic rate resulted in a dramatic drop of comprehension
455 performance. Interestingly, however, in both experiments performance decreased later than
456 previously observed, that is, beyond 9 syllables/s(49,65). Due to the nature of the word-order
457 task we expected better performance than in the classic intelligibility task. Additionally, our
458 control experiment rules out a potential confound by demonstrating that the high performance
459 in Experiment 2 is unlikely due to simple guessing of the correct word order (see Results sec-
460 tion and Supplementary Table 7). Similarly, other studies, also observed shallower decreases
461 in speech comprehension, with relatively high comprehension at higher syllable rates (~12
462 syllables/s) as measured by an intelligibility(49,66) and a 6IFC task(3). Another possibility,
463 explaining the high-speech-rate decline in comprehension performance is that naturally pro-
464 duced fast speech (with matched degrees of compression across syllabic rates, as used in
465 Experiment 2), in contrast to linearly compressed speech, results in more variance of the
466 speech rate and thus allows for part of the sentence to be understood. However, this explana-
467 tion does not account for Experiment 1, in which all stimuli were synthesized at the same rate
468 (varying in degrees of compression). Furthermore, the high performance level might be related
469 to different complexity between more naturalistic sentences providing stronger context infor-
470 mation to compensate loss of information, as compared to the words(18), digits(67) or simple
471 sentences(48) used in previous work. Finally, it is notable that while some studies conceptu-
472 alized the syllabic rate based on the 'theta-syllable' (an information unit defined by cortical
473 function (68)), we, following other studies, define syllabic rate as syllables per second (30).

474 Auditory-motor speech synchronization, a behavioral measure of auditory-motor cortex cou-
475 pling strength(5), had a modulatory -albeit small- effect on speech comprehension. We ob-
476 served that high compared to low synchronizers were characterized by better speech compre-
477 hension performance. These results expand findings which evidenced superior statistical word
478 learning(5) or syllable discrimination(6) for individuals with stronger auditory-motor coupling by
479 showing a similar effect for comprehending more naturalistic, continuous speech. Note that
480 this effect requires further validation as it did not survive the control for multiple comparisons
481 (Supplementary Table 5). Additionally, we expected an interaction of syllabic rate and auditory-
482 motor synchronization, as reported for rate discrimination in tone sequences(10). However,
483 the modulation observed here occurred across all syllabic rates suggesting that an interaction
484 effect may be masked and compensated for by context and linguistic information in continuous
485 speech comprehension. Alternatively, it is possible (although unlikely) that the interaction of

syllabic rate and auditory-motor synchronization was not observed here due to the different frequency resolution at low frequencies. The difference between HIGHs and LOWS in Kern et al.(10) manifested between 7.14 and 10.29 Hz. In contrast, in the present experiment, there was no frequency condition between 5 and 10.69 syllables/s.

Importantly, the spontaneous motor production rate affected speech comprehension, suggesting that individuals with higher spontaneous motor production rate had increased speech comprehension abilities. We replicated this finding in the second experiment. The finding might reflect a complex interplay of auditory and motor cortex during speech comprehension wherein not only the coupling strength, but also the preferred rates of the motor cortex affect speech perception. A possible role of the preferred speech motor rate for speech processing has been previously discussed(29). Furthermore, our findings are in line with an oscillatory model of speech comprehension(6). Interestingly, the preferred auditory rate (around 5.55 syllables/s) had no effect on speech comprehension in our study. A possible explanation is, that auditory cortex preferred rates are less flexible compared to motor cortex preferred rates and thus less prone to individual difference related improvements of speech comprehension. However, comparing the variances of the distribution of preferred auditory ($s^2 = 0.74$) and motor ($s^2 = 0.20$) rates revealed bigger variance in the auditory rate ($F(1,162) = 22.39$, $p < .001$). Another possibility is that due to the behavioral estimation in the current study, preferred auditory cortex rates were not optimally operationalized.

We show that continuous speech comprehension is additionally affected by other higher cognitive and linguistic factors. The relevance of linguistic predictability and working memory capacity have been shown in multiple studies(40,41). In agreement with these studies, higher cognitive variables explained a large amount of variance in speech comprehension. Interestingly, our findings suggests that the faciliatory effect of linguistic predictability first, is particularly effective at fast rates, and second, may be used differently depending on the individual strength of auditory-motor synchronization, with high auditory motor synchronizers showing a larger facilitation effect of linguistic predictability. A relevant question arising from this is: under what conditions is the impact of the motor system on speech comprehension the strongest?

In conclusion, speech comprehension is a highly predictive process which can be informed and affected by different sources of predictions. Here, we show that while speech comprehension is optimal in a preferred auditory temporal regime, the motor-system provides a central role for individual flexibility in continuous speech comprehension. Additionally, interestingly, we report that the well-known faciliatory effects of linguistic predictability on speech comprehension interact with individual differences in the motor system. This sets the stage for future

assessments of how predictions from these systems interact and under what circumstances the human brain relies more on one over the other.

Ethics

Experiment 1 was approved by the ethics committee of the School of Social Sciences, University of Dundee, UK (No. UoD-SoSS-PSY-UG-2019-88). Procedures for Experiment 2 and the Control Experiment were approved by the Max Planck Society (No. 2017_12).

Competing interests

We declare we have no competing interests.

Funding

This work was supported by the Max Planck Society.

Acknowledgments

We thank Anna Broggi and Harry Watt for help with data acquisition (Exp. 1) and Dr. Klaus Frieler for methodological advice.

References

1. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1992 Jun 29;336(1278):367–73.

2. Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, et al. Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. Journal of Neuroscience. 2009 Dec 9;29(49):15564–74.

3. Brungart DS, van Wassenhove V, Brandewie E, Romigh G. The effects of temporal acceleration and deceleration on AV speech perception. AVSP. 2007;27–34.

4. Morillon B, Schroeder CE, Wyart V. Motor contributions to the temporal precision of auditory attention. Nature Communications. 2014 Dec;5(1).

5. Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat Neurosci. 2019 Apr;22(4):627–32.

6. Assaneo MF, Rimmele JM, Sanz Perl Y, Poeppel D. Speaking rhythmically can shape hearing. Nat Hum Behav. 2021 Jan;5(1):71–82.

7. Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. Nat Rev Neurosci. 2020 May 6;

8. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. Neuroscience & Biobehavioral Reviews. 2017 Oct;81:181–7.

9. Pellegrino F, Coupé C, Marsico E. Across-Language Perspective on Speech Information Rate. Language. 2011;87(3):539–58.

10. Kern P, Assaneo MF, Endres D, Poeppel D, Rimmele JM. Preferred auditory temporal processing regimes and auditory-motor synchronization. Psychon Bull Rev. 2021 Jun 7;

11. Teng X, Tian X, Rowland J, Poeppel D. Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. Jensen O, editor. PLOS Biology. 2017 Nov 2;15(11):e2000812.

12. Drake C, Botte M-C. Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. Perception & Psychophysics. 1993 May;54(3):277–86.

13. Teng X, Poeppel D. Theta and Gamma Bands Encode Acoustic Dynamics over Wide-Ranging Timescales. Cerebral Cortex. 2020 Apr 14;30(4):2600–14.

14. Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. The Journal of the Acoustical Society of America. 1979 Nov;66(5):1364–80.

15. Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012 Apr;15(4):511–7.

16. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. Poeppel D, editor. PLoS Biol. 2013 Dec;11(12):e1001752.

17. Peelle JE, Davis MH. Neural Oscillations Carry Speech Rhythm through to Comprehension. Frontiers in Psychology. 2012;3.

18. Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica. 2009;66(1–2):113–26.

19. Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 2007 Dec;56(6):1127–34.

20. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007 Jun;54(6):1001–10.

21. Greenberg S, Carvey H, Hitchcock L, Chang S. Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics. 2003 Jul;31(3–4):465–85.

22. Cason N, Astésano C, Schön D. Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. Acta Psychologica. 2015 Feb;155:43–50.

23. Du Y, Zatorre RJ. Musical training sharpens and bonds ears and tongue to hear speech better. Proc Natl Acad Sci USA. 2017 Dec 19;114(51):13579–84.

24. Cheung C, Hamilton LS, Johnson K, Chang EF. The auditory representation of speech sounds in human motor cortex. eLife. 2016 Mar 4;5:e12577.

25. Evans S, Davis MH. Hierarchical Organization of Auditory and Motor Representations in Speech Perception: Evidence from Searchlight Similarity Analysis. Cereb Cortex. 2015 Dec;25(12):4772–88.

26. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007 May;8(5):393–402.

27. Morillon B, Arnal LH, Schroeder CE, Keitel A. Prominence of delta oscillatory rhythms

in the motor cortex and their relevance for auditory and speech perception. Neuroscience & Biobehavioral Reviews. 2019 Dec;107:136–42.

28. Scott SK, McGettigan C, Eisner F. A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. Nat Rev Neurosci. 2009 Apr;10(4):295–302.

29. Assaneo MF, Poeppel D. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Sci Adv. 2018 Feb;4(2):eaao3842.

30. Keitel A, Gross J, Kayser C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. Bizley J, editor. PLOS Biology. 2018 Mar;16(3):e2004473.

31. Keitel A, Gross J. Individual human brain areas can be identified from their characteristic spectral activation fingerprints. Engel AK, editor. PLOS Biology. 2016 Jun;14(6):e1002498.

32. Lubinus C, Orpella J, Keitel A, Gudi-Mindermann H, Engel AK, Roeder B, et al. Data-Driven Classification of Spectral Profiles Reveals Brain Region-Specific Plasticity in Blindness. Cerebral Cortex. 2021 Mar 31;31(5):2505–22.

33. Rosanova M, Casali A, Bellina V, Resta F, Mariotti M, Massimini M. Natural Frequencies of Human Corticothalamic Circuits. Journal of Neuroscience. 2009 Jun 17;29(24):7679–85.

34. Park H, Ince RAA, Schyns PG, Thut G, Gross J. Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. Current Biology. 2015 Jun;25(12):1649–53.

35. Haegens S, Zion Golumbic E. Rhythmic facilitation of sensory processing: A critical review. Neuroscience & Biobehavioral Reviews. 2018 Mar;86:150–65.

36. Rimmele JM, Morillon B, Poeppel D, Arnal LH. Proactive sensing of periodic and aperiodic auditory patterns. Trends in Cognitive Sciences. 2018 Oct;22(10):870–82.

37. Assaneo MF, Rimmele JM, Orpella J, Ripollés P, de Diego-Balaguer R, Poeppel D. The lateralization of speech-brain coupling Is differentially modulated by intrinsic auditory and top-down mechanisms. Front Integr Neurosci. 2019 Jul 17;13:28.

38. McPherson T, Berger D, Alagapan S, Fröhlich F. Intrinsic Rhythmicity Predicts Synchronization-Continuation Entrainment Performance. Scientific Reports. 2018 Dec;8(1).

39. Grant KW, Seitz PF. The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. The Journal of the Acoustical Society of America. 107.(2):1000–11.

40. Obleser J, Wise RJS, Alex Dresner M, Scott SK. Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. Journal of Neuroscience. 2007 Feb 28;27(9):2283–9.

41. Obleser J, Kotz SA. Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. Cerebral Cortex. 2010 Mar 1;20(3):633–40.

42. Arnal LH, Wyart V, Giraud A-L. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci. 2011 Jun;14(6):797–801.

43. Zhang Y, Frassinelli D, Tuomainen J, Skipper JI, Vigliocco G. More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. Proc R Soc B. 2021 Jul 28;288(1955):20210500.

44. Lewis AG, Bastiaansen M. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex. 2015 Jul;68:155–68.

45. Kuperberg GR, Jaeger TF. What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience. 2016 Jan 2;31(1):32–59.

46. Emmorey K, Giezen MR, Petrich JAF, Spurgeon E, O'Grady Farnady L. The relation between working memory and language comprehension in signers and speakers. Acta Psychologica. 2017 Jun;177:69–77.

47. ten Oever S, Martin AE. An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. eLife. 2021 Aug 2;10:e68066.

48. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences. 2001 Nov;98(23):6.

49. Dupoux E, Green K. Perceptual adjustment to highly compressed speech. Journal of Experimental Psychology: Human Perception and Performance. 1997;23(3):914–27.

50. Pefkou M, Arnal LH, Fontolan L, Giraud A-L. θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech. The Journal of Neuroscience. 2017 Aug 16;37(33):7930–8.

51. Richardson JTE. Measures of Short-Term Memory: A Historical Review. Cortex. 2007

Jan;43(5):635–50.

52. Olsthoorn NM, Andringa S, Hulstijn JH. Visual and auditory digit-span performance in native and non-native speakers. International Journal of Bilingualism. 2014 Dec;18(6):663–73.

53. de Jong NH, Wempe T. Praat script to detect syllable nuclei and measure speech rate automatically. Behavior Research Methods. 2009 May;41(2):385–90.

54. Boersma P, Weenik D. Praat: doing phonetics by computer [Computer program]. 2020 Aug;Version 6.0.40(retrieved from http://www.praat.org/).

55. Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. Human brain mapping. 1999;8(4):194–208.

56. Chi T, Shamma S. NSL MATLAB Toolbox https://isr.umd.edu/Labs/NSL/Software.htm. 2003;

57. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008 Nov;59(4):390–412.

58. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013 Apr;68(3):255–78.

59. Sassenhagen J, Alday PM. A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. Brain and Language. 2016 Nov;162:42–5.

60. Schielzeth H. Simple means to improve the interpretability of regression coefficients: *Interpretation of regression coefficients*. Methods in Ecology and Evolution. 2010 Jun;1(2):103–13.

61. Fox J, Weisberg S. An R companion to applied regression. 3rd ed. Thousand Oaks CA: Sage Publications; 2019.

62. Du Y, Buchsbaum BR, Grady CL, Alain C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proceedings of the National Academy of Sciences. 2014 May 13;111(19):7126–31.

63. Osnes B, Hugdahl K, Specht K. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. NeuroImage. 2011 Feb;54(3):2437–45.

64. Assaneo MF, Orpella J, Ripollés P, Noejovich L, López-Barroso D, de Diego-Balaguer R, et al. Population-level differences in the neural substrates supporting Statistical Learning. bioRxiv; 2020 Jul.

65. Ghitza O. Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. Front Psychol. 2014 Jul 4;5.

66. Giroud J, Pesnot Lerousseau J, Pellegrino F, Morillon B. The channel capacity of multilevel linguistic features constrains speech comprehension. bioRxiv. :30.

67. Doelling KB, Arnal LH, Ghitza O, Poeppel D. Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. NeuroImage. 2014 Jan;85:761–8.

68. Ghitza O. The theta-syllable: a unit of speech information defined by cortical function. Front Psychol. 2013;4.