Research

Author for correspondence:
Christina Lubinus
e-mail: christina.lubinus@ae.mpg.de

THE ROYAL SOCIETY
PUBLISHING

Explaining flexible continuous speech comprehension from individual motor rhythms

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

[1]Department of Neuroscience and Department of 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, and [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, New York, NY, USA
[7]Ernst Strüngmann Institute for Neuroscience (in Cooperation with Max Planck Society), Frankfurt am Main, Germany

CL, 0000-0002-7261-2490

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. Here we ask whether individual differences in auditory-motor coupling strength in part shape these temporal constraints. In two behavioural experiments, we characterize individual differences in the comprehension of naturalistic speech as function of the individual synchronization between the auditory and motor systems and the preferred frequencies of the systems. Obviously, speech comprehension declined at higher speech rates. Importantly, however, both higher auditory-motor synchronization and higher spontaneous speech motor production rates were predictive of better speech-comprehension performance. Furthermore, performance increased with higher working memory capacity (digit span) and higher linguistic, model-based sentence predictability—particularly so at higher speech rates and for individuals with high auditory-motor synchronization. The data provide evidence for a model of speech comprehension in which individual flexibility of not only the motor system but also auditory-motor synchronization may play a modulatory role.

1. Introduction

Speech comprehension relies on temporal processing, as speech and other naturalistic signals have a complex temporal structure with information at different timescales [1]. The temporal constraints of the auditory system limit our ability to understand speech at fast rates [2,3]. Interestingly, the motor system can under certain conditions provide temporal predictions that aid auditory perception [4,5]. 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 behavioural experiments, we investigate how the auditory, the motor system, and their synchronization shape individual flexibility of comprehending fast continuous speech.

Auditory temporal constraints have been observed as preferred rates of auditory speech [8,9] processing (but also of tones [10,11], and amplitude modulated sounds [11–14]) and explained in the context of neurocognitive models of speech perception. According to such proposals, humans capitalize on temporal

information by dynamically aligning ongoing brain activity in auditory cortex to the temporal patterns inherent to the acoustic speech signal [15–18]. By hypothesis, endogenous theta brain rhythms in auditory cortex partition the continuous auditory stream into smaller chunks at roughly the syllabic scale by tracking quasi-rhythmic temporal fluctuations in the speech envelope. This chunking mechanism allows for the decoding of segmental phonology – and ultimately linguistic meaning [15,18–20]. The decoding of the speech signal is accomplished seemingly effortlessly within an optimal range centred in the traditional theta band [18], whereas comprehension deteriorates strongly for speech presented beyond approximately 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], the *flexibility* in speech comprehension [9,21], that is, what constitutes individual differences in understanding fast speech rates, is poorly understood.

The motor system, and neural auditory-motor coupling in particular, is a plausible candidate to facilitate individual differences in auditory speech processing abilities. Two arguments supporting this notion are the motor systems' modulatory effect on auditory perception [22–24] and its susceptibility to training [25–27]. While there is evidence suggesting that the auditory and speech-motor brain areas are intertwined during speech comprehension [28–32], the extent to which speech-motor processing modulates auditory processing is debated [5,33,34]. Specifically, endogenous brain rhythms in both auditory [20,35] and motor [35,36] cortex have been observed to track the acoustic speech signal, and are characterized by preferred frequencies [19,37,38]. By contrast to neural measures of preferred frequencies [37–39], here we used a behavioural estimate termed 'preferred' or 'spontaneous' rate. Furthermore, neural coupling between auditory and motor brain areas during speech processing [35,36,40,41] has been hypothesized to provide temporal predictions about upcoming sensory events to the auditory cortex [4,41–43]. The precision of these predictions may be proportional to the strength of auditory-motor cortex coupling.

Auditory-motor cortex coupling strength varies across the population, as shown by recent work [6,10,40,44,45]. Assaneo *et al.* [40] developed a behavioural protocol (spontaneous speech synchronization test; SSS-test) which quantifies the strength of auditory-to-motor synchronization during speech production in individuals. The authors reported that auditory-motor synchronization is characterized by a bimodal distribution in the population, classifying individuals into high versus low synchronizers. (The rejection of unimodality has been previously shown with large sample sizes [40] (see also [46]). Importantly, in addition to superior behavioural synchronization, high synchronizers have stronger structural and functional connectivity between auditory and speech motor cortices (see [40]; figure 3a,b). Thus, the SSS-test provides not only a behavioural measure but also approximates 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 [40], syllable detection [6], and rate discrimination [10], as well as the individual variability in preferred auditory and motor rate, predicts differences in an individuals' ability to comprehend continuous speech at fast syllabic rates.

The influence of individual auditory-motor coupling strength on behavioural performance has so far been established

for behavioural paradigms using rather basic auditory and speech stimuli (e.g. tones or syllables) [6,10,40]. The current study assesses its importance in a more naturalistic context: during the comprehension of continuous speech. This adds several layers of complexity. First, 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 [47]. Second, rich linguistic context is used to derive linguistic predictions about upcoming words and sentences [48–51]. When linguistic predictability of a sentence is high [52], speech comprehension is improved, even in adverse listening situations [53,54]. Thus, similar to auditory-motor synchronization, linguistic predictability offers a compensatory mechanism when comprehension is difficult.

In summary, we investigate the role of auditory-motor synchronization with the 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 established literature [3,18,55–57], we expected a decline in comprehension performance at syllabic rates beyond the theta range. Second, as a faciliatory effect of auditory-motor coupling on auditory processing has been observed [6,10,40], we hypothesized that individual differences in comprehension performance could be predicted by individual auditory-motor synchronization, with superior speech comprehension for high synchronizers. Such a faciliatory effect might be strongest in demanding listening situations, such as at fast syllabic rates [5,10]. 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 [35] we expected a systematic relation of both preferred auditory and motor rates with individual speech comprehension performance. Finally, we hypothesized that linguistic predictability and working memory span should positively affect speech comprehension. Similar to auditory-motor synchronization, we expected linguistic predictability to interact with syllabic rate, such that both systems would become stronger predictors for speech comprehension as syllabic rate increases.

2. Methods

Two behavioural experiments and a control experiment were conducted: experiment 1 was performed in the laboratory and investigated the influence of the spontaneous speech motor production rate on speech comprehension performance. In experiment 2 we aimed to understand the complex interplay of multiple variables during speech comprehension beyond the spontaneous speech motor production rate. To this end, we additionally measured participants' preferred auditory rate, auditory-motor synchronization, and working memory capacity. experiment 2 and the control experiment were online studies.

(a) Participants

Participants were English native speakers with normal hearing and no neurological or psychological disorders (experiment 1: $n = 34$, experiment 2: $n = 82$, control: $n = 39$). Participation was voluntary. For a detailed description of participants, stimuli, exclusion criteria and tasks please refer to electronic supplementary material, methods, figures S1 and S2, and tables S1 and S2.
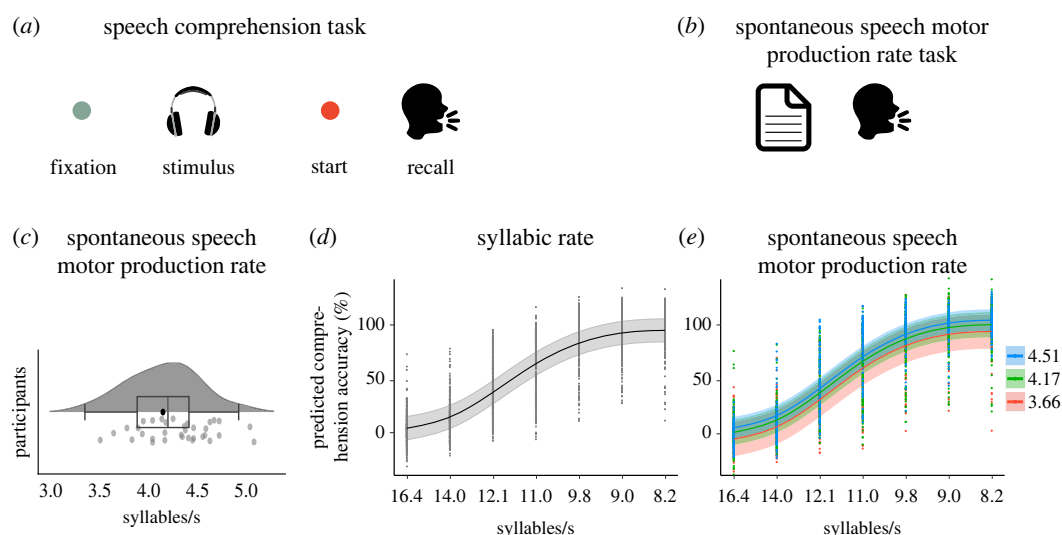
Figure 1. (*a*) Example trial for the speech comprehension task. Participants fixated on a green fixation dot while presented auditorily with a sentence. On stimulus offset the fixation dot turned red, indicating to commence recall, i.e. reporting the sentence back. (*b*) Spontaneous speech motor production rate task. Participants read a stimulus paragraph from a paper. (*c*) Spontaneous speech motor production rate. We observed spontaneous speech motor production rates between 3.35 and 4.85 syllables per second ($M = 4.11$ syllables per second, left). The violin and boxplot show summary statistics and density: the median centre line, 25th to 75th percentile hinges, whiskers indicate minimum and maximum within $1.5 \times$ interquartile range. Grey dots represent participants individual speech motor productions rates, averaged across 6 trials. (*d*) Main effect of syllabic rate. Plot shows the predicted main effect of syllabic rate from the generalized additive mixed model (GAMM). Black line indicates the predicted effect with 95% confidence interval in grey. Black dots show trial-level speech comprehension performance per subject and rate condition. (*e*) Main effect of spontaneous speech motor production rate. Plot shows the predicted main effect of spontaneous speech motor production rate from the GAMM. Coloured lines indicate the predicted effect with 95% confidence interval in the corresponding colour. Coloured dots show trial-level speech comprehension performance per subject and rate condition.

(b) Design and materials

(i) Speech comprehension task

In two speech comprehension tasks, we measured participants ability to comprehend sentences at various syllabic rates. Sentences were presented at 7 (experiment 1: [8.2, 9.0, 9.8, 11.0, 12.1, 14.0, 16.4]) or 6 (experiment 2: [5.00, 10.69, 12.48, 13.58, 14.38, 15.00]) rates. In experiment 1, participants performed a classic intelligibility task, also termed 'word identification task' [58,59] (review in [60]). On each trial ($n = 70$), a sentence was presented through headphones and participants verbally repeated the sentence as accurately as possible (figure 1*a*). 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 (figure 2*a*).

(ii) 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) from a printout. Participants were instructed to read the text excerpt out loud at a comfortable and natural pace while their speech was recorded (figure 1*b*).

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 electronic supplementary material, table S2). Each response period lasted 30 s and trials were separated by self-paced breaks (figure 2*c*). While speaking, participants simultaneously listened to white noise. The white noise was introduced to measure the preferred rate of the motor system, without potential interference from auditory feedback. A second reason was to be consistent with the protocol from the

SSS-test ([40,61]; also see below). Note that this procedure was not applied in experiment 1.

(iii) Auditory rate task (only experiment 2)

To measure participants preferred auditory rate, we implemented a two-interval forced choice (2IFC) task, presenting a reference and a comparison stimulus in random order in each trial. Participants indicated via button press which stimulus they preferred (figure 2*b*). Stimuli were presented at syllabic rates from 3.00 to 8.50 syllables per second (3.00, 3.92, 4.83, 5.75, 6.67, 7.58, 8.50). A reference rate, e.g. 3.00 syllables per second, was compared to all syllabic rates, including itself. For each reference/comparison pair the same sentence was presented – that is, the two stimuli in any given trial only differed in their syllabic rate. Additionally, the task included catch trials to measure participant's engagement (see electronic supplementary material, Methods for details).

(iv) Spontaneous speech synchronization test (only experiment 2)

We measured participant's auditory-motor synchronization using the spontaneous speech synchronization test (SSS-test) (for details see [40]). 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 (figure 2*d*). The syllable rate in the auditory stimulus progressively increased in frequency from 4.3 to 4.7 syllables per second in increments of 0.1 syllables per second, every 60 syllables. Participants completed two trials, while the whispering was recorded.

Participants' syllable production was masked by the simultaneously presented auditory syllable train. The masking procedure suppresses auditory feedback, allowing us better to isolate the synchronization of motor production to the auditory input, without interference of auditory feedback [44].
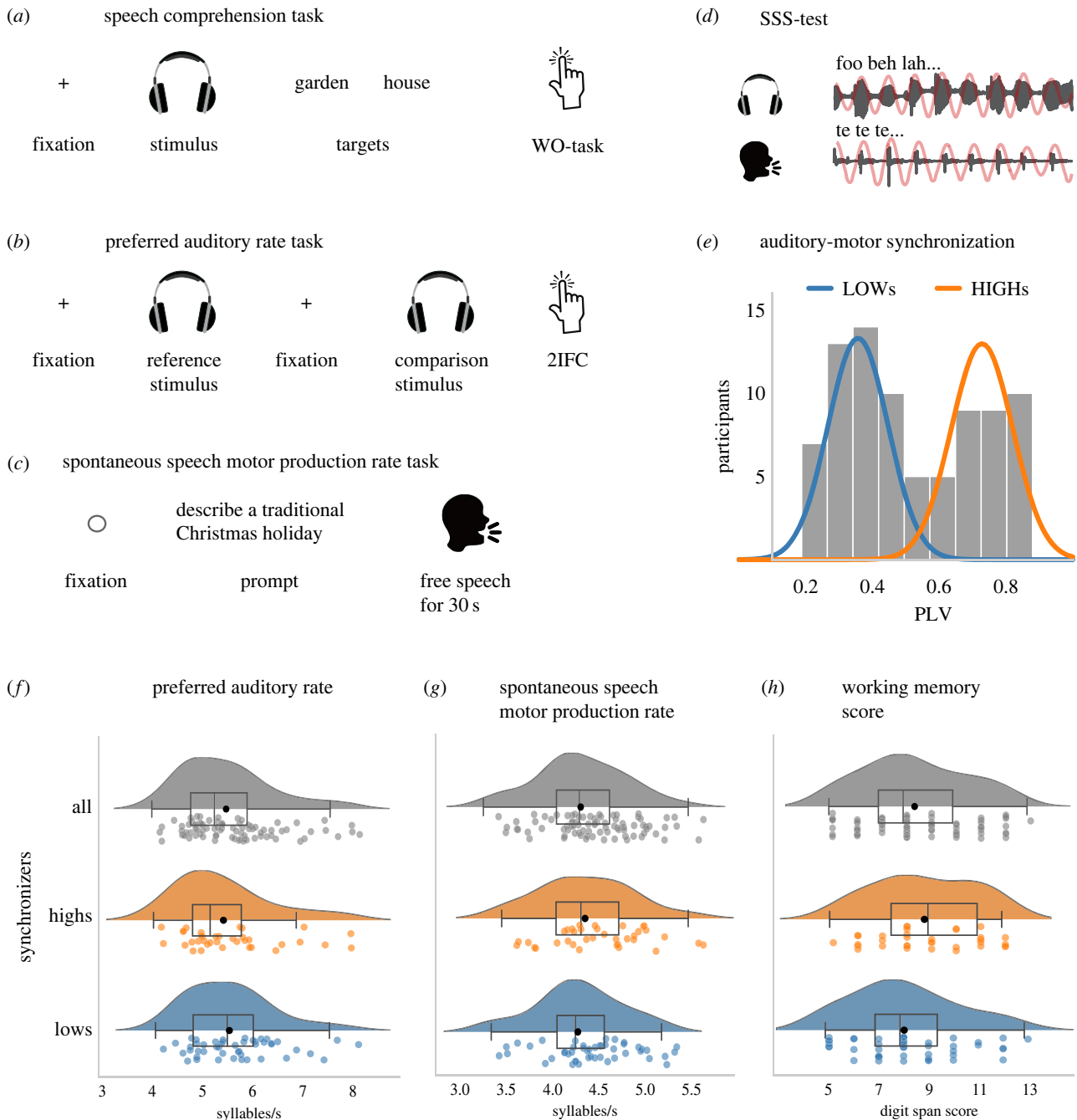
Figure 2. Example trials for (*a*) the speech comprehension task, (*b*) the preferred auditory rate task and (*c*) the spontaneous speech motor production task. (*d*) Schematic representation of the SSS-test used to measure auditory-motor synchronization. Participants whisper a syllable (here /te/). (*e*) Histogram of 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 coloured lines representing fitted normal distributions. (*f*) Participants showed a mean preferred auditory rate of 5.57 syllables per second (s.d. = 0.86), with no differences between high and low synchronizers ($U = 897.5$, $p = 0.48$). (*g*) We observed a spontaneous speech motor production rate between 3.36 and 5.38 syllables per second ($M = 4.32$, s.d. = 0.45) and no group difference between high and low synchronizers ($U = 751.0$, $p = 0.51$). (*h*) Working memory capacity was indicated by a mean digit-span forward score of 8.46 (s.d. = 2.12) and the score did not differ between high and low synchronizers ($U = 666.5$, $p = 0.14$).

(v) Digit span test (only experiment 2)

Working memory capacity was quantified using the forward and backward [62] digit span test. 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 [63].

(vi) Control experiment

We designed a control experiment to test if the correct word order from the word order task of experiment 2 could be guessed from the target words alone, that is, without understanding the sentence. The task consisted in judging which of two words would be more likely to occur first in a *hypothetical sentence*. On

each trial, two words were presented on screen and participants indicated their choice via button press. Importantly, (1) participants did not listen to a full sentence at any time and (2) the target words were taken from the stimulus materials actually presented in experiment 2.

(c) Analysis

(i) Spontaneous speech motor production rate (experiment 1 + 2)

The individual *spontaneous speech motor production rate* (i.e. articulation rate [64]) was computed using Praat software [65] by
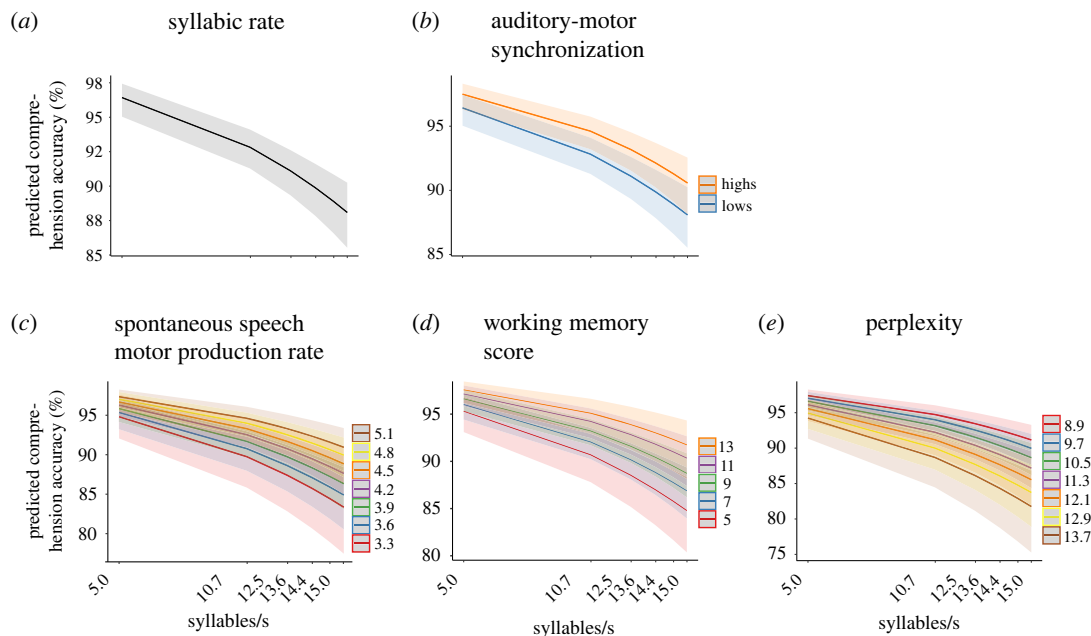
Figure 3. Significant main effects predicting speech comprehension performance. The generalized linear mixed effects model revealed a negative main effect of (*a*) syllabic rate and positive main effects of (*b*) auditory-motor synchronization, (*c*) spontaneous speech motor production rate and (*d*) and working memory score. (*e*) For stimulus perplexity we observed a negative main effect. In all panels, error shades indicate 95% confidence intervals. Note that the predictors are shown as a function of syllabic rate for visualization purposes only.

automatically detecting syllable nuclei. The number of 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. The motor rate was then averaged across all trials.

(ii) Preferred auditory rate (experiment 2)

First, participants with low performance in the catch trials of the preferred auditory rate task (below 75% correct) were excluded; among the remaining participants ($n = 82$) catch trial performance was very high ($M = 98.48\%$, $s.d. = 3.71$). To compute the preferred auditory rate, a distribution of preferred frequencies was derived from all trials (except catch 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).

(iii) Auditory-motor synchronization (experiment 2)

From the SSS-test [40] we derived the participant's auditory-motor synchronization by calculating the phase-locking value (PLV) [66] between the (cochlea) envelopes of the auditory and the speech signals.

$$\text{PLV} = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i(\theta_1(t) - \theta_2(t))}\right|, \tag{2.1}$$

where T is the total number of time points, t denotes the discretized time, and θ_1 and θ_2 are the phase of the first and the second signals, respectively.

To obtain the cochlear envelope of the syllable train (auditory channels: 180–7246 Hz), we used the Chimera Software toolbox [67]. 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 5 s, overlap 2 s) and then averaged across runs, resulting in a mean PLV. The distribution of mean PLV values was subjected to a k-means algorithm [68] ($k = 2$) to split participants into a high- and a low-synchronizer group. Speech auditory-motor synchronization (PLV) was treated as bimodal variable based on previous research that rejected unimodality based on larger samples [40] (see also [46]).

(iv) Linguistic predictability—recurrent neural network (experiment 2)

Linguistic predictability of all stimulus sentences was measured by deriving single-sentence perplexity from a recurrent neural network language model. A language model, such as a recurrent neural network, assigns probabilities to all words in a sequence of words. From the single-word probabilities, we derived one value per sentence, quantifying its predictability [69,70]. This so-called perplexity is the most common intrinsic evaluation metric of language models [71–73]. It is computed as the inverse of the mean probability of a sentence weighted by sentence length [69] (i.e. lower perplexity values equal higher sentence predictability; see electronic supplementary material, Methods for full details on RNN and perplexity).

(v) Mixed-effects models

For both experiments, we performed mixed effects analyses to quantify how speech comprehension was affected by all variables of interest. Mixed models were computed using the R packages *lme4* (v. 1.1–29) and *mgcv* (v. 1.8–39), as set up in Rstudio (v. 2022.2.1.461). 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) [74,75]. Thus, both models included random intercepts for *participant* and *items*.

In experiment 1, we computed a generalized additive mixed-effects model (GAMM) using the *mgcv:gam* function. 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 approximately 77% of the variance.

In experiment 2, the dependent variable *speech comprehension* was binary (*correct* versus *incorrect* word order judgement). Thus, we employed a generalized linear mixed-effects model (GLMM; *lme4:glmer* function) with a binomial logit link function. In terms of fixed effects, the model included all variables of interest: *syllabic rate, spontaneous speech motor production rate, preferred auditory rate, auditory-motor synchronization, working memory, sentence predictability*. Additionally, we introduced several linguistic and other covariates for nuisance control [76]: *predictability target 1, predictability target 2, sentence length* (number of words), *target distance* (i.e. distance in words between the target words), *compression/dilation* of audio file. 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 [74,75]. Continuous predictor variables were z-transformed to facilitate the interpretation and comparison of the strength of the different predictors [77]. Thus, the coefficients of all continuous predictors reflect log changes in comprehension for each unit (*s.d.*) increase in a given predictor. We observed no problems with (multi-)collinearity, all variance inflation factors were less than 1.2 (package car v. 3.0–10 [78]). Overall, the model explained approximately 38% of the variance.

(vi) Control experiment

For each trial, we computed how many participants correctly guessed the word order (as a percentage, '*word order index*'). In a new GLMM analysis, this *word order index* was added as covariate into the model from the main analysis while all other parameters remained the same.

3. Results

(a) 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). We observed a mean *spontaneous speech motor production rate* of 4.11 syllables per second (s.d. = 0.35, min = 3.35, max = 4.85) across participants (figure 1c).

As expected, the GAMM revealed a main effect of *syllabic rate*: slower speech stimuli were associated with better speech comprehension (edf. = 4.91, $F = 1222.01$, $p < 0.001$; figure 1d; see electronic supplementary material, table S3). 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, $F = 4.25$, $p = 0.039$; figure 1e).

(b) experiment 2

First, in line with the first experiment, we observed a mean *spontaneous speech motor production rate* of 4.30 syllables per second across participants (s.d. = 0.45, min = 3.35, max = 5.33 syllables per second; figure 2g). Within-subject variance

was low (electronic supplementary material, figure S3), suggesting that participants' articulation rate was stable across trials. Second, participants showed a *preferred auditory rate* of 5.57 syllables per second (peak: $M = 5.57$, s.d. = 0.86, min = 4.16, max = 7.92; FWHM, $M = 4.89$, s.d. = 0.50, min = 3.23, max = 5.50; figure 2f). Single-subject raw data can be inspected in electronic supplementary material, figure S4. Third, *auditory-to-motor speech synchronization* was quantified using the SSS-test [40], classifying participants as HIGH or LOW synchronizers (mean PLV HIGHs = 0.73, s.d. = 0.09, mean PLV LOWs = 0.36, s.d. = 0.09; figure 2e). Fourth, *working memory* was measured by means of the digit span test [62] which revealed a mean forward digit score of $M = 8.46$ (s.d. = 2.12, min = 5.00, max = 13.00; figure 2h).

The GLMM revealed that *syllabic rate* significantly influenced participants' comprehension accuracy: for each increase of syllabic rate by one syllable/s, the odds of a correct word order judgement decreased (odds ratio (OR = 0.65, std. error (s.e.) = 0.04, $p < 0.001$; figure 3a). This main effect of syllabic rate is consistent with a decline of speech comprehension performance at higher syllabic rates [3]. In line with our hypothesis, we observed main effects for *spontaneous speech motor production rate* and *auditory-motor synchronization*. The higher a participant's *spontaneous speech motor production rate*, the better the performance in the word order task (OR = 1.19, s.e. = 0.09, $p = 0.014$, figure 3c), replicating our finding from the first experiment. For *auditory-motor synchronization*, being a dichotomous variable (i.e. HIGH versus LOW) [40], performance in the word order judgement task was higher for high compared to low synchronizers (OR = 1.34, s.e. = 0.20, $p = 0.048$; figure 3b). That is, across all trials, high synchronizers were more likely to correctly perform the task. Additionally, the model revealed a positive effect for *working memory score* (OR = 1.20, s.e. = 0.09, $p = 0.012$; figure 3d). This main effect suggests that 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 (OR = 1.14, s.e. = 0.08, $p = 0.072$). By contrast to our hypothesis, we observed no interaction effect of *syllabic rate* and *auditory-motor synchronization* on speech comprehension (OR = 0.97, s.e. = 0.07, $p = 0.602$).

(i) Linguistic predictability and further linguistic variables

To account for the effect of linguistic attributes, we expanded the GLMM by adding several (information-theoretic) 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 ($\chi^2 = 184.24$, $p < 0.001$; see electronic supplementary material, table S4).

The full GLMM revealed that *perplexity* had a statistically reliable, negative effect on speech comprehension (OR = 0.84, s.e. = 0.04, $p = 0.001$; figure 3e) 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* (OR = 0.93, s.e. = 0.03, $p = 0.026$) and *target word 2* (OR = 0.92, s.e.: 0.03, $p = 0.021$). 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 effect for *target distance* (OR = 1.48, s.e.: 0.05, $p < 0.001$), suggesting that larger distance between targets was associated with better speech comprehension performance. By contrast, suggesting the opposite relation, for *stimulus length* we observed a negative effect (OR = 0.61, s.e.: 0.03, $p < 0.001$), 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; for full results see electronic supplementary material, table S5). All effects remained robust after FDR correction: syllabic rate: $p < 0.001$; spontaneous speech motor production rate: $p = 0.023$; preferred auditory rate: $p = 0.078$; working memory score: $p = 0.022$; perplexity: $p = 0.003$; probability target 1: $p = 0.034$; probability target 2: $p = 0.030$; compression: $p < 0.001$; sentence length: $p < 0.001$; target distance: $p < 0.001$. Only auditory-motor synchronization changed from a significant effect to a trend ($p = 0.057$) (note that this was a planned comparison and therefore is discussed).

Finally, we explored interaction effects between *syllabic rate*, *auditory-motor synchronization* and *perplexity*. Adding the interaction term improved model fit ($\chi^2 = 13.84$, $p = 0.004$ (AIC without interaction term: 12675; AIC with interaction term: 12668)). The model revealed two significant 2-way interaction effects: *syllabic rate × perplexity* (OR = 0.88, s.e. = 0.05, $p = 0.015$) and *auditory-motor synchronization × perplexity* (OR = 0.86, s.e. = 0.04, $p = 0.003$; see electronic supplementary material, figure S5 and table S6). 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 (OR = 0.94, s.e. = 0.07, $p = 0.392$), as tested before, and the three-way interaction effect of *syllabic rate × auditory-motor interaction × perplexity* (OR = 1.09, s.e. = 0.06, $p = 0.106$) did not show a statistically reliable effect on speech comprehension.

(ii) Control experiment

In experiment 2, speech comprehension performance was exceptionally good, even at high syllabic rates. To ensure the high performance was not an artefact of the task or stimuli, we conducted a control experiment. The analysis revealed that *word order index* did not influence speech comprehension in a statistically meaningful way (OR = 0.96, s.e. = 0.07, $p = 0.219$; see electronic supplementary material, table S7).

4. Discussion

In two behavioural experiments, we show clear effects of 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 per second) was predicted 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 predicted comprehension. By contrast, the preferred speech perception rate was not related to speech comprehension performance. Together, these findings suggest that while speech comprehension is limited by general processing characteristics of the auditory system, interindividual differences in comprehension flexibility are intertwined with characteristics of the motor system and auditory-motor interactions (figure 4). 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 [6,10,40,79], to more natural continuous speech.

As expected [2,18,55–57], we observed that speech comprehension accuracy declined as syllabic rate increased. Although speech comprehension dropped at higher rates in both paradigms, the overall level of comprehension accuracy was much higher in experiment 2, with accuracy remaining very high (approx. 85%), even for speech as fast as 15 syllables per second. By contrast, in experiment 1 the increase in syllabic rate resulted in a dramatic drop of comprehension performance. This is in line with our expectations, as the nature of the word-order task is likely to yield overall better performance than the classic intelligibility task. Additionally, our control experiment rules out a potential confound by demonstrating that the high performance in experiment 2 is not due to simple guessing of the correct word order (see Results section and electronic supplementary material, table S7). Interestingly, however, in both experiments performance decreased later than previously observed, that is, beyond rates of 9 syllables per second [56,80]. However, in line with our findings, several other studies, also observed shallower decreases in speech comprehension, with relatively high comprehension at higher syllable rates (approx. 12 syllables per second) [3,56,81,82]. We consider several possible explanations for these discrepancies. One explanation for the different and higher speech-rate decline in comprehension performance is that naturally produced fast speech (with matched degrees of compression across syllabic rates, as used in experiment 2), in contrast to linearly compressed speech, results in more variance of the speech rate and thus allows for part of the sentences to be understood. However, this explanation does not account for experiment 1, in which all stimuli were synthesized at the same rate (varying in degrees of compression). Furthermore, the high performance level might be related to different complexity between more naturalistic sentences, providing stronger context information to compensate loss of information, as compared to the words [18], digits [83], or simple sentences [55] used in previous work. Finally, it is notable that while some studies conceptualized the syllabic rate based on the 'theta-syllable' (an information unit defined by cortical function [84]), we define syllabic rate as linguistically defined syllables per second, following other studies [36].

Auditory-motor speech synchronization, a behavioural estimate of auditory-motor cortex coupling strength [40], had a modulatory (albeit small) effect on speech comprehension. We observed that high compared to low synchronizers exhibited better speech comprehension performance. These
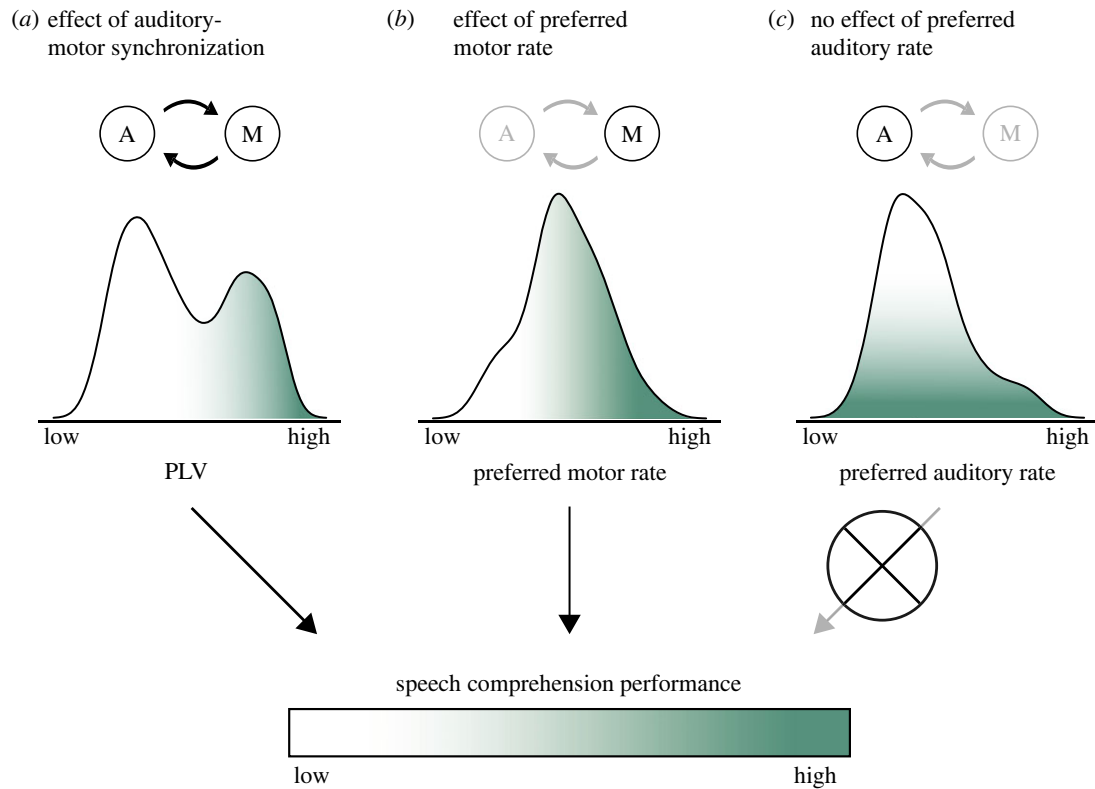
Figure 4. The relationship between speech comprehension performance and (*a*) auditory-motor synchronization, (*b*) preferred rate of the motor and (*c*) preferred rate of the auditory systems. All three predictor variables are represented by the corresponding distribution generated from our experimental data. The present data propose that better speech comprehension at demanding rates—and by hypothesis, auditory behaviour more generally—is accompanied by a higher preferred rate of the motor system as well as stronger auditory-motor synchronization. By contrast, the preferred rate of the auditory system seems not to determine auditory behaviour. Circled A and M illustrate the auditory and motor systems. The arrows connecting them express the relevance of synchronization between the systems for the variable in question.

results expand on findings which showed superior statistical word learning [40] or syllable discrimination [6] for individuals with stronger auditory-motor coupling by showing a similar effect for comprehending more naturalistic, continuous speech. Note that this effect requires further validation as it did not survive control for multiple comparisons (electronic supplementary material, table S5). Additionally, we expected an interaction of syllabic rate and auditory-motor synchronization, as reported for rate discrimination in tone sequences [10]. However, the modulation observed here occurred across all syllabic rates, suggesting that an interaction effect may be masked and compensated for by context and linguistic information in continuous 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. By contrast, in the present experiment, there was no frequency condition between 5 and 10.69 syllables per second.

Importantly, the spontaneous motor production rate affected speech comprehension, suggesting that individuals with a higher spontaneous motor production rate have increased speech comprehension (at the higher range). We replicated this finding in the second experiment. The finding is likely to 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 [35]. Furthermore, our findings are in line with an oscillatory model of speech comprehension [6]. An alternative interpretation of our findings might be that general processes such as vigilance and fatigue are equally reflected in the spontaneous speech motor production rate and the speech comprehension performance. This could be because speech comprehension is tightly intertwined with production, and vigilance effects on production, for example, might similarly affect comprehension. Spontaneous production rates might also be more prone to vigilance effects compared to measures of production performance (e.g. [45]). It is notable that although the preferred spontaneous motor production rates observed here are close to the rates at which speech comprehension has been reported to decline in earlier studies [2,3,18,55], these rates are further apart in our study. The behavioural protocol does not allow to rule out such an alternative interpretation. However, given that no correlation of a demanding cognitive task (digit span) with the spontaneous speech motor production rates was observed (see electronic supplementary material), we consider this unlikely. Furthermore, for the effects of speech auditory-motor synchronization on syllable discrimination, others have ruled out such an interpretation [6].

Interestingly, the preferred auditory rate (approx. 5.57 syllables per second) had no effect on speech comprehension in our study. A possible explanation is that preferred rates in auditory cortex are less flexible compared to preferred rates in motor cortex 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 ($F_{1,162} = 22.39$, $p < 0.001$). Another possibility is that the behavioural estimation of preferred auditory cortex rates were not optimally operationalized. This might also explain the lack of correlation between preferred auditory and spontaneous speech production rates (see electronic supplementary material), which we expected to be correlated. Generally, our behavioural protocol only allows for an indirect assessment of preferred neural rates. Nevertheless, behavioural measures have been regarded as proxy for underlying intrinsic brain rhythms [45,85–87]. Finally, the rates at which speech comprehension decreases are much higher than the preferred auditory and spontaneous speech motor production rates. While the preferred rates were well within the expected range [7,8], the mismatch between maximal comprehension rates and preferred rates was due to the high speech comprehension ability of participants even at high rates.

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 [53,54]. In agreement with these studies, such cognitive variables explained a large amount of variance in speech comprehension. Interestingly, our findings suggest that the faciliatory effect of linguistic predictability is particularly effective at fast rates. Second, we tentatively interpret that facilitation due to linguistic predictability may be used more efficiently from individuals with stronger auditory-motor synchronization. A relevant question arising from this is: under what conditions is the impact of the motor system on speech comprehension the strongest? Previous work observed an impact of the motor system on speech comprehension in demanding listening conditions, such as listening to speech in noise [5,33]. Our data suggest that this view might extend toward conditions of fast speech (which requires more experiments) or might interact with linguistic predictability.

Speech comprehension is a highly predictive process which is affected by different sources of predictions. Here we show that, while speech comprehension is optimal in a preferred auditory temporal regime, the motor-system possibly provides a source for individual flexibility in continuous speech comprehension. Additionally, we report that the well-known facilitatory effect of linguistic predictability on speech comprehension interacts with individual differences in the motor system. This motivates future assessments of how predictions from these systems interact and under what circumstances the human brain relies more on one over the other.

References

1. Rosen S. 1992 Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil. Trans. R. Soc. Lond. B* **336**, 367–373. (doi:10.1098/rstb.1992.0070)

2. Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, Brugge JF. 2009 Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* **29**, 15 564–15 574. (doi:10.1523/JNEUROSCI.3065-09.2009)

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

4. Morillon B, Schroeder CE, Wyart V. 2014 Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* **5**, 5255. (doi:10.1038/ncomms6255)

5. Stokes RC, Venezia JH, Hickok G. 2019 The motor system's [modest] contribution to speech

perception. *Psychon. Bull. Rev.* **26**, 1354–1366. (doi:10.3758/s13423-019-01580-2)

6. Assaneo MF, Rimmele JM, Sanz Perl Y, Poeppel D. 2021 Speaking rhythmically can shape hearing. *Nat. Hum. Behav.* **5**, 71–82. (doi:10.1038/s41562-020-00962-0)

7. Poeppel D, Assaneo MF. 2020 Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* **21**, 322–334. (doi:10.1038/s41583-020-0304-4)

8. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. 2017 Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* **81**, 181–187. (doi:10.1016/j.neubiorev.2017.02.011)

9. Pellegrino F, Coupé C, Marsico E. 2011 Across-language perspective on speech information rate. *Language* **87**, 539–558. (doi:10.1353/lan.2011.0057)

10. Kern P, Assaneo MF, Endres D, Poeppel D, Rimmele JM. 2021 Preferred auditory temporal processing regimes and auditory-motor synchronization.

Psychon. Bull. Rev. **28**, 1860–1873. (doi:10.3758/s13423-021-01933-w)

11. Teng X, Tian X, Rowland J, Poeppel D. 2017 Concurrent temporal channels for auditory processing: oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol.* **15**, e2000812. (doi:10.1371/journal.pbio.2000812)

12. Drake C, Botte MC. 1993 Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Percept. Psychophys.* **54**, 277–286. (doi:10.3758/BF03205262)

13. Teng X, Poeppel D. 2020 Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cereb. Cortex* **30**, 2600–2614. (doi:10.1093/cercor/bhz263)

14. Viemeister NF. 1979 Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.* **66**, 1364–1380. (doi:10.1121/1.383531)

10

royalsocietypublishing.org/journal/rspb   *Proc. R. Soc. B* **290**: 20222410

15. Giraud AL, Poeppel D. 2012 Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517. (doi:10.1038/nn.3063)

16. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S. 2013 Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* **11**, e1001752. (doi:10.1371/journal.pbio.1001752)

17. Peelle JE, Davis MH. 2012 Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* **3**, 320. (doi:10.3389/fpsyg.2012.00320)

18. Ghitza O, Greenberg S. 2009 On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113–126. (doi:10.1159/000208934)

19. Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. 2007 Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127–1134. (doi:10.1016/j.neuron.2007.09.038)

20. Luo H, Poeppel D. 2007 Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001–1010. (doi:10.1016/j.neuron.2007.06.004)

21. Greenberg S, Carvey H, Hitchcock L, Chang S. 2003 Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phonetics* **31**, 465–485. (doi:10.1016/j.wocn.2003.09.005)

22. Morillon B, Baillet S. 2017 Motor origin of temporal predictions in auditory attention. *Proc. Natl Acad. Sci. USA* **114**, E8913–E8921. (doi:10.1073/pnas.1705373114)

23. Schubotz RI. 2007 Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* **11**, 211–218. (doi:10.1016/j.tics.2007.02.006)

24. Grahn JA, Brett M. 2007 Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* **19**, 893–906. (doi:10.1162/jocn.2007.19.5.893)

25. Cason N, Astésano C, Schön D. 2015 Bridging music and speech rhythm: rhythmic priming and audio–motor training affect speech perception. *Acta Psychol.* **155**, 43–50. (doi:10.1016/j.actpsy.2014.12.002)

26. Du Y, Zatorre RJ. 2017 Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. USA* **114**, 13 579–13 584. (doi:10.1073/pnas.1712223114)

27. Grahn JA, Rowe JB. 2009 Feeling the Beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* **29**, 7540–7548. (doi:10.1523/JNEUROSCI.2018-08.2009)

28. Cheung C, Hamilton LS, Johnson K, Chang EF. 2016 The auditory representation of speech sounds in human motor cortex. *Elife* **5**, e12577. (doi:10.7554/eLife.12577)

29. Evans S, Davis MH. 2015 Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. *Cereb. Cortex* **25**, 4772–4788. (doi:10.1093/cercor/bhv136)

30. Hickok G, Poeppel D. 2007 The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402. (doi:10.1038/nrn2113)

31. Morillon B, Arnal LH, Schroeder CE, Keitel A. 2019 Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci. Biobehav. Rev.* **107**, 136–142. (doi:10.1016/j.neubiorev.2019.09.012)

32. Scott SK, McGettigan C, Eisner F. 2009 A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* **10**, 295–302. (doi:10.1038/nrn2603)

33. Wu ZM, Chen ML, Wu XH, Li L. 2014 Interaction between auditory and motor systems in speech perception. *Neurosci. Bull.* **30**, 490–496. (doi:10.1007/s12264-013-1428-6)

34. Rogalsky C *et al.* 2022 The neuroanatomy of speech processing: a large-scale lesion study. *J. Cogn. Neurosci.* **34**, 1355–1375.

35. Assaneo MF, Poeppel D. 2018 The coupling between auditory and motor cortices is rate-restricted: evidence for an intrinsic speech-motor rhythm. *Sci. Adv.* **4**, eaao3842. (doi:10.1126/sciadv.aao3842)

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

37. Keitel A, Gross J. 2016 Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biol.* **14**, e1002498. (doi:10.1371/journal.pbio.1002498)

38. Lubinus C, Orpella J, Keitel A, Gudi-Mindermann H, Engel AK, Roeder B, Rimmele JM. 2021 Data-driven classification of spectral profiles reveals brain region-specific plasticity in blindness. *Cereb. Cortex* **31**, 2505–2522. (doi:10.1093/cercor/bhaa370)

39. Rosanova M, Casali A, Bellina V, Resta F, Mariotti M, Massimini M. 2009 Natural frequencies of human corticothalamic circuits. *J. Neurosci.* **29**, 7679–7685. (doi:10.1523/JNEUROSCI.0445-09.2009)

40. Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D. 2019 Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* **22**, 627–632. (doi:10.1038/s41593-019-0353-z)

41. Park H, Ince RAA, Schyns PG, Thut G, Gross J. 2015 Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**, 1649–1653. (doi:10.1016/j.cub.2015.04.049)

42. Haegens S, Zion Golumbic E. 2018 Rhythmic facilitation of sensory processing: a critical review. *Neurosci. Biobehav. Rev.* **86**, 150–165. (doi:10.1016/j.neubiorev.2017.12.002)

43. Rimmele JM, Morillon B, Poeppel D, Arnal LH. 2018 Proactive sensing of periodic and aperiodic auditory patterns. *Trends Cogn. Sci.* **22**, 870–882. (doi:10.1016/j.tics.2018.08.003)

44. Assaneo MF, Rimmele JM, Orpella J, Ripollés P, de Diego-Balaguer R, Poeppel D. 2019 The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Front. Integr. Neurosci.* **13**, 28. (doi:10.3389/fnint.2019.00028)

45. McPherson T, Berger D, Alagapan S, Fröhlich F. 2018 Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. *Sci. Rep.* **8**, 11782. (doi:10.1038/s41598-018-29267-z)

46. Rimmele JM, Kern P, Lubinus C, Frieler K, Poeppel D, Assaneo MF. 2022 Musical sophistication and speech auditory-motor coupling: easy tests for quick answers. *Front. Neurosci.* **15**, 764342. (doi:10.3389/fnins.2021.764342)

47. Emmorey K, Giezen MR, Petrich JAF, Spurgeon E, O'Grady Farnady L. 2017 The relation between working memory and language comprehension in signers and speakers. *Acta Psychol.* **177**, 69–77. (doi:10.1016/j.actpsy.2017.04.014)

48. Arnal LH, Wyart V, Giraud AL. 2011 Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* **14**, 797–801. (doi:10.1038/nn.2810)

49. Zhang Y, Frassinelli D, Tuomainen J, Skipper JI, Vigliocco G. 2021 More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proc. R. Soc. B.* **288**, 20210500. (doi:10.1098/rspb.2021.0500)

50. Lewis AG, Bastiaansen M. 2015 A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* **68**, 155–168. (doi:10.1016/j.cortex.2015.02.014)

51. Kuperberg GR, Jaeger TF. 2016 What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* **31**, 32–59. (doi:10.1080/23273798.2015.1102299)

52. Grant KW, Seitz PF. 2000 The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *J. Acoust. Soc. Am.* **107**, 1000–1011. (doi:10.1121/1.428280)

53. Obleser J, Wise RJS, Alex Dresner M, Scott SK. 2007 Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* **27**, 2283–2289. (doi:10.1523/JNEUROSCI.4663-06.2007)

54. Obleser J, Kotz SA. 2010 Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex.* **20**, 633–640. (doi:10.1093/cercor/bhp128)

55. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. 2001 Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl Acad. Sci. USA.* **98**, 6.

56. Dupoux E, Green K. 1997 Perceptual adjustment to highly compressed speech. *J. Exp. Psychol.: Hum. Percept. Perform.* **23**, 914–927. (doi:10.1037/0096-1523.23.3.914)

57. Pefkou M, Arnal LH, Fontolan L, Giraud AL. 2017 θ-band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *J. Neurosci.* **37**, 7930–7938. (doi:10.1523/JNEUROSCI.2882-16.2017)

11

royalsocietypublishing.org/journal/rspb   *Proc. R. Soc. B* **290**: 20222410

58. Beukelman DR, Yorkston KM. 1979 The relationship between information transfer and speech intelligibility of dysarthric speakers. *J. Commun. Disord.* **12**, 189–196. (doi:10.1016/0021-9924(79)90040-6)

59. Schiavetti N, Sitler RW, Metz DE, Houde RA. 1984 Prediction of contextual speech intelligibility from isolated word intelligibility measures. *J. Speech Lang. Hear Res.* **27**, 623–626. (doi:10.1044/jshr.2704.623)

60. Schiavetti N. 1992 Scaling procedures for the measurement of speech intelligibility. In *Studies in speech pathology and clinical linguistics* (ed. RD Kent), p. 11. Amsterdam, The Netherlands: John Benjamins Publishing Company.

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

62. Richardson JTE. 2007 Measures of short-term memory: a historical review. *Cortex* **43**, 635–650. (doi:10.1016/S0010-9452(08)70493-3)

63. Olsthoorn NM, Andringa S, Hulstijn JH. 2014 Visual and auditory digit-span performance in native and non-native speakers. *Int. J. Bilingual..* **18**, 663–673. (doi:10.1177/1367006912466314)

64. de Jong NH, Wempe T. 2009 Praat script to detect syllable nuclei and measure speech rate automatically. *Behav. Res. Methods* **41**, 385–390. (doi:10.3758/BRM.41.2.385)

65. Boersma P, Weenik D. 2020 Praat: doing phonetics by computer [Computer program]. Version 6.0.40. Retrieved from http://www.praat.org/.

66. Lachaux JP, Rodriguez E, Martinerie J, Varela FJ. 1999 Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* **8**, 194–208. (doi:10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C)

67. Smith ZM, Delgutte B, Oxenham AJ. 2002 Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90. (doi:10.1038/416087a)

68. MacQueen J. 1967 Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.* **1**, 281–297.

69. Jurafsky D, Martin JH. 2009 *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

70. Chien JT, Ku YC. 2016 Bayesian recurrent neural network for language modeling. *IEEE Trans. Neural Netw. Learn. Syst.* **27**, 361–374. (doi:10.1109/TNNLS.2015.2499302)

71. Mikolov T. 2010 Recurrent neural network based language model. See http://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf.

72. Merity S, Keskar NS, Socher R. 2017 Regularizing and optimizing LSTM language models. *arXiv*. See http://arxiv.org/abs/1708.02182.

73. Fernandez J, Downey D. 2018 Sampling informative training data for RNN language models. In *Proceedings of ACL 2018, Student Research Workshop*, pp. 9–13. Melbourne, Australia: Association for Computational Linguistics. See http://aclweb.org/anthology/P18-3002.

74. Baayen RH, Davidson DJ, Bates DM. 2008 Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**, 390–412. (doi:10.1016/j.jml.2007.12.005)

75. Barr DJ, Levy R, Scheepers C, Tily HJ. 2013 Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* **68**, 255–278. (doi:10.1016/j.jml.2012.11.001)

76. Sassenhagen J, Alday PM. 2016 A common misapplication of statistical inference: nuisance control with null-hypothesis significance tests. *Brain Lang.* **162**, 42–45. (doi:10.1016/j.bandl.2016.08.001)

77. Schielzeth H. 2010 Simple means to improve the interpretability of regression coefficients: *interpretation of regression coefficients*. *Methods Ecol. Evol.* **1**, 103–113. (doi:10.1111/j.2041-210X.2010.00012.x)

78. Fox J, Weisberg S. 2019 *An R companion to applied regression*, 3rd edn. Thousand Oaks, CA: Sage Publications.

79. Assaneo MF, Orpella J, Ripollés P, Noejovich L, López-Barroso D, de Diego-Balaguer R, Poeppel D. 2020 Population-level differences in the neural substrates supporting statistical learning. *bioRxiv*. (doi:10.1101/2020.07.03.187260)

80. Ghitza O. 2014 Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. *Front. Psychol.* **5**, 652. (doi:10.3389/fpsyg.2014.00652)

81. Giroud J, Lerousseau JP, Pellegrino F, Morillon B. 2021 The channel capacity of multilevel linguistic features constrains speech comprehension. *Cognition* **232**, 105345. (doi:10.1016/j.cognition.2022.105345)

82. Verschueren E, Gillis M, Decruy L, Vanthornhout J, Francart T. 2022 Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. *J. Neurosci.* **42**, 7442–7453. (doi:10.1523/JNEUROSCI.0259-22.2022)

83. Doelling KB, Arnal LH, Ghitza O, Poeppel D. 2014 Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85**, 761–768. (doi:10.1016/j.neuroimage.2013.06.035)

84. Ghitza O. 2013 The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* **4**, 138. (doi:10.3389/fpsyg.2013.00138)

85. McAuley JD, Jones MR, Holub S, Johnston HM, Miller NS. 2006 The time of our lives: life span development of timing and event tracking. *J. Exp. Psychol.: General* **135**, 348–367. (doi:10.1037/0096-3445.135.3.348)

86. Michaelis K, Wiener M, Thompson JC. 2014 Passive listening to preferred motor tempo modulates corticospinal excitability. *Front. Hum. Neurosci.* **8**, 252. (doi:10.3389/fnhum.2014.00252)

87. Provasi J, Anderson DI, Barbu-Roth M. 2014 Rhythm perception, production, and synchronization during the perinatal period. *Front. Psychol.* **5**, 1048. (doi:10.3389/fpsyg.2014.01048)

88. Lubinus C, Keitel A, Obleser J, Poeppel D, Rimmele JM. In press. Explaining flexible continuous speech comprehension from individual motor rhythms. *Neuroscience.* (doi:10.1101/2022.04.01.486685)

89. Lubinus C, Keitel A, Obleser J, Poeppel D, Rimmele JM. 2023 Explaining flexible continuous speech comprehension from individual motor rhythms. Figshare. (doi:10.6084/m9.figshare.c.6431747)