



Preferred auditory temporal processing regimes and auditory-motor synchronization

Pius Kern¹ · M. Florencia Assaneo² · Dominik Endres³ · David Poeppel^{1,4,5} · Johanna M. Rimmele^{1,5}

Accepted: 7 April 2021
© The Author(s) 2021

Abstract

Decoding the rich temporal dynamics of complex sounds such as speech is constrained by the underlying neuronal-processing mechanisms. Oscillatory theories suggest the existence of one optimal perceptual performance regime at auditory stimulation rates in the delta to theta range (< 10 Hz), but reduced performance in the alpha range (10–14 Hz) is controversial. Additionally, the widely discussed motor system contribution to timing remains unclear. We measured rate discrimination thresholds between 4 and 15 Hz, and auditory-motor coupling strength was estimated through a behavioral auditory-motor synchronization task. In a Bayesian model comparison, high auditory-motor synchronizers showed a larger range of constant optimal temporal judgments than low synchronizers, with performance decreasing in the alpha range. This evidence for optimal processing in the theta range is consistent with preferred oscillatory regimes in auditory cortex that compartmentalize stimulus encoding and processing. The findings suggest, remarkably, that increased auditory-motor synchronization might extend such an optimal range towards faster rates.

Keywords Auditory · Perceptual constraints · Temporal sensitivity · Auditory-motor synchronization

Introduction

Natural sounds such as speech or music contain temporal structure at multiple time scales. Particularly slow acoustic modulations in the delta-theta range (2–9 Hz) are considered crucial for speech and music processing (Ding et al., 2017; Pellegrino et al., 2011; Singh & Theunissen, 2003). Such natural statistics are arguably not accidental and co-occur with

potential neuronal coding principles in auditory cortex (Ravignani et al., 2019; Singh & Theunissen, 2003). Support for this proposal comes from electrophysiological studies that identified endogenous oscillations in auditory cortex in the delta-theta band (Giraud et al., 2007; Keitel & Gross, 2016; Lakatos et al., 2005; Lubinus et al., 2019). By entraining to acoustic signals at these time scales, neuronal oscillations in auditory cortex might contribute to the processing of temporal information in sound (Ghitza, 2012; Giraud & Poeppel, 2012; Gross et al., 2013; McAuley & Jones, 2003; Miller & McAuley, 2005; Rimmele, Gross, et al., 2018). The optimal processing range of neuronal populations should, therefore, constrain auditory perception by facilitating auditory temporal processing within this range (Haegens & Zion Golumbic, 2018; Rimmele, Morillon, et al., 2018).

Perceptual constraints, such as decreased neuronal tracking of speech and reduced speech comprehension at fast rates outside of the presumably optimal range, have been shown previously (Ahissar et al., 2001; Brungart et al., 2007; Doelling et al., 2014; Ghitza & Greenberg, 2009). Similarly, for amplitude-modulated sounds and isochronous tone sequences, reduced neuronal tracking (Teng et al., 2017; Teng & Poeppel, 2020) and reduced temporal sensitivity have been observed at stimulus rates associated with the higher alpha

✉ Johanna M. Rimmele
johanna.rimmele@ae.mpg.de

¹ Department of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt/M, Germany

² Instituto de Neurobiología, Universidad Nacional Autónoma de México Juriquilla, Campus UNAM 3001, 76230 Juriquilla, Qro., Mexico

³ Department of Psychology, Philipps University Marburg, Gutenbergstraße 18, 35032 Marburg, Germany

⁴ Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA

⁵ Max Planck NYU Center for Language, Music, and Emotion, Frankfurt/M, Germany, NY, USA

range compared to lower rates (Drake & Botte, 1993; Friberg & Sundberg, 1995; Teng et al., 2017; Viemeister, 1979). Overall, there is considerable evidence from studies on rate perception for “constant” optimal auditory temporal processing in the theta range (4–8 Hz, including lower rates in the delta range 2–4 Hz). This has typically been assessed through relative difference thresholds for rate discrimination, that is, the minimal difference between two stimulation rates necessary for discrimination normalized by the standard rate. Relative difference thresholds have been shown to be lowest and constant in the theta range, which is commonly interpreted as a zone of optimal temporal processing, referring to Weber’s law (Drake & Botte, 1993; Friberg & Sundberg, 1995; Viemeister, 1979). According to Weber’s law, the ability to distinguish stimulation rates is proportional to the frequency of the presentation rate. The absolute rate difference necessary for discrimination, thus, scales with the stimulation rate, resulting in a constant relative difference threshold. Although the temporal sensitivity has been shown to be constant at low stimulation rates (corresponding roughly to the delta-theta band, in neural terms), the onset of the decrease in temporal sensitivity is controversial. While some studies already find higher relative thresholds for rates around 8 Hz (Drake & Botte, 1993; Friberg & Sundberg, 1995; ten Hoopen et al., 1994; ten Hoopen et al., 2011), others report a threshold increase at 10 Hz (McAuley & Kidd, 1998; Michon, 1964), 12 Hz (Ehrlé & Samson, 2005; Elliott & Theunissen, 2009), 16 Hz (Nordmark, 1968; Viemeister, 1979), or even 40 Hz (Dau et al., 1997; Sheft & Yost, 1990). In these studies, typically only a coarse range of standard (modulation) rates in the upper theta and alpha range was tested.

Here, we investigate whether interindividual differences in auditory-motor coupling strength might contribute to the controversial or mixed findings regarding the (onset of) sensitivity changes, using a behavioral paradigm. “Temporal predictions” from motor cortex have been shown to modulate auditory processing in studies presenting periodic tone sequences (Arnal et al., 2015; Morillon et al., 2014; Morillon & Baillet, 2017) or continuous speech (Keitel et al., 2018; Park et al., 2015), even during passive listening (Chen et al., 2008; Grahn & Rowe, 2013). Top-down effects from motor cortex seem to particularly affect auditory processing and facilitate behavior during demanding listening situations (Stokes et al., 2019; Wu et al., 2014). Interestingly, it has been proposed based on MEG findings that the right hemispheric lateralization of speech processing is reduced at more challenging fast stimulation rates (Assaneo, Rimmele, et al., 2019), whereas such lateralization might be linked to motor top-down predictions (Tang et al., 2020). Recently, Assaneo et al. (2019) developed a simple behavioral protocol to measure spontaneous auditory-motor synchronization. The synchronization of one’s own speech production to a perceived speech stream differed

widely but systematically across individuals, revealing a bimodal distribution of high and low synchronizers. On a neuronal level, high synchronizers showed stronger functional and structural connectivity between frontal speech-motor and auditory cortices, rendering this behavioral protocol suitable to estimate individual differences in auditory-motor cortex coupling strength. Furthermore, rhythmic speech production more strongly modulated perception in high synchronizers compared to low synchronizers, suggesting increased motor top-down predictions with increasing auditory-motor coupling strength (Assaneo et al., 2020; Assaneo, Ripollés, et al., 2019).

Here, we investigate the following hypotheses: (1) There is a particular auditory sensitivity for processing stimulation rates in the theta range that decreases at higher rates in the alpha range, reflected in increasing rate-discrimination thresholds. (2) Interindividual differences in the auditory-motor synchronization, measured behaviorally with the spontaneous speech synchronization test (SSS-test; as an estimate of fronto-temporal structural connectivity strength) (Assaneo, Ripollés, et al., 2019), modulate auditory temporal processing sensitivity. Based on previous findings, we expect (1) an overall increased auditory temporal sensitivity in high compared to low synchronizers due to increased temporal top-down predictions from production areas, or, alternatively, (2) the benefit might particularly occur at higher “non-optimal” rates. In an adaptive weighted up-down staircase procedure, we tested auditory discrimination thresholds for a fine-grained range of rates in the theta to alpha range (4–15 Hz). Bayesian model comparison was used to test our hypotheses. To validate the threshold measure, we used the method of constant stimuli to measure rate discrimination performance at two rates in the theta and alpha range (4, 11.86 Hz).

Method

Participants

All participants reported normal hearing and absence of any type of dyslexic, neurological, or psychiatric disorder or intake of psychotropic substances during the last 6 months. The experimental procedures were ethically approved by the Ethics Council of the Max Planck Society (No. 2017_12). All participants gave written informed consent prior to the study and received monetary compensation. Participants were excluded from the statistical analyses because of outlier behavioral performance ($n = 3$; mean relative difference threshold in at least one psychophysical procedure exceeded the median ± 3 median absolute deviation; Leys et al., 2013) or because the two runs in the SSS-test were inconsistent ($n = 2$). The final sample included 55 participants (age range: 19–32

years ($M = 25.38$, $SD = 3.78$), 27 female, three left-handed), clustered into 35 high and 20 low synchronizers.

Procedure

The experiment was conducted in a sound-attenuated experiment booth. Stimulus presentation and response acquisition were run on a Windows 7 computer using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for Matlab R2017a (version 9.2). Responses were collected on a standard computer keyboard. Auditory stimuli were presented binaurally at a 44,100 Hz sampling rate via a RME Fireface UCX audio interface, a Phone Amp Lake People G109 headphone amplifier, and Beyerdynamic DT770 Pro (80 Ohm) closed-back headphones. In the SSS-test, we used Etymotic Research 3C insert earphones (50 Ohm) with foam ear tips and a directional gooseneck condenser microphone (Shure MX418 Microflex) placed about 5 cm in front of the participant for speech recording.

Rate-discrimination task

We adopted an auditory two-interval forced-choice (2IFC) rate discrimination task (Fig. 1a), in which a standard and a faster comparison sequence were presented in random order (Drake & Botte, 1993). Participants were asked to indicate as accurately and quickly as possible by keypress whether the first or second sequence was faster. They were instructed to guess whenever they failed to discriminate the rate and to refrain from any movement to the beat such as tapping.

We generated isochronous sequences by concatenating a 15-ms 440-Hz sinusoidal pure tone (5-ms rise/fall of the sound envelope) with silent intervals, depending on the stimulation rate. One of the two sequences consisted of five and the other of seven tones, randomized trial-by-trial. Auditory stimuli were presented at 70 dB SPL (peak amplitude normalization).

On each trial, a fixation cross appeared 500–750 ms prior to stimulus onset (random uniform distribution). Standard and comparison sequences were presented in random order separated by an inter-sequence interval (random uniform distribution between 4.5 and 5.5 times the tone inter-onset interval of the first sequence). After giving their response, participants received visual feedback (correct, incorrect) in the weighted up-down procedure, but not in the method of constant stimuli. The intertrial interval was 1,000–1,500 ms (random uniform distribution).

Weighted up-down procedure

We used an adaptive weighted up-down (WUD) staircase method implemented in the Palamedes toolbox (Kingdom & Prins, 2010) to measure discrimination thresholds block-wise for eight standard rates linearly spaced from 4 to 15 Hz (Fig.

1b). The relative difference between the standard and the comparison rate was reduced by a certain step size given a correct response and was increased by three times that step size following an incorrect response to converge on a 75% correct level, defined as the threshold (Kaernbach, 1991). All parameters were chosen in terms of the relative difference between standard and comparison rate (in Hz), given by $d_r = (rate_{comparison} - rate_{standard})/rate_{standard}$. Start values for each standard rate condition (linearly increasing across rates from 20–27%) were selected well above the expected threshold to observe a convergence to threshold level and to capture the threshold area in all participants at each standard rate (Green, 1990) based on previous studies and pilot results. We chose linearly increasing start values to avoid differential adaptation and learning effects as higher thresholds were expected at higher rates. The initial step size (1%) was divided in half after six (0.5%) and again after 12 reversals (0.25%), that is, after 12 changes of direction in the adaptive run (e.g., up to down). Progressively decreasing step sizes were expected to yield reliable and fine-grained threshold estimates more efficiently (Levitt, 1971; Rammseyer, 1992). In case the relative rate difference became smaller or equal to one step-size, the current difference was divided in half after a correct response.

The relative difference threshold for each standard rate was computed as the mean of the last six reversals. Short breaks were included every 20 trials. One adaptive run ended after 18 reversals. To familiarize participants with the task, a minimum of five and a maximum of 15 training trials (75% correct criterion) were completed at the beginning of each block. In training trials, the relative difference between the standard and the comparison rate was 10% higher compared to the start values in the main part (i.e. 30–37%, 1% steps).

Method of constant stimuli procedure

We used the method of constant stimuli (CS) to validate the WUD threshold estimates by fitting the psychometric function at two standard rates of main interest, 4 and 11.86 Hz (Fig. 1c). The psychometric function describes the relationship between a participant's response behavior and a stimulus feature (Wichmann & Hill, 2001). We measured performance (percentage correct) at seven comparison levels (relative rate difference in the comparison sequence). For every participant, we calculated the rate of the comparison sequences by multiplying a seven-point scale logarithmically spaced from 0.2 to 3 with the individual WUD threshold to capture the whole range of the psychometric function from 50% chance level to 100% correct (see: Herbst & Obleser, 2019). Every run contained one trial of each comparison level in random order (mixed presentation). There were 30 runs in total, amounting to 30 trials measured per comparison level. Pauses were included after 21 trials. The

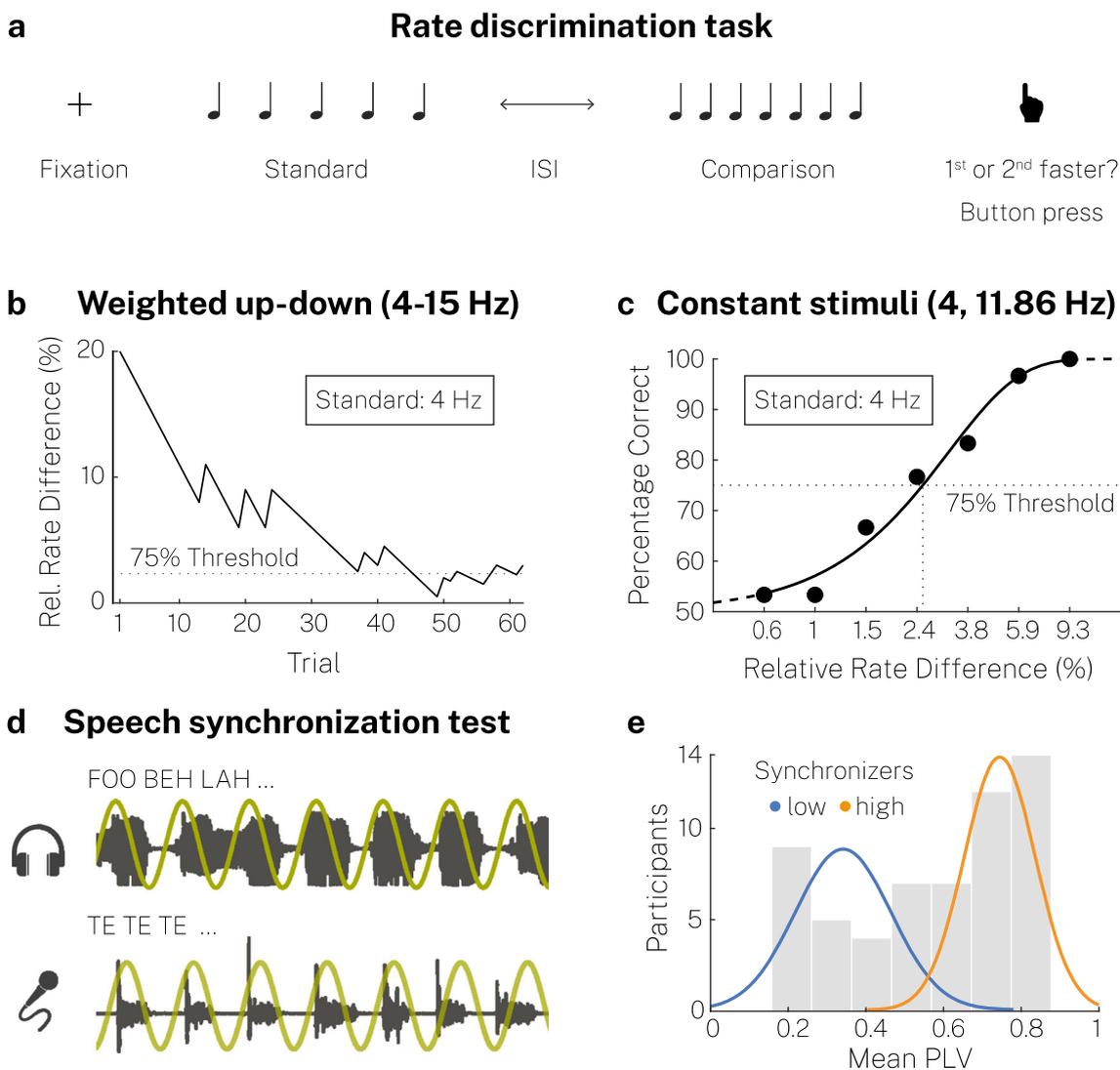


Fig. 1 Experimental design. **(a)** In a two-interval forced-choice (2-IFC) task, participants judged whether the first or second of two isochronous tone sequences was faster. Standard and comparison sequence were randomly presented at either position. **(b)** Using a weighted up-down staircase procedure, we measured relative difference thresholds at eight standard rates between 4 and 15 Hz (exemplary run from one participant at a standard rate of 4 Hz). **(c)** Additionally, the discrimination performance was measured in a constant stimuli procedure for two standard rates at seven stimulus levels (i.e., levels of the comparison sequence). The psychometric function was fitted to estimate the relative difference threshold (exemplary data and psychometric function from one participant at a

standard rate of 4 Hz). **(d)** In the auditory-motor speech synchronization test, participants repeatedly whispered the syllable /te/ while listening to an isochronous train of random syllables (Assaneo, Ripollés, et al., 2019). The synchronization between the produced and presented speech was measured (icons based on resources from [Flaticon.com](https://flaticon.com)). Each participant completed two runs. **(e)** Histogram of the synchronization measurements, computed as the phase-locking value (PLV) between the envelope of the produced and perceived speech signals. PLV values were averaged across runs. We clustered participants into two groups, low and high synchronizers. Lines represent normal distributions fitted to each cluster

procedure was completed for each standard rate separately in randomized order across participants.

We fitted a Weibull function to the individual data at both standard rates using Bayesian inference methods implemented in the Psignifit Toolbox 4 (Schütt et al., 2016). The guess rate was fixed at 50%, which is the chance level given the 2IFC task, and three parameters were estimated: the 75% correct threshold, the width, and the lapse rate. We assessed the goodness-of-fit for individual

data by comparing the deviance to a parametric bootstrap sample distribution ($N = 10,000$; 95% percentile).

Auditory-motor speech-synchronization test

In the SSS-test (Fig. 1d; for details see Assaneo, Ripollés, et al., 2019) we measured participants' ability to synchronize their speech production to a heard syllable sequence. First, participants wearing earphones were instructed to cautiously

increase the volume of a background babble while continuously whispering the syllable /te/, until they could not hear their own speech anymore (start sound level: 70 dB SPL, maximum: 95 dB SPL). In two training trials, participants listened to a 10-s periodic syllable /te/ sequence (at 4.5 Hz) and were asked to whisper /te/ at the same rate for 10 s directly afterwards. Finally, a 70-s random syllable train with a progressively increasing rate ($M = 4.5$ Hz, range: 4.3–4.7 Hz, steps: 0.1 Hz after 60 syllables) was presented. The random syllable stream audio file was created using the MBROLA synthesizer (male American English voice, 200-Hz pitch) (Dutoit et al., 1996). Participants were instructed to whisper /te/ continuously in synchrony with the audio. Each participant completed two blocks of the test.

We quantified auditory-motor speech synchronization by calculating the phase-locking value (PLV) (Lachaux et al., 1999) between the cochlear envelope of the audio stimulus and the envelope of the recorded speech signal (time windows 5 s, overlap 2 s). The cochlear envelope was determined using the NSL Auditory Model toolbox for Matlab (auditory channels: 180–7,246 Hz). Phases of both envelopes were calculated via the Hilbert transform after resampling to 100 Hz and band-pass filtering (3.5–5.5 Hz). To classify high and low synchronizers, we performed k-means clustering (Arthur & Vassilvskii, 2007) of participants into two clusters based on mean PLVs across both blocks (Fig. 1e, mean PLV across high synchronizers = 0.74, $SD = 0.1$, mean PLV across low synchronizers = 0.34, $SD = 0.12$). Note that one participant was not assigned to one cluster consistently given the nature of the clustering algorithm. Our findings were not altered by the classification of this participant as a high synchronizer (reported here, because this was the case in ~58% of 10,000 simulation runs), a low synchronizer, or an exclusion.

Finally, participants filled out a questionnaire on demographic data, task strategies, and musical sophistication (Goldsmiths Musical Sophistication Index (Gold-MSI); Müllensiefen et al., 2014).

Analysis

Data analysis was performed in Matlab 2018b (version 9.5) and in R (version 3.6.1). In order to specifically test our hypotheses on temporal sensitivity in the WUD procedure, we performed a Bayesian model comparison using the RStan software (version 2.19.2) (Carpenter et al., 2017). We compared five different model types (Fig. 2a), each specified by a set of parameters θ , to explain the observations D (i.e., the relative difference thresholds) depending on auditory-motor speech synchronization behavior and the rate of the standard tone sequences. For model simplicity, we assumed that the observations D are independent and identically distributed given the models' parameters. We approximated the data generation process by log-normal distributions, given that

threshold values could only be positive and were expectedly right-skewed. In an additional control analysis, Gaussian normal distributions truncated at zero yielded lower posterior probabilities, while the main results were equivalent. All model parameters were set to be lower bounded by 0, given the definition of the relative difference threshold and our prior assumptions. We further assumed that the variance in each standard rate condition and subgroup (high and low synchronizers) was drawn from the same distribution.

In the null model M_1 , the relative difference threshold was constant at μ across all standard rates.

$$D \sim \text{LogNormal}(\log(\mu), \sigma^2), \text{ where } \mu, \sigma^2 \in \mathbb{R}_0^+$$

The prior distribution for μ was set to be a normal distribution with prior mean μ_0 and prior variance $\sigma_{\mu_0}^2$. Summarizing previous results on rate discrimination in isochronous sequences (Drake & Botte, 1993; Ehrlé & Samson, 2005; Friberg & Sundberg, 1995; ten Hoopen et al., 1994; ten Hoopen et al., 2011), we expected the mean threshold μ across the tested frequency range to be 5% on average, with single observations ranging from minimally 0.1% to maximally 50%. We doubled the maximal variance possible given that range to obtain a prior estimate for $\sigma_{\mu_0}^2$. The same prior distribution for μ was used in the remaining models as well.

$$\begin{aligned} \mu &\sim \text{Normal}(\mu_0, \sigma_{\mu_0}^2), \text{ where } \mu_0 = 0.05, \sigma_{\mu_0}^2 \\ &= 2 \frac{(0.5 - 0.001)^2}{4} \approx 0.125 \end{aligned}$$

The prior distribution for σ^2 was specified as a uniform distribution with a lower bound $\alpha = 0$. The upper bound γ was determined by the maximum expected variance, given an expected minimal threshold of 0.1% and a maximum threshold of 50%. The same prior distribution for σ^2 was used in the remaining models as well.

$$\sigma^2 \sim \text{Uniform}(\alpha, \gamma), \text{ where } \alpha = 0, \gamma$$

$$= \frac{(\log(0.5) - \log(0.001))^2}{4} \approx 9.66$$

Model M_2 included the group baseline difference in relative difference thresholds between high and low synchronizers, coded by a group indicator variable $j = \{1, 2\}$ for every data point and two group-specific parameters μ_j .

$$D \sim \text{LogNormal}(\log(\mu_j), \sigma^2), \text{ where } j \in \{1, 2\} \text{ and } \mu_j \in \mathbb{R}_0^+$$

Based on differences between musicians and non-musicians in previous studies, we expected the threshold to be on average 4% higher in low synchronizers. Correspondingly, we set the mean of the Gaussian prior to

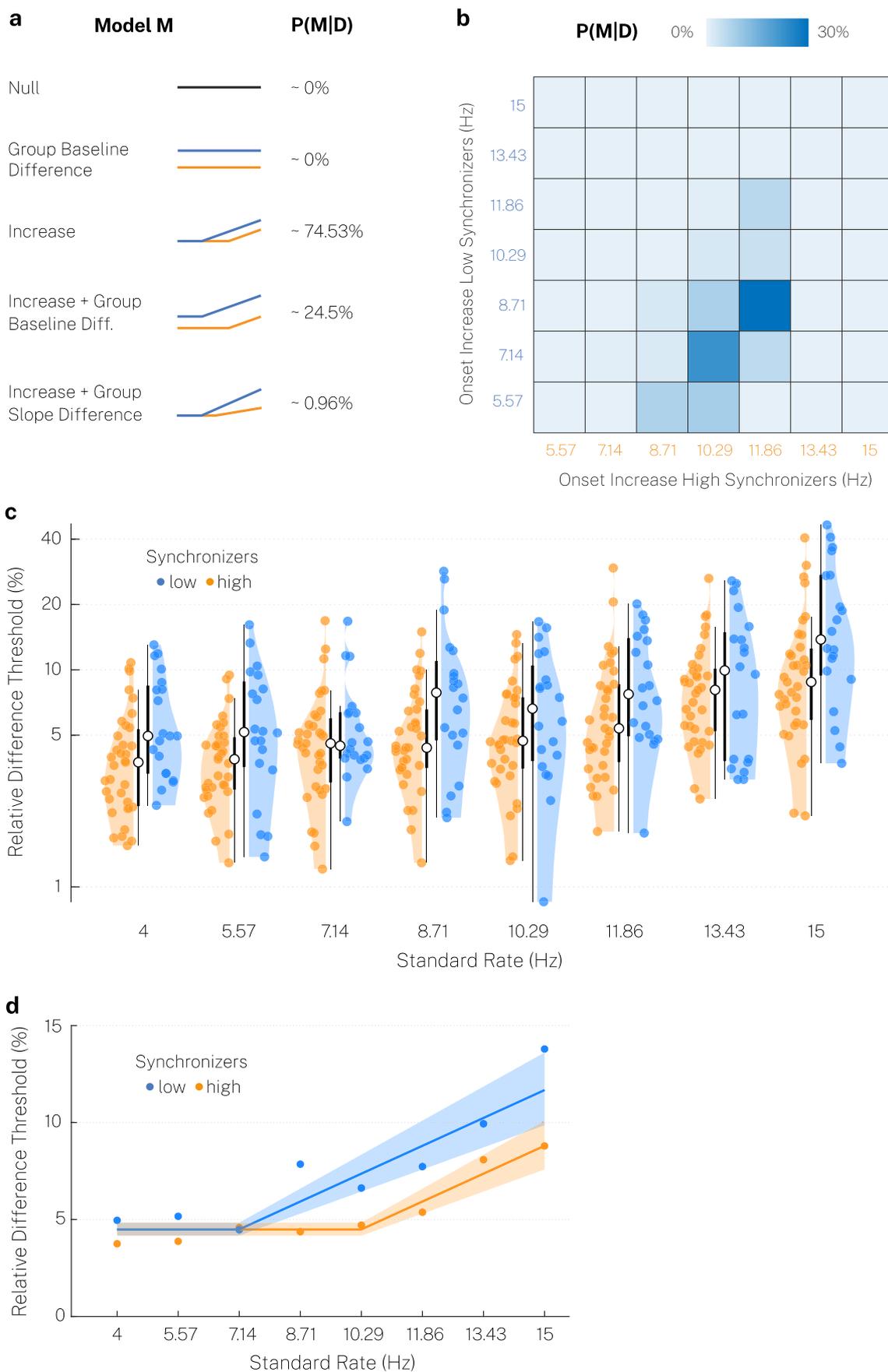


Fig. 2 Weighted up-down method indicates an optimal processing range modulated by auditory-motor coupling. (a) We tested five model types, which included a constant threshold (Null), a difference between high and low synchronizers (Group Baseline Difference), a linear threshold increase (Increase), an additional baseline difference (Increase + Group Baseline Difference), or an additional variable slope (Increase + Group Slope Difference). In the models that included a threshold increase, we considered all seven possible different onset points (see b), that could vary between groups. P(M|D) gives the marginal posterior probability of each model type given the data marginalized over possible onset points. (b) Heatmap illustrating the marginal posterior probabilities P(M|D) of each onset point of the threshold increase in high and low synchronizers marginalized over model types. (c) Relative difference thresholds for high and low synchronizers. Colored dot: individual participant, white dot: median, thick line: quartiles, thin line: quartiles $\pm 1.5 \times$ interquartile range. (d) Median relative difference thresholds (dots) and model predictions of the best model (line) with 95% confidence interval (shaded area) for high and low synchronizers

5% for high synchronizers and 9% for low synchronizers. The prior variance $\sigma_{\mu_{j-0}}^2$ was assumed to be the same as in M_j for both groups.

$$\mu_j \sim \text{Normal}\left(\mu_{j-0}, \sigma_{\mu_{j-0}}^2\right), \text{ where } j \in \{1, 2\}, \mu_{1-0} = 0.05, \mu_{2-0} = 0.09, \sigma_{\mu_{j-0}}^2 \approx 0.125$$

In the models of type ‘‘Increase,’’ we included the hypothesized departure from a constant relative threshold μ in the theta range and modelled a linear threshold increase by including a slope parameter β . For each participant, the threshold increase at a particular starting point was coded by an indicator variable x_k (length corresponding to the eight standard rates tested). We considered all possible seven points k , where the threshold could start to differ from a constant baseline threshold (5.57–15 Hz). Thus, the first point of the indicator variable, corresponding to the standard rate 4 Hz, was always zero. For example, an increase starting at 5.57 Hz was coded by $x_{5.57} = \{0, 1, 2, 3, 4, 5, 6, 7\}$. The starting point could vary in high versus low synchronizers (again coded by the indicator variable $j: x_{k,j}$), which resulted in a total of 49 possible combinations (see also Fig. 2b), tested in models M_3 to M_{51} .

$$D \sim \text{LogNormal}(\mu + x_{k,j}\beta, \sigma^2), \text{ where } \mu, \beta, \sigma^2 \in R_0^+, k \in \{5.57, \dots, 15\}, j \in \{1, 2\}$$

The prior distribution of β was a normal distribution with a prior mean β_0 of 2% predicated upon previous results on threshold increases in the alpha range. The prior variance $\sigma_{\beta_0}^2$ was set equal to $\sigma_{\mu_0}^2$. The same prior distribution was adopted in the remaining models as well.

$$\beta \sim \text{Normal}\left(\beta_0, \sigma_{\beta_0}^2\right), \text{ where } \beta_0 = 0.02, \sigma_{\beta_0}^2 \approx 0.125$$

The fourth type of model (models M_{52} to M_{100}) combined a constant baseline group difference and a linear threshold increase at varying starting points between high and low synchronizers.

$$D \sim \text{LogNormal}(\mu_j + x_{k,j}\beta, \sigma^2), \text{ where } j \in \{1, 2\}, \mu_j, \beta, \sigma^2 \in R_0^+, k \in \{5.57, \dots, 15\}$$

Upon suggestion of a reviewer, we included another model type, in which the slope β_j of the threshold increase could vary between high and low synchronizers. In models M_{101} to M_{149} , we tested this possibility again for all possible starting points.

$$D \sim \text{LogNormal}(\mu + x_{k,j}\beta, \sigma^2), \text{ where } j \in \{1, 2\}, \mu, \beta_j, \sigma^2 \in R_0^+, k \in \{5.57, \dots, 15\}$$

The previous prior for the slope parameter was applied for both groups.

$$\beta_j \sim \text{Normal}\left(\beta_0, \sigma_{\beta_0}^2\right), \text{ where } \beta_0 = 0.02, \sigma_{\beta_0}^2 \approx 0.125$$

In total, we compared 149 models, which we assigned equal prior probability, $P(M_i) = \frac{1}{149}$. According to Bayes’ rule, the posterior probability of a model M_i is given by the product of its prior probability and the marginal likelihood of the observations given the model divided by the marginal probability of the observations:

$$p(M_i|D) = \frac{p(M_i)p(D|M_i)}{p(D)}, \text{ where } p(D) = \sum_{j=1}^m p(D|M_j)p(M_j)$$

The Bayes factor (BF) was calculated to determine the amount of evidence in favor of one model as the ratio between its posterior probability and the posterior probability of the remaining models (Kass & Raftery, 1995). Similarly, we obtained the probability for a specific effect by marginalizing over all secondary effects and comparing the resulting posterior probability.

We used numerical methods implemented in the RStan and the bridgesampling package (Gronau et al., 2017) in R to estimate the log posterior probability. We first obtained samples of the posterior distribution (log density function) by running each model in RStan, which uses a No-U-Turn (NUTS) sampling Markov chain Monte Carlo (MCMC) algorithm (Carpenter et al., 2017). We ran five chains (maximal tree depth = 10, target average acceptance probability for adaptation = 0.95), each containing 2,000 iterations, including 1,000 warm-up iterations (stan default settings). Further increasing the number of iterations did not change the parameter estimates and posterior probabilities, indicating convergence. The accuracy of the MCMC algorithm was checked via the diagnostics provided in the RStan environment (rhat statistics, divergences, saturation of the maximum three depth, and the Bayesian fraction of missing information). Finally, we used the RStan output to compute the log marginal likelihood via bridge sampling and, based on that, the posterior probability (Gronau et al., 2017).

Results

Weighted up-down procedure

Median relative difference thresholds in the weighted up-down (WUD) procedure ranged from 4.08% at 4 Hz to 10.5% at 15 Hz (threshold averaged across standard rates in each participant: $Mdn = 5.25%$, $MAD = 2.54%$). High synchronizers displayed lower thresholds averaged across all standard rates ($Mdn = 5.52%$, $MAD = 2.15%$) compared to low synchronizers ($Mdn = 8.21%$, $MAD = 3.7%$) (Fig. 2c; Wilcoxon rank-sum test, one-sided, $W = -2.25$, $p = .012$, $r = -0.21$). We used a Bayesian model comparison to test the influence of stimulation rate and auditory-motor speech synchronization behavior more specifically. The Bayesian sampling algorithm displayed appropriate behavior for all of the 149 models (rhat statistics < 1.001 ; no divergence in any of the fits; no saturations of the maximal tree depth 10; non-suspicious Bayesian fraction of missing information). Bridge sampling therefore allowed for accurate estimation of the posterior probability of each model given the data (see Table S1 in the Online Supplementary Materials). As we used a uniform prior probability for a large number of models ($P(M_i) = \frac{1}{149}$), we did not expect a high BF for a single model. More importantly, the posterior probability of the most likely model (M_{33}) increased considerably from a prior probability of about 0.67% to 28.61% ($BF \approx 0.4$). This model included a constant baseline threshold ($\mu \approx 4.48%$, 95% CI [4.17%, 4.84%]) from 4 to 7.14 Hz in low synchronizers and from 4 to 10.29 Hz in high synchronizers, and a subsequent increase with the same slope in both groups ($\beta \approx 1.44%$, 95% CI [1.13%, 1.75%]) (Fig. 2d). Overall, the model type “Increase” yielded the highest posterior probability marginalized over onset points, $P(M_{3-51}|D) \approx 74.53%$ ($BF \approx 2.93$) (Fig. 2A). In contrast, an additional group difference between high and low synchronizers in the baseline threshold ($P(M_{52-100}|D) \approx 24.5%$, $BF \approx 0.32$) or the slope of the threshold increase ($P(M_{101-149}|D) \approx 0.96%$, $BF \approx 0.01$) was not supported by our data. Crucially, we found considerable evidence for an earlier threshold increase in low synchronizers compared to high synchronizers, when comparing the corresponding models to all the remaining ones, marginalizing over model types ($P(M_{(Onset\ Low < Onset\ High)}|D) \approx 87.97%$, $BF \approx 7.31$) (Fig. 2b). The results suggest that the threshold increase is most likely to start at around 7.14–8.71 Hz in low synchronizers and around 10.29–11.86 Hz in high synchronizers.

Above we reported the measured relative rate discrimination thresholds. In order to relate our findings to oscillatory theories of auditory perception (e.g., Poeppel, 2003; Teng et al., 2017; Teng & Poeppel, 2020), the relative rate difference can be expressed as a relative fraction of the respective oscillatory cycle (in ms) required for reliable rate

discrimination. This can be straightforwardly computed from the relative difference thresholds reported here and yields similar results (Table S2, Online Supplementary Materials).

Constant stimuli procedure

To validate the WUD threshold measure procedure, we examined rate discrimination sensitivity in the constant stimuli (CS) procedure by fitting the psychometric function at 4 Hz and 11.86 Hz (Fig. 1c). Inspection of goodness-of-fit for individual data demonstrated proper fit for all 110 individual fits (55 subjects, two standard rates; see Fig. 3a for mean data in each standard rate condition across all participants). The lapse rate was generally low, with a trend for higher lapse rates at 11.86 Hz ($Mdn = 1.16 \times e^{-07\%}$, $MAD = 2.07%$) compared to the 4-Hz standard rate condition ($Mdn = 6.05 \times e^{-08\%}$, $MAD = 1.3%$) (Wilcoxon signed-rank test, two-sided, $T = -1.93$, $p = .054$). This suggests that participants remained attentive throughout the whole procedure, producing only a few lapses.

Relative thresholds at 4 Hz and 11.86 Hz strongly correlated between the WUD and CS method (average across standard rates: Spearman-rank correlation, high/low synchronizer group-demeaned variables, $r_s = 0.68$, $p < .001$, Fig. 3b; 4 Hz: $r_s = 0.66$, $p < .001$; 11.86 Hz: $r_s = 0.55$, $p < .001$), as well as the difference in thresholds between 4 Hz and 11.86 Hz ($r_s = 0.39$, $p = .003$). Thresholds were overall lower in the CS procedure compared to the WUD procedure at both 4 Hz (Wilcoxon signed-rank test, two-sided, $T = -4.37$, $p < .001$, $r = -0.42$) and 11.86 Hz (Wilcoxon signed-rank test, two-sided, $T = -3.24$, $p = .001$, $r = -0.31$), presumably reflecting increased familiarity with the task. But the difference between relative difference thresholds at 4 and 11.86 Hz (indicating a threshold increase at rates in the alpha range) did not differ between methods (Wilcoxon signed-rank test, two-sided, $T = 0.91$, $p = .361$). Together, these results confirm that temporal sensitivity was measured adequately in the adaptive weighted up-down procedure, which was our main procedure to test temporal sensitivity across the theta and alpha range.

Thresholds in the CS method were also higher at 11.86 Hz ($Mdn = 4.84%$, $MAD = 2.38%$) compared to 4 Hz ($Mdn = 3.13%$, $MAD = 1.55%$) (Fig. 3c; Wilcoxon signed-rank test, one-sided, $T = 4.95$, $p < .001$, $r = 0.47$). This provides further evidence for a decrease in temporal sensitivity from the theta to the alpha range at 11.29 Hz. Average thresholds across standard rate conditions were descriptively lower in high ($Mdn = 3.82%$, $MAD = 1.42%$) versus low synchronizers ($Mdn = 4.87%$, $MAD = 2.05%$), but there was only a trend for statistical significance (Wilcoxon rank-sum test, one-sided, $W = -1.58$, $p = .057$) (Fig. 3d). (Note that in the CS method we only tested two rates, and thus could not access group differences as nuanced as in the WUD threshold measure).

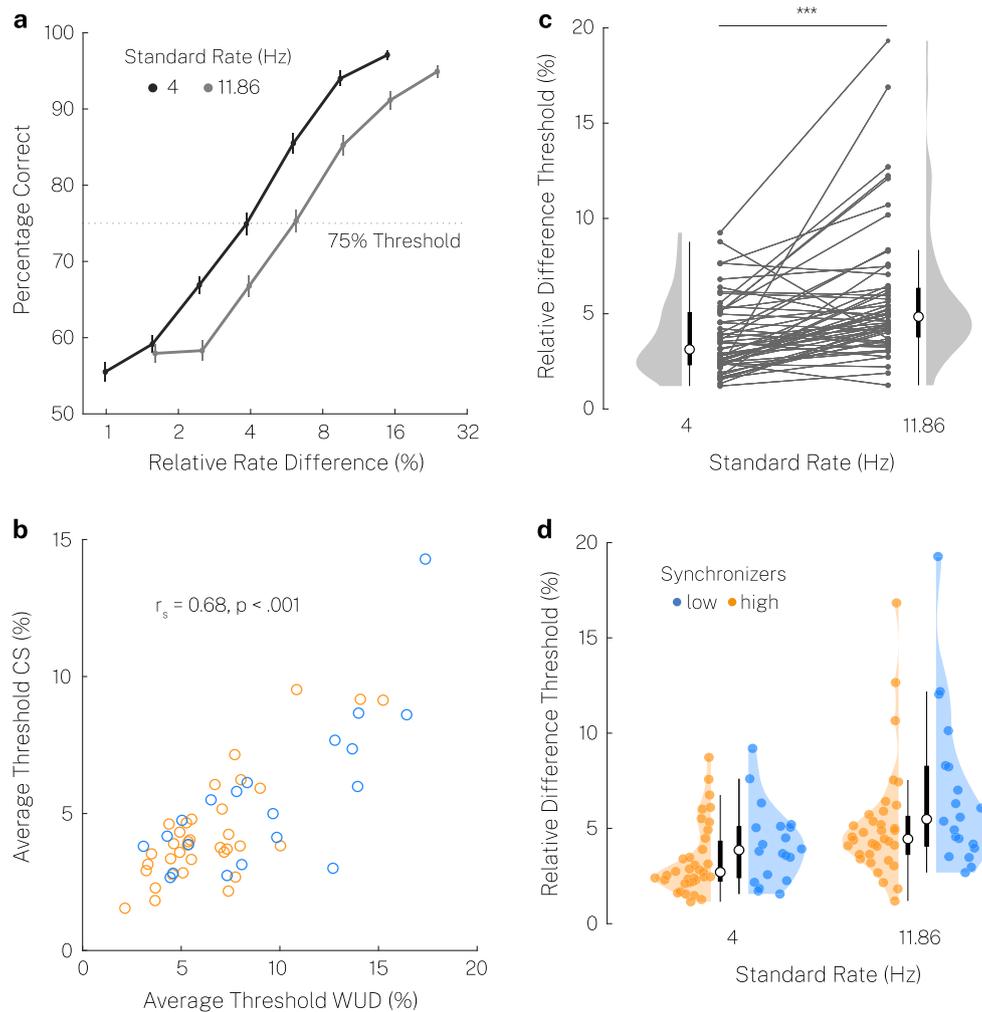


Fig. 3 Constant stimuli method suggests optimal processing range in the theta range. **(a)** Mean psychometric functions across all participants at two standard rates and seven comparison levels. Error bars represent the standard error of the mean. **(b)** Scatter plots of average thresholds across standard rates in the weighted up-down (WUD) and constant stimuli (CS) procedure. Correlation calculated on high/low group-demeaned variables.

Results from the CS procedure expressed as the relative fraction of the oscillatory cycle (in ms) are reported in the Online Supplementary Materials (Table S3).

Correlation of temporal sensitivity, auditory-motor synchronization, and musicality

Our previous analyses suggest a close relationship between auditory-motor speech synchronization behavior and auditory temporal sensitivity. Here, we analyzed the impact of general musical sophistication (Gold-MSI) on these measures. High synchronizers reported higher musicality compared to low synchronizers (Wilcoxon rank-sum test, one-sided, $W = 3.27, p < .001, r = 0.3$). Given the bimodal distribution of auditory-motor speech synchronization behavior, we demeaned all variables in the high and low synchronizer clusters

(c) Relative difference thresholds at two standard rates across all participants. Grey dot and horizontal line: individual participant, white dot: median, thick line: quartiles, thin line: quartiles $\pm 1.5 \times$ interquartile range. Thresholds were significantly lower at rates of 4 Hz compared to 11.86 Hz. *** $p < .001$. **(d)** Thresholds for high and low synchronizers. Colored dot: individual participant

to control for spurious correlations when calculating correlations among the variables (Fig. 4). Musical sophistication correlated positively with mean PLVs in the SSS-test ($r_s = 0.46, p < .001$), indicating that higher musicality went along with improved auditory-motor synchronization behavior. There was only a trend for a negative correlation between musical sophistication and average relative threshold levels across standard rates in the CS procedure ($r_s = -0.26, p = .056$) and the WUD procedure ($r_s = -0.23, p = .092$). Crucially, threshold levels in the WUD method correlated moderately with mean PLVs in the SSS-test ($r_s = -0.36, p = .008$), even when controlled for musical sophistication (partial correlation $r_s = -0.29, p = .034$). This suggests that differences in auditory-motor synchronization behavior explained individual variance in temporal sensitivity beyond musical sophistication. However, the correlation between average threshold levels

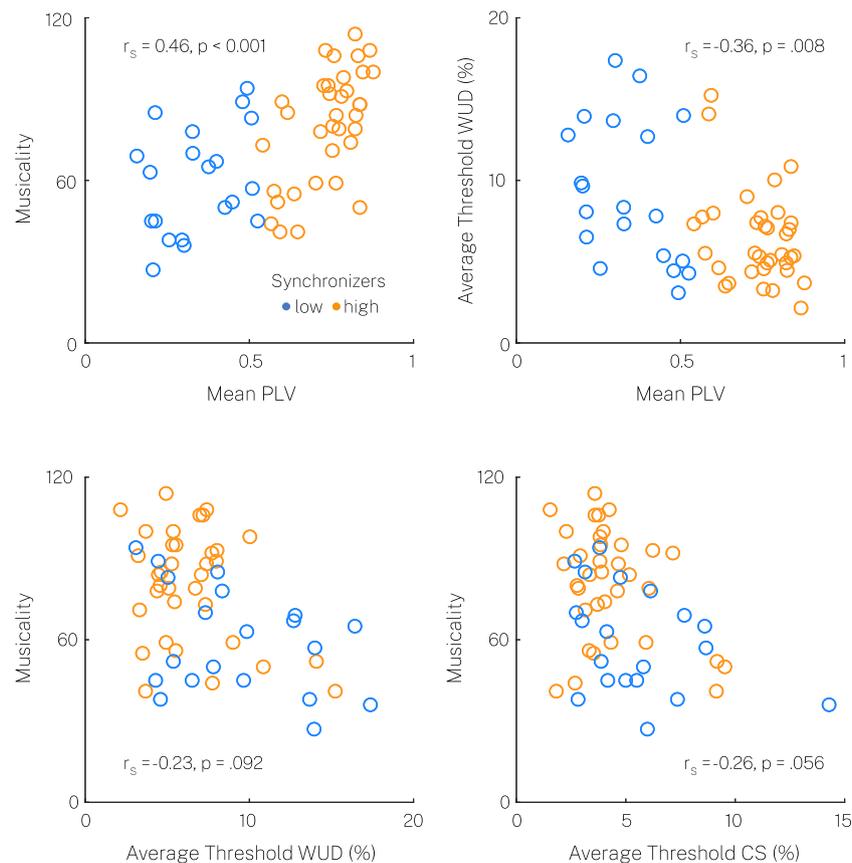


Fig. 4 Scatter plots and correlations between rate-discrimination threshold, auditory-motor coupling, and musicality measures. Correlations were calculated based on high/low group-demeaned variables to account for spurious correlations

in the CS procedure and mean PLVs did not reach statistical significance ($r_s = -0.21, p = .13$).

Discussion

The experiments we present here lie at the intersection of two fundamental questions in perception that have been investigated by and large independently. On the one hand, rate sensitivity is a basic question in auditory perception. On the other hand, the interaction between perception and action systems has been increasingly studied, with new insights about auditory-motor coupling. We combine these two lines of research in psychophysical experiments and discover a surprising generalization: listeners with increased synchronization of speech perception and production have broader sensitivity to acoustic rates. Our evidence invites two conclusions: first, there are indeed preferred, fine-grained regimes of temporal processing in hearing at low-frequency modulation rates in the theta range. Second, as the speech synchronization test has been shown to estimate coupling between auditory and speech motor cortices (Assaneo, Ripollés, et al., 2019), our findings suggest that the motor system influences perceptual

timing; in particular, individual variability in auditory-motor coupling seems to undergird listeners' temporal processing thresholds, with a larger range of optimal processing in high compared to low auditory-motor synchronizers.

Overall, individuals showed constant low relative rate discrimination thresholds in the theta range. The range of optimal processing varied for high and low synchronizers, with the most likely optimal range for low synchronizers from 4 to 7.14 Hz and high synchronizers from 4 to 10.29 Hz. Thresholds increased at frequencies above the optimal range up to 15 Hz, indicating reduced temporal sensitivity in the alpha range (and above). Some previous studies report reduced temporal sensitivity for rate discrimination at higher rates in the alpha range (Drake & Botte, 1993; Friberg & Sundberg, 1995; McAuley & Kidd, 1998; ten Hoopen et al., 1994; ten Hoopen et al., 2011), whereas others did not (Dau et al., 1997; Elliott & Theunissen, 2009; Sheft & Yost, 1990; Viemeister, 1979). Compared to previous reports, we tested rate discrimination within a fine-grained range of rates in a larger sample of participants. Furthermore, we applied an additional procedure, the constant stimuli method, to confirm that the thresholds were reliably estimated by the adaptive WUD procedure. Our data support previous behavioral findings, and neurophysiological evidence, suggesting that

neuronal oscillatory activity in auditory cortex shows optimal temporal processing in the theta range (Teng et al., 2017; Teng & Poeppel, 2020). The constant relative threshold in the theta range corresponds to a linear increase in absolute rate difference required for discrimination, congruent with Weber's law. In contrast, Weber's law did not apply for the observed increase of relative thresholds at higher rates, suggesting a decreased temporal resolution at higher rates. In other words, this finding indicates that at higher rates the absolute rate difference required for discrimination failed to scale with the rate, that is, the temporal resolution at higher rates seems to be constrained by a minimum "time constant," which differed for high and low synchronizers (at ~7–10% relative difference threshold, corresponding to ~6–8 ms). On a neuronal level, the constant relative thresholds might reflect the preferred frequencies of the auditory cortex neuronal population (Teng et al., 2017; Teng & Poeppel, 2020). More specifically, this leads to the prediction that the relative difference threshold is constant - reflecting a constant fraction of the oscillatory cycle (Table S2, S3, Online Supplementary Materials) - only for optimal rates that can be tracked reliably.

Interestingly, in addition to the theta range, an optimal auditory processing range in the gamma band has been reported using amplitude-modulated sound (Galambos, 1992; Teng et al., 2017; Teng & Poeppel, 2020; see also: Poeppel, 2003). In our paradigm - which uses pure tone sequences - investigating rate perception in the gamma range is complicated, as rate perception would cross over to (roughness and) pitch perception in the gamma range (Oxenham, 2012). Further research, with a different experimental paradigm, is required to investigate whether this second auditory processing regime in the gamma band (Giraud, 2020; Hoonhorst et al., 2009; Joliot et al., 1994; Teng et al., 2017; Teng & Poeppel, 2020) is affected by auditory-motor coupling.

Crucially, auditory temporal sensitivity was modulated by interindividual differences in the auditory-motor speech synchronization behavior, which is indicative of the auditory-motor cortex coupling strength (Assaneo et al., 2020; Assaneo, Ripollés, et al., 2019). High compared to low synchronizers showed lower relative difference thresholds at higher rates, supporting a model that suggests a larger range of optimal temporal resolution in the high synchronizers. A possibility is that our findings reflect top-down temporal predictions from the motor system facilitating auditory processing. Previous research suggests that top-down predictions from the motor system can modulate auditory processing even during passive listening (Chen et al., 2008; Grahn & Rowe, 2013). We utilized the previously introduced behavioral spontaneous speech synchronization test (SSS-test) to distinguish individuals based on their spontaneous auditory-motor synchronization behavior (Assaneo, Ripollés, et al., 2019). The original study reported the test-retest reliability and replicability of the bimodal distribution of the synchronization

behavior. Crucially, differences in functional and structural frontal-motor to auditory cortex connectivity have been shown to underly this group distinction, rendering it a suitable estimator of auditory-motor coupling strength. (Note that although the SSS-test measures auditory-motor synchronization at ~4.5 Hz, the relation to structural connectivity suggest that more general rate-independent auditory-motor interactions are reflected by the test outcome, see also: Assaneo et al., 2020.) Furthermore, the auditory-motor synchronization behavior most likely reflects an oscillatory mechanism, whereas high synchronizers have shown stronger perceptual facilitation likely related to top-down predictions from the motor system (Assaneo et al., 2020). Based on these previous findings, we propose that the increased rate discrimination sensitivity in high synchronizers reflects a facilitating effect of increased motor cortex recruitment in high compared to low synchronizers. Motor top-down effects particularly show a facilitating effect during demanding listening situations (Stokes et al., 2019; Wu et al., 2014; e.g. such as at "non-optimal" fast stimulation rates). Previously it has been proposed that motor top-down effects can affect processing in demanding listening situations by reducing the hemispheric right lateralization resulting in bilateral auditory cortex recruitment (Assaneo, Rimmele, et al., 2019). However, whether such a neuronal mechanism can account for the observed larger optimal range for rate discrimination in high synchronizers is unknown and requires further research.

An alternative explanation for our findings is that the differences in auditory temporal sensitivity in high and low synchronizers was due to population differences in auditory processing rather than differences in auditory-motor coupling. We cannot entirely exclude this explanation. However, previous evidence shows that the high/low synchronizer distinction comes with differences in functional and structural cortical auditory-motor interactions, rendering this "purely auditory processing" explanation unlikely. Furthermore, musical expertise might affect the temporal sensitivity for rate discrimination. For example, there is evidence for enhanced auditory-motor integration in musicians benefitting auditory processing (Du & Zatorre, 2017). Accordingly, we found that high and low synchronizers differed with respect to musical sophistication. However, the auditory-motor coupling accessed with the SSS-test (Assaneo, Ripollés, et al., 2019) correlated with interindividual differences in temporal sensitivity beyond musicality. Our findings suggest that the SSS-test provides a more specific estimate of auditory-motor synchronization compared to the measure of musical sophistication, which is likely a more indirect estimate of auditory-motor coupling.

In summary, we report a constant auditory temporal sensitivity for rate discrimination in the theta range, with a decrease in sensitivity for higher rates. Individuals with strong compared to weak synchronization of speech perception and

production showed higher rate-discrimination sensitivity particularly at faster rates, indicating a larger range of optimal temporal processing. Our behavioral findings suggest constraints of auditory perception consistent with oscillatory theories, which propose neuronal populations in auditory cortex with preferred frequencies in the theta range. Furthermore, our data suggest a crucial role of auditory-motor coupling for enhancing temporal sensitivity by improving temporal resolution at higher stimulation rates.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-021-01933-w>.

Acknowledgements/Funding Open Access funding enabled and organized by Projekt DEAL. This work was funded by the Max-Planck-Institute for Empirical Aesthetics and supported by CLaME Max Planck NYU Center for Language Music and Emotion. The authors have no relevant financial or non-financial interests to disclose.

Authors' contributions P. Kern, J. M. Rimmele, and D. Poeppel designed the research. P. Kern collected the experimental data. P. Kern, M. F. Assaneo, D. Endres, and J. M. Rimmele analyzed the data. P. Kern wrote the initial draft of the manuscript. P. Kern, M. F. Assaneo, D. Endres, D. Poeppel, and J. M. Rimmele discussed the results and contributed to the final manuscript.

Declarations

Ethics approval The experimental procedures were ethically approved by the Ethics Council of the Max Planck Society (No. 2017_12).

Consent to participate All participants gave written informed consent prior to the study.

Consent for publication All participants gave written informed consent for publication.

Preregistration The study was not formally preregistered.

Data availability Anonymized data (except for the speech audio recordings) are publicly available on the Open Science Framework (OSF).

Code availability All data analysis scripts are publicly available on OSF.

Conflicts of interest The authors declare no conflicts of interest/competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372. <https://doi.org/10.1073/pnas.201400998>
- Amal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cerebral Cortex*, *25*(9), 3077–3085. <https://doi.org/10.1093/cercor/bhu103>
- Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*.
- Assaneo, M. F., Rimmele, J. M., Orpella, J., Ripollés, P., de Diego-Balaguier, R., & Poeppel, D. (2019). The Lateralization of Speech-Brain Coupling Is Differentially Modulated by Intrinsic Auditory and Top-Down Mechanisms. *Frontiers in Integrative Neuroscience*, *13*. <https://doi.org/10.3389/fnint.2019.00028>
- Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. *Nature Human Behaviour*, *5*, 71–82. <https://doi.org/10.1038/s41562-020-00962-0>
- Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguier, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*(4), 627–632. <https://doi.org/10.1038/s41593-019-0353-z>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Brungart, D., Wassenhove, V. van, Brandewie, E., & Romigh, G. D. (2007). The effects of temporal acceleration and deceleration on AV speech perception. *AVSP*.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), 1–32. <https://doi.org/10.18637/jss.v076.i01>
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex*, *18*(12), 2844–2854. <https://doi.org/10.1093/cercor/bhn042>
- Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America*, *102*(5), 2906–2919. <https://doi.org/10.1121/1.420345>
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience*

- and *Biobehavioral Reviews*, 81(Pt B), 181–187. <https://doi.org/10.1016/j.neubiorev.2017.02.011>
- Doelling, K. B., Amal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, 761–768. <https://doi.org/10.1016/j.neuroimage.2013.06.035>
- Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54(3), 277–286. <https://doi.org/10.3758/BF03205262>
- Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, 114(51), 13579–13584. <https://doi.org/10.1073/pnas.1712223114>
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, 3, 1393–1396. <https://doi.org/10.1109/ICSLP.1996.607874>
- Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, 58(1), 133–147. <https://doi.org/10.1016/j.bandc.2004.09.014>
- Elliott, T. M., & Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLOS Computational Biology*, 5(3), e1000302. <https://doi.org/10.1371/journal.pcbi.1000302>
- Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *The Journal of the Acoustical Society of America*, 98(5), 2524–2531. <https://doi.org/10.1121/1.413218>
- Galambos, R. (1992). A Comparison of Certain Gamma Band (40-HZ) Brain Rhythms in Cat and Man. In E. Başar & T. H. Bullock (Eds.), *Induced Rhythms in the Brain* (pp. 201–216). Birkhäuser. https://doi.org/10.1007/978-1-4757-1281-0_11
- Ghitza, O. (2012). On the Role of Theta-Driven Syllabic Parsing in Decoding Speech: Intelligibility of Speech with a Manipulated Modulation Spectrum. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00238>
- 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(1–2), 113–126. <https://doi.org/10.1159/000208934>
- Giraud, A.-L. (2020). Oscillations for all A commentary on Meyer, Sun & Martin. *Language, Cognition and Neuroscience*, 0(0), 1–8. <https://doi.org/10.1080/23273798.2020.1764990>
- Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron*, 56(6), 1127–1134. <https://doi.org/10.1016/j.neuron.2007.09.038>
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. <https://doi.org/10.1038/nn.3063>
- Grahn, J. A., & Rowe, J. B. (2013). Finding and Feeling the Musical Beat: Striatal Dissociations between Detection and Prediction of Regularity. *Cerebral Cortex*, 23(4), 913–921. <https://doi.org/10.1093/cercor/bhs083>
- Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *The Journal of the Acoustical Society of America*, 87(6), 2662–2674. <https://doi.org/10.1121/1.399058>
- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingrover, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81, 80–97. <https://doi.org/10.1016/j.jmp.2017.09.005>
- 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 Biology*, 11(12), e1001752. <https://doi.org/10.1371/journal.pbio.1001752>
- Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, 86, 150–165. <https://doi.org/10.1016/j.neubiorev.2017.12.002>
- Herbst, S. K., & Obleser, J. (2019). Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. *NeuroImage*, 203, 116198. <https://doi.org/10.1016/j.neuroimage.2019.116198>
- Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., & Deltenre, P. (2009). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in French-speaking subjects. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 120(5), 897–903. <https://doi.org/10.1016/j.clinph.2009.02.174>
- Joliot, M., Ribary, U., & Llinás, R. (1994). Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), 11748–11751. <https://doi.org/10.1073/pnas.91.24.11748>
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227–229. <https://doi.org/10.3758/BF03214307>
- Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Keitel, A., & Gross, J. (2016). Individual Human Brain Areas Can Be Identified from Their Characteristic Spectral Activation Fingerprints. *PLOS Biology*, 14(6), e1002498. <https://doi.org/10.1371/journal.pbio.1002498>
- Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, 16(3), e2004473. <https://doi.org/10.1371/journal.pbio.2004473>
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction* (pp. xv, 279). Elsevier Academic Press.
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:4<194::AID-HBM4>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C)
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology*, 94(3), 1904–1911. <https://doi.org/10.1152/jn.00263.2005>
- Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Lubinus, C., Orpella, J., Keitel, A., Gudi-Mindermann, H., Engel, A. K., Roeder, B., & Rimmele, J. M. (2021). Data-driven classification of spectral profiles reveals brain region-specific plasticity in blindness. *Cerebral Cortex*, 31(5), 2505–2522. <https://doi.org/10.1093/cercor/bhaa370>
- MATLAB. (2018). *Version 9.5 (R2018b)*. Natick, Massachusetts: The MathWorks Inc.
- McAuley, J. D., & Jones, M. R. (2003). Modeling Effects of Rhythmic Context on Perceived Duration: A Comparison of

- Interval and Entrainment Approaches to Short-Interval Timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102–1125. <https://doi.org/10.1037/0096-1523.29.6.1102>
- McAuley, J. D., & Kidd, G. R. (1998). Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1786–1800. <https://doi.org/10.1037/0096-1523.24.6.1786>
- Michon, J. A. (1964). Temporal Structure of Letter Groups and Span of Perception. *Quarterly Journal of Experimental Psychology*, 16(3), 232–240. <https://doi.org/10.1080/17470216408416373>
- Miller, N. S., & McAuley, J. D. (2005). Tempo sensitivity in isochronous tone sequences: The multiple-look model revisited. *Perception & Psychophysics*, 67(7), 1150–1160. <https://doi.org/10.3758/BF03193548>
- Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences*, 114(42), E8913–E8921. <https://doi.org/10.1073/pnas.1705373114>
- Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, 5(1), 1–9. <https://doi.org/10.1038/ncomms6255>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nordmark, J. O. (1968). Mechanisms of Frequency Discrimination. *The Journal of the Acoustical Society of America*, 44(6), 1533–1540. <https://doi.org/10.1121/1.1911293>
- Oxenham, A. J. (2012). Pitch Perception. *Journal of Neuroscience*, 32(39), 13335–13338. <https://doi.org/10.1523/JNEUROSCI.3815-12.2012>
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, 25(12), 1649–1653. <https://doi.org/10.1016/j.cub.2015.04.049>
- Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3), 539–558. JSTOR.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Poepfel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Rammsayer, T. H. (1992). An experimental comparison of the weighted up-down method and the transformed up-down method. *Bulletin of the Psychonomic Society*, 30(5), 425–427. <https://doi.org/10.3758/BF03334107>
- Ravignani, A., Bella, S. D., Falk, S., Kello, C. T., Noriega, F., & Kotz, S. A. (2019). Rhythm in speech and animal vocalizations: A cross-species perspective. *Annals of the New York Academy of Sciences*, 1453(1), 79–98. <https://doi.org/10.1111/nyas.14166>
- Rimmele, J. M., Gross, J., Molholm, S., & Keitel, A. (2018). Editorial: Brain oscillations in human communication. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00039>
- Rimmele, J. M., Morillon, B., Poepfel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, 22(10), 870–882. <https://doi.org/10.1016/j.tics.2018.08.003>
- Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, 122, 105–123. <https://doi.org/10.1016/j.visres.2016.02.002>
- Sheft, S., & Yost, W. A. (1990). Temporal integration in amplitude modulation detection. *The Journal of the Acoustical Society of America*, 88(2), 796–805. <https://doi.org/10.1121/1.399729>
- Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6), 3394–3411. <https://doi.org/10.1121/1.1624067>
- Stokes, R. C., Venezia, J. H., & Hickok, G. (2019). The motor system’s [modest] contribution to speech perception. *Psychonomic Bulletin & Review*, 26(4), 1354–1366. <https://doi.org/10.3758/s13423-019-01580-2>
- Tang, D., Möttönen, R., Asaridou, S. S., & Watkins, K. E. (2020). Asymmetry of auditory-motor speech processing is determined by language experience. *BioRxiv*, 2020.06.05.137067. <https://doi.org/10.1101/2020.06.05.137067>
- ten Hoopen, G., Boelaerts, L., Gruisen, A., Apon, I., Donders, K., Mul, N., & Akerboom, S. (1994). The detection of anisochrony in monaural and interaural sound sequences. *Perception & Psychophysics*, 56(1), 110–120. <https://doi.org/10.3758/BF03211694>
- ten Hoopen, G., Van Den Berg, S., Memelink, J., Bocanegra, B., & Boon, R. (2011). Multiple-look effects on temporal discrimination within sound sequences. *Attention, Perception, & Psychophysics*, 73(7), 2249–2269. <https://doi.org/10.3758/s13414-011-0171-1>
- Teng, X., & Poepfel, D. (2020). Theta and Gamma Bands Encode Acoustic Dynamics over Wide-Ranging Timescales. *Cerebral Cortex*, 30(4), 2600–2614. <https://doi.org/10.1093/cercor/bhz263>
- Teng, X., Tian, X., Rowland, J., & Poepfel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLOS Biology*, 15(11), e2000812. <https://doi.org/10.1371/journal.pbio.2000812>
- Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, 66(5), 1364–1380. <https://doi.org/10.1121/1.383531>
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313. <https://doi.org/10.3758/BF03194544>
- Wu, Z.-M., Chen, M.-L., Wu, X.-H., & Li, L. (2014). Interaction between auditory and motor systems in speech perception. *Neuroscience Bulletin*, 30(3), 490–496. <https://doi.org/10.1007/s12264-013-1428-6>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.