Neural Entrainment Determines the Words We Hear

Anne Kösem, Hans Rutger Bosker, Atsuko Takashima, Antje Meyer, Ole Jensen, Peter Hagoort

Neural entrainment determines the words we hear

Anne Kösem[1-3*], Hans Rutger Bosker[1,2], Atsuko Takashima[1,2], Antje Meyer[1,2], Ole Jensen[2,4], Peter Hagoort[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Radboud University, Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands
[3]Lyon Neuroscience Research Center (CRNL), Brain Dynamics and Cognition Team, INSERM U1028, CNRS UMR5292, Université Claude Bernard Lyon 1, Lyon, France
[4]University of Birmingham, Centre for Human Brain Health, Birmingham, United Kingdom

*Corresponding author, lead contact: kosem.anne@gmail.com

SUMMARY

Low-frequency neural entrainment to rhythmic input has been hypothesized as a canonical mechanism that shapes sensory perception in time. Neural entrainment is deemed particularly relevant for speech analysis, as it would contribute to the extraction of discrete linguistic elements from continuous acoustic signals. Yet, its causal influence in speech perception has been difficult to establish. Here, we provide evidence that oscillations build temporal predictions about the duration of speech tokens that affect perception. Using magnetoencephalography (MEG), we studied neural dynamics during listening to sentences that changed in speech rate. We observed neural entrainment to preceding speech rhythms persisting for several cycles after the change in rate. The sustained entrainment was associated with changes in the perceived duration of the last word's vowel, resulting in the perception of words with different meanings. These findings support oscillatory models of speech processing, suggesting that neural oscillations actively shape speech perception.

INTRODUCTION

Brain activity is known to follow the rhythmic structure of sensory signals, and this for various stimulation ranges and sensory modalities [1,2]. Yet, it is still unclear whether the observed oscillatory activity in electrophysiological recordings truly reflects the recruitment of endogenous neural oscillations that are entrained to the stimulations rhythms, and whether these oscillations causally influence sensory processing and perception [1]. Neural entrainment that relies on the recruitment of endogenous oscillations should be dynamic and self-sustained, meaning that it should adapt to the dynamics of current sensory rhythms and should persist for several cycles after stimulation. Crucially, the sustained neural entrainment would be functionally relevant for sensory processing as it would provide a temporal predictive mechanism [2,3]: neural entrainment would reflect the internalization of past sensory rhythms to optimize sensory processing by predicting the timing of future sensory events. So far evidence for sustained entrainment is scarce, and has only been reported in occipital cortices for visual alpha oscillations, and in temporal cortices after auditory entrainment in monkey recordings [4,5]. A crucial open question is whether sustained entrainment occurs during the presentation of complex ecological signals such as speech, and, if so, how it would impact perception [6,7].

Neural entrainment could provide important temporal information for speech processing, given that the acoustic signal presents periodicities of the same temporal granularity as relevant linguistic units, e.g. syllables [6,7]. Specifically, low-frequency neural entrainment has been proposed to contribute to parsing, and to defining the duration of discrete speech information extracted from the continuous auditory input [8–10]. Being recruited at the earliest stages of speech analysis, entrained oscillations should ultimately influence the perception of the spoken utterances. As for other entrainment schemes, their causal efficacy in speech processing remains debated [11–14]. Because neural oscillations match the dynamics of speech during entrainment, it is unclear whether oscillatory activity observed in electrophysiological recordings during speech processing reflects the involvement of neural oscillators for speech analysis, or, alternatively, is the consequence of non-oscillatory based mechanisms that modulate the evoked response to the rhythmic speech signal [13,15,16]. For instance, stronger neural entrainment has repeatedly been observed for more intelligible speech signals [17–20], but these observations could either originate from the stronger recruitment of oscillatory mechanisms, or from the enhanced evoked response to the speech acoustic features.

To demonstrate the causal role of neural entrainment in speech perception, the oscillatory activity has to be disentangled from the driving stimulus's dynamics. Neural oscillatory models suggest that this dissociation is possible when speech temporal characteristics are suddenly changing. Sustained entrainment to the preceding speech dynamics should be observed after a change in speech rate, meaning that the observed neural entrainment to speech is dependent on contextual rhythmic information. If neural oscillations causally influence speech processing, different neural oscillatory dynamics should lead to different percepts for the same speech material. This predicts that entrainment to past speech rhythms should influence subsequent perception. In line with this proposal, contextual speech rate has been shown to affect the detection of subsequent words [21], word segmentation boundaries [22], and perceived constituent durations [23–25]. We propose that these effects could originate from the presence of sustained neural oscillatory activity that defines the parsing window of linguistic segments from continuous speech [8,13,23]. The frequency of sustained entrainment should then affect the onset, offset and size of the discretized items, so that a change in frequency leads to distinct percepts of the extracted linguistic units.

We tested this hypothesis in an MEG study in which native Dutch participants listened to Dutch sentences with varying speech rates (see Audio S1 and S2 for exemplars). The beginning of the sentence (carrier window) was either compressed or expanded in duration, leading to a fast or a slow speech rate (Figure 1A). Specifically, during the carrier window, the speech envelopes in the slow and fast rate conditions had a strong rhythmic component at 3 Hz and 5.5 Hz respectively (Figure 1B). The final three words (target window) were consistently presented at the original recorded speech rate (Figure 1C). Participants were asked to report their perception of the last word of the sentence (target word), which contained a vowel ambiguous between a short /ɑ/ and a long /a:/, and could be perceived as two distinct Dutch words (e.g., *tak* /tɑk/ "branch" or *taak* /ta:k/ "task"). We investigated whether sustained neural entrainment to speech could be visible after a speech rate change (during the target window), and if the sustained entrainment affected the perception of the target word.

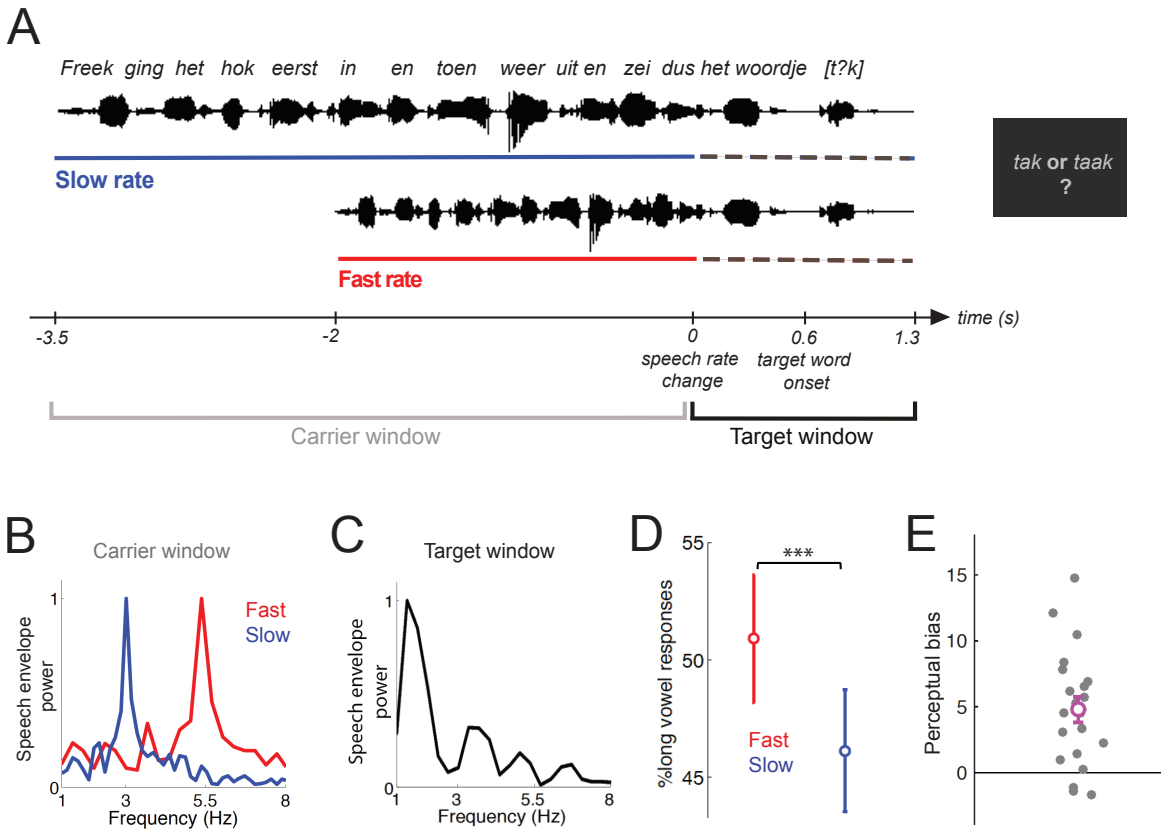

Figure 1: Experimental design and behavioral results. A) The participants listened to Dutch sentences with two distinct speech rates. The beginning of the sentence (carrier window) was either presented at a slow (blue) or fast (red) speech rate (example: "Freek ging het hok eerst in en toen weer uit en zei dus…", meaning *Freek first went in to the shack and then out again and said*...). The last three words ("het woord [target]", meaning *the word [target]*, target window) were spoken at the same pace between conditions. Participants were asked to report their perception of the last word of the sentence (target). The words presented in the carrier window did not contain semantic information that could bias target perception and did not contain any /ɑ/ or /a:/ vowels. B) Speech envelope power spectra in the Carrier window (average across all carrier sentences). The speech envelopes showed a strong oscillatory component at 3 Hz for the Slow (blue) condition, and at 5.5 Hz for the Fast (red) speech rate condition (the two rates correspond to the syllabic presentation rate of the stimuli). For visualization, the speech envelope power spectra have been normalized by dividing the power spectra by their maximum power value. C) Speech envelope power spectra in the Target window (averaged across all sentence endings). 3 Hz and 5.5 Hz oscillatory components were not prominently observed in the power spectra during the Target window. D) Proportion of long vowel percepts in the Fast (red) and Slow (blue) speech rate conditions. Error bars represent s.e.m. The perception of the target word was influenced by the carrier window speech rate: more long vowel percepts were reported when the word was preceded by a fast speech rate. E) Perceptual Bias. We defined the perceptual bias as the difference in percentage long vowel reports between the Fast and Slow speech rate conditions. Each grey dot corresponds to one participant. The magenta dot corresponds to the average perceptual bias across participants. Error bars represent s.e.m. The horizontal display of the dots is dispersed to prevent too much overlap. See also Audio S1 and Audio S2.

RESULTS

Speech perception is influenced by contextual speech rate

Target words always contained an ambiguous vowel that could either be categorized as a short /ɑ/ or as a long /a:/ vowel. Note that the two vowels are distinguishable by both temporal (duration) and spectral characteristics (e.g. second formant frequency; F2) in Dutch [24]. In the design, vowels were kept at a constant duration, but were presented at three distinct F2 frequencies (one ambiguous F2 value, one F2 value biasing participant reports towards short /ɑ/ responses, one F2 value biasing participant reports towards long /a:/ responses). The F2 was varied to control for the participant's engagement in the task, and as expected participants relied on this acoustic cue to discriminate the two vowels (main effect of F2: $F(2,40) = 124.5$, $p < 0.001$, partial Eta squared $\eta p^2 = 0.86$). Crucially, the preceding speech rate in the carrier window affected the perception of the target word (main effect of speech rate: $F(1,20) = 24.4$, $p < 0.001$, $\eta p^2 = 0.55$). Participants were more biased to perceiving the word with a long /a:/ vowel (e.g., *taak*) after a fast speech rate, and the word with a short /ɑ/ vowel (e.g., *tak*) after a slow speech rate (Figure 1D). We quantified how strongly each participant was affected by the preceding speech rate in his/her behavioral report with the Perceptual Bias, which corresponds to the difference in the percentage of long /a:/ vowel reports between the Fast and Slow rate conditions (Figure 1E). The behavioral effect of contextual speech rate was not significantly different across the various F2s tested (interaction F2 by speech rate: $F(2,40) = 0.6$, $p = 0.58$, $\eta p^2 = 0.03$), as previously observed [24]. It suggests that the effect of context is not significantly different when F2 cues bias perception towards either a short vowel or a long vowel percept. We thus pooled the data across F2 conditions for the following MEG analyses.

Neural oscillations follow speech envelope dynamics during the carrier window

The MEG analysis was performed at two distinct time windows: the carrier window (sentence presentation up to the change in speech rate), and the target window (sentence endings after the change in speech rate). During the carrier window, the speech envelopes in the Slow and Fast rate conditions had a strong oscillatory component at 3 Hz and 5.5 Hz respectively. Therefore, neural oscillatory responses were expected to peak at 3 Hz for the Slow rate condition and at 5.5 Hz for the Fast rate condition. To test this, we introduced the Entrainment Index (EI, see Materials and Methods). EI is based on the ratio of total neural oscillatory power at the 3 Hz and at 5.5 Hz between Fast and Slow conditions. EI is larger than 1 when neural oscillations follow the initial speech rate for both Fast and Slow conditions (i.e., stronger 3 Hz power for Slow condition and stronger 5.5 Hz power for Fast condition). Significant neural oscillatory response to the speech rate was observed during the carrier window ($EI > 1$, $p < 0.05$ cluster-based correction), demonstrating that low-frequency brain activity efficiently tracked the dynamics of speech (Figure 2A). Strong EI was observed for all participants (Figure 2B), and effectively captured the oscillatory response to the speech rate in both conditions: the 3 Hz power was relatively stronger in the Slow rate condition than in the Fast rate condition, and 5.5 Hz power was stronger in the Fast rate condition (Figure 2C-D). In this window, the EI may primarily capture the evoked response to speech. In line with this assumption, the strongest EI was most prominently observed in auditory cortices (Figure 2A, Figure S1A).

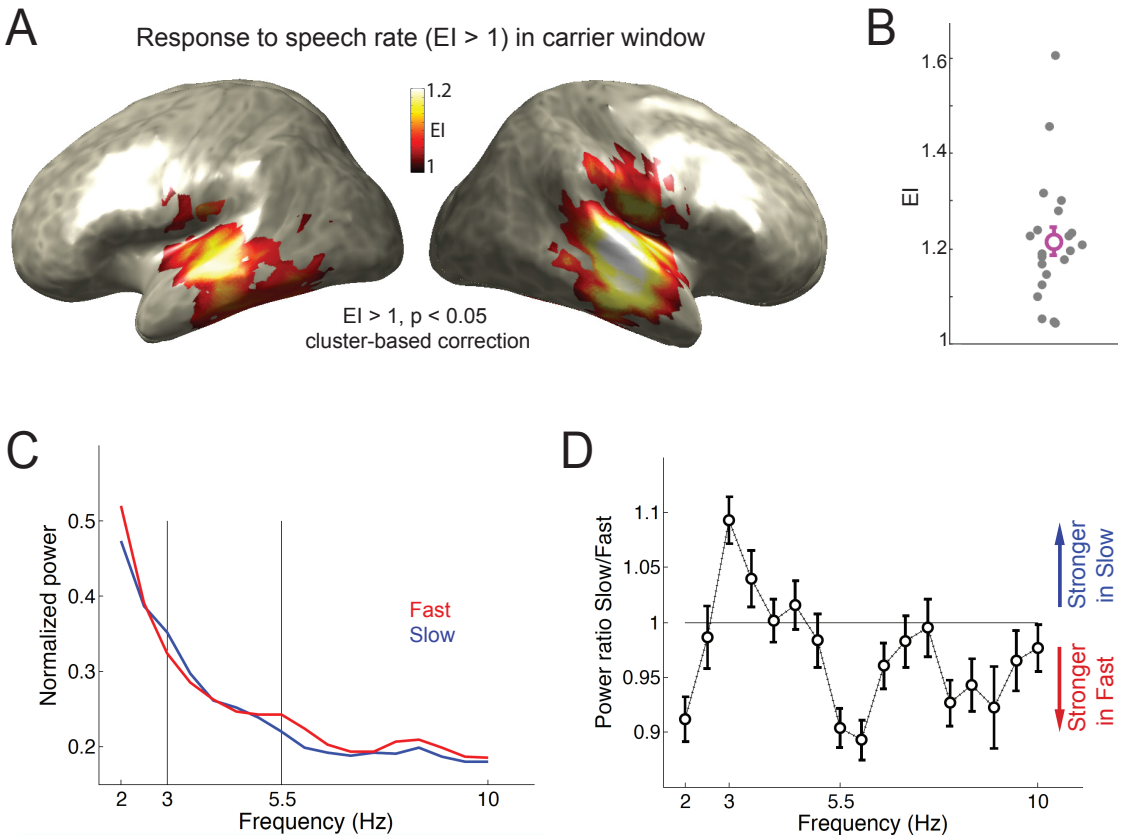

Figure 2: Neural oscillatory response to speech rate during carrier window. A) During the carrier window, neural dynamics in auditory areas follow the current speech rate (i.e. EI > 1). The EI values are thresholded at p<0.05, controlled for multiple comparisons using cluster-based permutation tests. B) Entrainment Index within the most strongly activated grid point (MNI coordinates: 60, -20, -10, Right Superior Temporal Cortex). Each grey dot corresponds to one participant. The magenta circle corresponds to the average EI across participants. Error bars represent s.e.m. The horizontal display of the dots is dispersed to prevent too much overlap. C) Power spectra of neural activity for the Fast (red) and Slow (blue) speech rate conditions within the most strongly activated grid point. D) Contrast in power between the two speech rate conditions (using the power ratio Slow/Fast) within the most strongly activated grid point. Ratios higher than 1 reflect stronger power in Slow speech rate condition, ratios lower than 1 reflect stronger power in Fast speech rate condition. Error bars denote s.e.m. See also Figure S1.

Neural entrainment to past speech dynamics persists after the change in speech rate and affects comprehension

EI was also significantly larger than 1 during the target window ($p < 0.05$ cluster-based correction), in which the speech acoustics were identical across Fast and Slow rate conditions (Figure 3A-B, Figure S1B). Larger EI (>1) reflected a stronger oscillatory response that corresponded in frequency to the preceding speech rate (3 Hz power in the Slow rate condition and 5.5 Hz power in the Fast rate condition, Figure 3C-D) even though the speech signals did not contain a pronounced 3 or 5.5 Hz component (Figure 1C), suggesting that neural entrainment to the preceding speech rhythm persisted. Sustained entrainment was most prominently observed along the right superior temporal and inferior temporal sulci, with the significant cluster extending to the right infero-frontal areas (Figure 3A). No significant sustained entrainment was observed in the left hemisphere during the Target Window (Figure 3A, Figure S1B). Primary auditory cortices (where stimulus-driven responses are largest) do not show significant EI>1. This could be due to the fact that, in the target window, auditory evoked responses interfere with sustained entrainment response. This also suggests that the observed EIs in the Carrier and Target windows reflect distinct neural responses. Specifically, the EI in the Target window could have potentially captured the evoked response to the last

words of the Carrier window. Yet, if it were the case, the EI would have been expected to be at a similar location in both Carrier and Target windows, i.e. in primary auditory areas, which is not what we observed.

As shown in the previous section, listening to Fast and Slow speech rate conditions lead to different target word percepts. Hence, it could be argued that the observed EI potentially reflects the brain response to the stimulus percept, rather than sustained entrainment to past speech rate. To discard this hypothesis, we ran the same EI analyses but categorized the trials based on the reported perception of the target word percept (Long/Short vowel word percept) and not based on past speech rate condition (Slow/ Fast). When contrasting word percepts, no significant cluster was found using whole brain source statistics. Power in the right middle temporal cortex at 3 Hz and 5.5 Hz did not significantly differ between Short and Long vowel percept conditions (Figure S2).

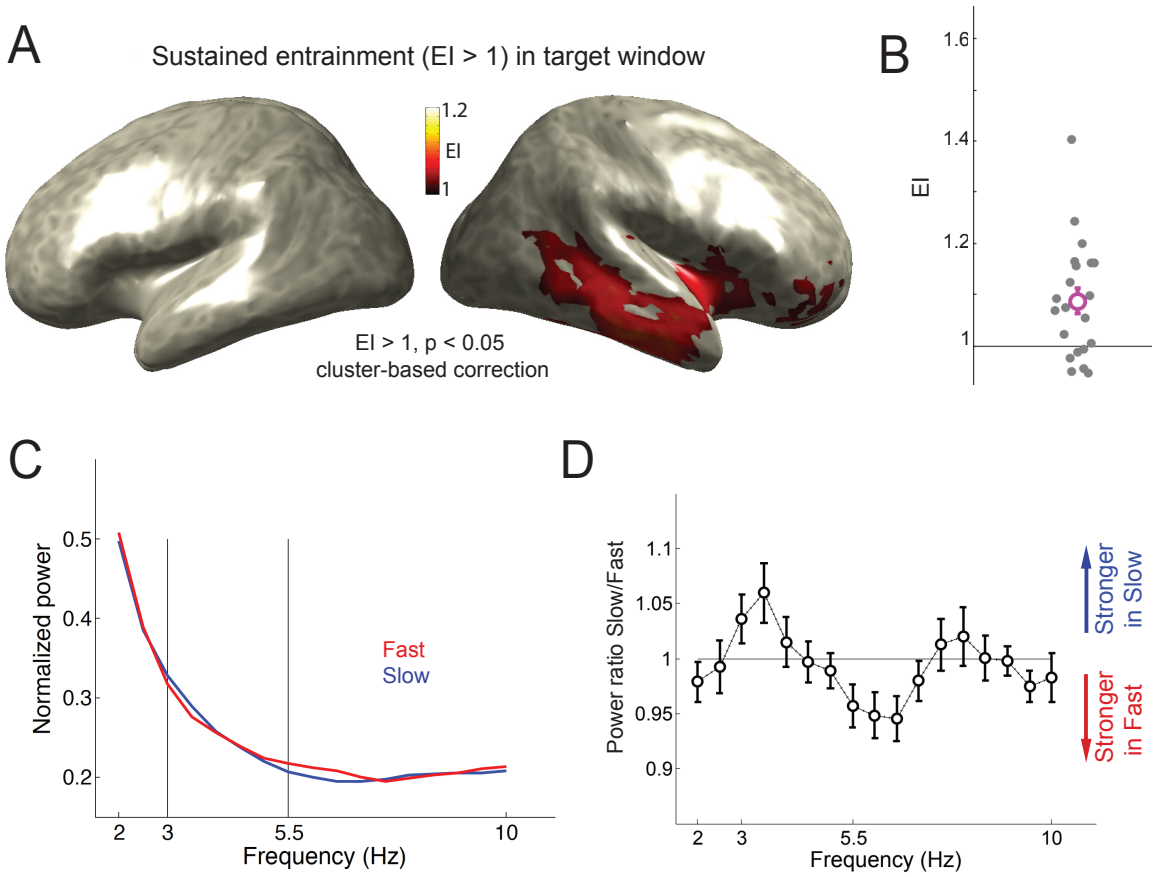

Figure 3: Sustained neural entrainment during target window. A) During the target window, sustained entrainment to the preceding speech rate was observed, most prominently in right middle-temporal and right infero-frontal areas. EI values are thresholded at $p < 0.05$, controlled for multiple comparisons using cluster-based permutation tests. B) Entrainment Index within the most strongly activated grid point (MNI coordinates: 50, -40, -10, Right Middle Temporal Cortex). Each grey dot corresponds to one participant. The magenta circle corresponds to the average EI across participants. Error bars represent s.e.m. The horizontal display of the dots is dispersed to prevent too much overlap. C) Power spectra of neural activity for the Fast (red) and Slow (blue) speech rate conditions within the most strongly activated grid point. D) Contrast in power between the two speech rate conditions (using the power ratio Slow/Fast) within the most strongly activated grid point. Ratios higher than 1 reflect stronger power in Slow speech rate condition, ratios lower than 1 reflect stronger power in Fast speech rate condition. Error bars denote s.e.m. See also Figure S1 and Figure S2.

Crucially, sustained entrainment correlated with behavioral performance, so that participants with stronger entrainment were also more strongly biased in their perceptual reports by the contextual speech rate. We correlated the EI observed in the most activated grid point of the significant cluster (in right middle temporal cortex, MNI coordinates: 50, -40, -10) to the Perceptual Bias of each participant. A significant positive correlation was observed between the EI in the Target window and the Perceptual Bias (Spearman's rho: 0.54, $p = 0.018$, Figure 4A), suggesting that participants with stronger sustained entrainment (i.e. high EI in the target window) had a stronger Perceptual Bias, i.e., were more influenced by the preceding speech rate in the perception of the target word (more likely to perceive a short /ɑ/ after a slow speech rate, and a long /a:/ after a fast speech rate). Hence, inter-subject variability in the strength of sustained entrainment was observed and could predict how susceptible participants' judgments on the target word were affected by contextual speech rate. In contrast, the EI in the Carrier window did not correlate with the Perceptual Bias (at rMTC grid point: rho = -0.10, $p = 0.67$; at rSTC grid point: rho = -0.09, $p = 0.69$) nor with the EI in the Target window (rho = -0.06, $p = 0.78$). As discussed earlier, the EI in the Carrier window may capture both endogenous entrainment and stimulus-driven evoked responses to speech. Hence, in the Carrier window we cannot isolate the sustained entrainment response, as the EI likely reflects different mechanisms that overshadow each other. In particular, there is no ground to assume that the strength of the sustained entrainment is linked to the strength of the stimulus-driven response to speech. The strength of auditory stimulus-driven responses varies with the sound quality, loudness, or presence of background noise. Sustained entrainment should speculatively not behave like stimulus-driven responses and should not be affected by adverse listening conditions if it reflects a default temporal predictive mechanism.

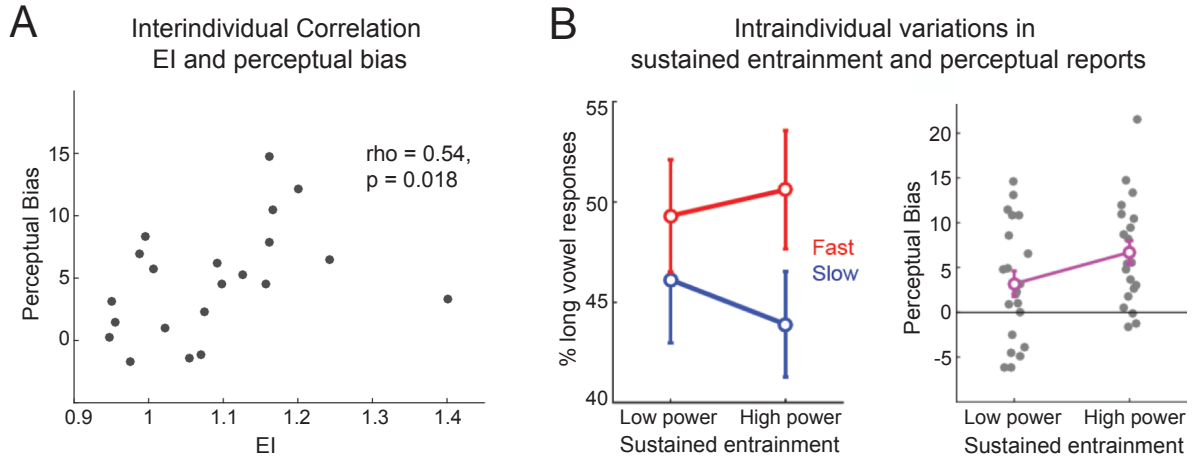

Figure 4: Sustained neural entrainment during target window predicts speech perception. A) Correlation between sustained entrainment (as measured by EI) and Perceptual Bias. Each dot corresponds to one participant. The stronger the sustained entrainment, the stronger the influence of preceding speech rate on target word percept. B) For each participant, the data were median-split based on the strength of sustained entrainment to the preceding speech rate: for the Fast rate condition (red), the trials were median-split based on the observed 5.5 Hz power. For the Slow rate condition, the trials were median-split based on the observed 3 Hz power. Left panel: Proportion of long vowel responses as a function of the strength of the sustained entrainment for the Fast (red) and Slow (blue) rate conditions. Error bars denote s.e.m. More long vowel percepts were observed for trials with strong sustained entrainment to the Fast speech rate; conversely more short vowel percepts were observed for trials with strong sustained entrainment to the Slow speech rate. Right panel: Perceptual Bias as a function of the strength of sustained entrainment. Each dot corresponds to one participant. The magenta circle corresponds to the average perceptual bias across participants. Error bars denote s.e.m. The stronger the sustained entrainment to the preceding speech rate, the stronger the perceptual bias.

We also asked whether sustained entrainment positively correlated with the Perceptual Bias on a trial-by-trial basis. For each participant, the individual data were split into two groups of trials based on the strength of sustained entrainment in the Target window. For the Fast rate condition, the trials were median-split based on the power of sustained 5.5 Hz oscillations. For the Slow rate condition, the trials were similarly divided based on the observed 3 Hz power. We observed that the strength of sustained entrainment impacted the perceptual reports at the trial level. More long vowel percepts were observed for trials with strong sustained entrainment to the Fast speech rate; conversely more short vowel percepts were observed for trials with strong sustained entrainment to the Slow speech rate (Figure 4B, left panel, interaction between Speech rate and Strength of sustained entrainment $F(1,20) = 3.77$; $p = 0.066$, marginally significant, $\eta p^2 = 0.16$). Stronger sustained entrainment was thus associated with a stronger Perceptual Bias (Figure 4B, right panel).

Finally, to estimate how long the sustained entrainment persisted, we performed time-resolved analyses of the Entrainment Index. These analyses only provide a broad descriptive evolution of the sustained entrainment. The analyses were performed at two representative grid points, corresponding to the maximum EI values in Carrier window (located in right Superior Temporal Cortex, rSTC) and Target window (located in right Middle Temporal Cortex, rMTC). At rMTC grid point (Figure 5A), the EI in the Target window showed a tendency in the expected direction (values were higher than 1) up to 750 ms post speech rate change, but the effects were not significant (p-values of one-sample t-test EI >1 without correction for multiple comparisons, 350 ms: $p = 0.052$, 450 ms: $p = 0.049$, 550 ms: $p = 0.084$, 650 ms: $p = 0.110$, 750 ms: $p = 0.055$, 850 ms: $p = 0.495$, 950 ms: $p = 0.659$). The statistics, though in the expected direction, are not conclusive. We attribute this to a loss in the precision of the frequency analysis as we had to use a shorter time window (700 ms) for the time-resolved analyses. In rSTC grid point, the EI tends to go back to baseline immediately after the speech rate change (Figure 5B).

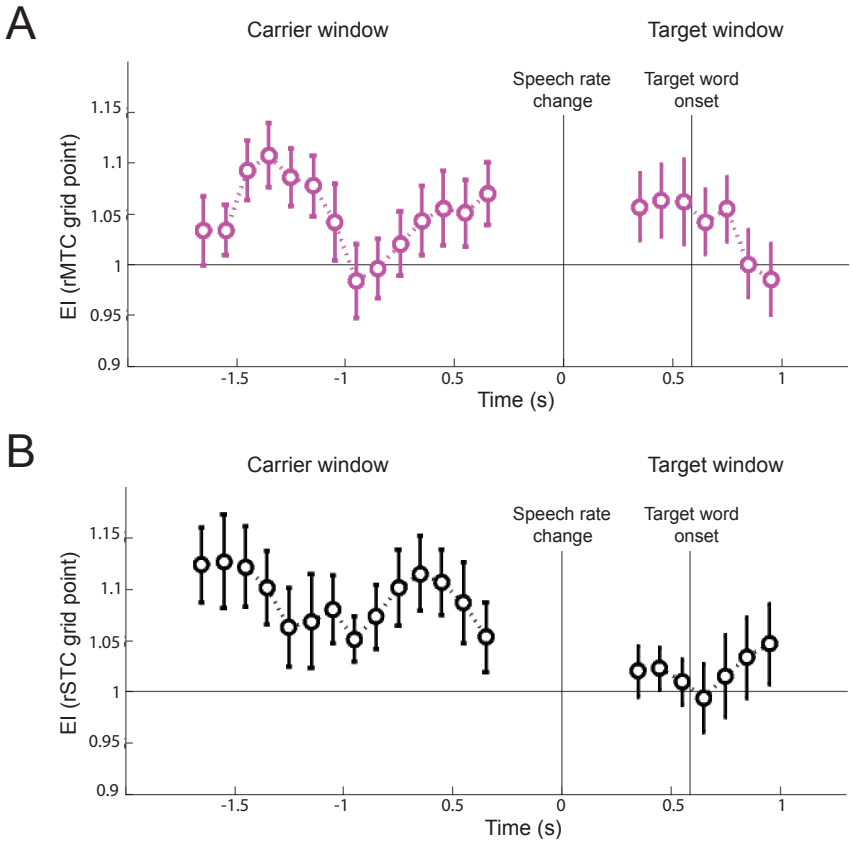

Figure 5: Evolution of Entrainment Index as the sentence unfolds. A) EI at rMTC grid point (most responsive grid point in Target window). B) EI at rSTC grid point (most responsive grid point in Carrier window). Error bars denote s.e.m.

DISCUSSION

We investigated neural oscillatory activity during listening to sentences with changing speech rates. We observed neural oscillatory responses to the syllabic rhythm at the beginning of the sentence (Carrier window). Crucially, entrainment to the preceding speech rate persisted after the speech rate had suddenly changed (i.e., in the Target window), and the observed sustained entrainment biased the perception of ambiguous words in the Target window. The participants who demonstrated stronger sustained entrainment were also more influenced by the preceding speech rate in their perceptual reports. Strong sustained slow rate entrainment was associated with a bias towards short vowel word percepts, and strong sustained fast rate entrainment biased towards more long vowel percepts. Hence, our findings suggest that the neural tracking of the speech temporal dynamics is a predictive mechanism, which is involved in the processing of subsequent speech input and directly influences perception.

To our knowledge, the present results provide the first evidence in human recordings that neural entrainment to speech outlasts the stimulation. Sustained neural entrainment is a crucial prediction in support of the active nature of neural entrainment [1]. First, the sustained entrainment, being independent of the dynamics of the speech tokens, shows that low frequency entrainment to speech rhythms is not purely stimulus driven [26]. Second, sustained entrainment to the temporal statistics of past sensory information supports the hypothesis that neural entrainment builds temporal predictions [2,6]. Recent reports have shown that parieto-occipital alpha oscillations outlast rhythmic visual presentation [5] or brain stimulation [27]. In an electrophysiological study with monkeys, Lakatos and colleagues [4] showed that auditory entrainment in the delta band (1.6 – 1.8 Hz) outlasts the stimulus train for several cycles and argued that the reported sustained entrainment could be of crucial importance for speech processing. The current findings support this view, showing that sustained entrainment is observable in human temporal cortex and influences speech perception.

The present findings support oscillatory models of speech processing [8–10], which suggest that neural entrainment is a mechanism recruited for speech parsing. In these models, neural theta oscillations (3-8 Hz; entraining to syllabic rates) flexibly adapt to the ongoing speech rate, and each cycle of the entrained oscillations chunks a discrete acoustic token of syllabic length from the continuous acoustic signal. The entrained oscillations hence serve as temporal reference frames for speech perception [13,28]: as the entrained frequency matches the spoken syllabic rate, one cycle of the neural oscillation corresponds to the expected duration of syllabic tokens. Modulations in the frequency of entrained theta oscillations should then modify the expected average syllable duration, potentially affecting the perception of some words. In the present study, the observed effects of speech rate on the perceived vowel of the target word are then interpretable as a mismatch in the actual duration of incoming syllables and the predicted syllabic duration defined by the frequency of entrained oscillations [8,9,23,29,30]. A preceding fast rate generates sustained neural entrainment of a faster rate (i.e. shorter expected syllable duration) than the monosyllabic word being parsed. This leads to an overestimation of the word duration – in particular of its vowel in our design – and biases percepts towards a word with a long vowel. Conversely, slower sustained neural entrainment could lead to underestimation of the syllable's duration biasing perception towards a short vowel word.

We speculate that sustained entrainment could also be at the origin of other perceptual effects of contextual speech rate: if entrainment delineates parsed tokens within continuous speech, then distinct sustained entrainment frequencies could lead to changes in the perceived word segmentation [22], and sustained entrainment could even cause the omission of certain words

[21] if occurring at the phase of entrained oscillations that marks the boundary between discretized tokens. However, neural entrainment mechanisms may not account for all known contextual effects in speech processing. For instance, exposure to certain spectral features, such as specific formant frequencies, can also influence the perception of ongoing words [31]. In this case, neural entrainment is not expected to play any particular role because the adaptation does not rely on temporal regularities. Lastly, it has been reported that exposure to distinct speech rates entails long lasting effects on speech perception (up to one hour) [32,33]. These long-lasting effects of speech rate are speculatively still compatible with the oscillatory parsing hypothesis. Auditory cortices have a preferred frequency of entrainment in the theta range [34,35], and this preferred frequency of entrainment is thought to be shaped by experience (in particular via speech exposure [36]). Hence, exposure to speech at specific rates could modulate the preferred 'resonance' frequency of auditory cortices and influence how accurately they entrain to new signals.

Our findings provide neurophysiological evidence that neural oscillations build predictions on past rhythmic sensory information [15,16,37] that affects speech comprehension [13,23,32] Specifically, we used sensory history as a relevant means to modulate ongoing oscillatory activity without any difference in sensory stimulation during the Target window, and showed that sustained entrainment correlates with perceptual word reports. These results are in line with recent transcranial stimulation studies suggesting a causal link between neural entrainment and speech perception [38–40]. Yet, in our study as well as in brain stimulation studies, the causal link between the observed sustained neural entrainment and speech perception cannot unambiguously be determined. For our study, we cannot fully exclude the existence of a hidden neural variable dependent on speech rate history that would be the causal factor affecting speech perception. Similarly, transcranial brain stimulation reports assume that the stimulation effects originate from the recruitment of neural oscillations in the cortex. However, this assumption is still debated [41], and it cannot be excluded that the stimulation affects non-oscillatory brain mechanisms as well. Nevertheless, we believe that our data point to a causal chain that is initiated by the neural entrainment, and results in consequences for perception. That is, despite an identical speech signal in the Target window in the two entrainment conditions, a difference in speech perception is observed in relation to a change in the entrainment scheme. We also think that our data cannot easily be explained by alternative hypotheses such as auditory habituation or neural fatigue. First, it is unclear how an auditory habituation account would predict sustained entrainment, specifically how neural fatigue would generate 5.5 Hz neural oscillations in the target window for the Fast rate condition, and 3 Hz oscillations in the Slow rate condition. Second, we observed sustained entrainment outside primary auditory cortices, while auditory habituation would predict sensory history to modulate neural activity in primary auditory cortex.

Sustained entrainment was most prominently observed in the right middle temporal areas. Though the lateralization of the sustained entrainment was not explicitly tested and is not the main focus of this study, this observation is in line with evidence that the right superior temporal sulcus is specialized in processing sound events of syllabic length (~250 ms) [42,43], and that the tracking of the speech envelope [44–46], and of slow spectral transitions [47,48] or prosodic cues [49] are known to be stronger in right auditory cortices [50]. It should be noted that the observed sustained entrainment might mainly impact the perception of long speech segments like vowels. Vowels form the nuclei of the syllable, as such they are segments of long duration (100 - 200 ms) and carry the strongest energy fluctuations of the envelope of the acoustic signal. It is unclear how the observed sustained entrainment would impact consonantal processing. The Asymmetric Sampling Theory would suggest that the perception of consonants rely on other

tracking mechanisms, based on oscillations of a higher frequency (in the gamma range) and which would be left lateralized [43,46]. In contrast, experimental reports have shown that theta oscillations in bilateral auditory cortices reflect consonant-level processing [51,52]. In particular, Ten Oever and colleagues have found that the pre-stimulus phase of ongoing theta oscillations determines the perceived consonants of spoken syllables [52].

The results confirm that the tracking of auditory temporal regularities affects speech processing. Yet, the relevance of neural oscillations in building temporal predictions based on past temporal statistics may be a general property of sensory processing [53–55], in line with the idea that oscillations provide temporal metrics for perception [28,56]. Additionally, the current study was focused on the neural entrainment to the strongest rhythmic cues in the speech envelope, i.e., syllabic rhythms, operated by theta oscillations (3-8 Hz). We hypothesize that the observed sustained entrainment would primarily influence the processing of speech acoustic features considering that theta oscillations are linked to acoustic parsing [57] and phonemic processing [51,52], while they do not seem to be involved in parsing of words in the absence of relevant acoustic cues [30]. Theta oscillations would then serve a distinct role compared to oscillations in the delta range (1-3 Hz): theta would be involved in the acoustic parsing of continuous speech into words, while delta oscillations would combine the segmented words into larger linguistic discrete structures based on procedures underlying syntactic and semantic combinatoriality [13,58,59].

In summary, the present results show neural entrainment to speech is not purely stimulus driven and is modulated by past speech rate information. Sustained neural entrainment to past speech rate is observed, and it influences how ongoing words are heard. The results thus support the hypothesis that neural oscillations actively track the dynamics of speech to generate temporal predictions that bias the processing of ongoing speech input.

ACKNOWLEDGEMENTS

We would like to thank Annelies van Wijngaarden for the recordings of her voice and Anne van Hoek for help with pretesting. This research was supported by the Netherlands Organisation for Scientific Research (NWO) Gravitation Grant 24.001.006 to the Language in Interaction Consortium, and by the NWO Spinoza Prize and by the Academy Professorship Award of the Royal Netherlands Academy of Arts and Sciences to P.H.

AUTHORS CONTRIBUTIONS

Conceptualization, A.K., H.R.B, A.M., O.J., P.H.; Methodology, A.K., H.R.B., A.T.; Software, A.K., H.R.B., A.T.; Investigation, A.K., H.R.B., A.T.; Resources, P.H.; Writing – Original Draft, A.K.; Writing – Review & Editing, A.K., H.R.B, A.T., A.M., O.J., P.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Thut, G., Schyns, P.G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. Front. Psychol. *2*, 170.
2. Schroeder, C.E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. *32*, 9–18.
3. Large, E.W., and Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. Psychol. Rev. *106*, 119.
4. Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., and Schroeder, C.E. (2013). The Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron *77*, 750–761.
5. Spaak, E., de Lange, F.P., and Jensen, O. (2014). Local entrainment of α oscillations by visual stimuli causes cyclic modulation of perception. J. Neurosci. *34*, 3536–44.
6. Morillon, B., and Schroeder, C. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. Ann. N. Y. Acad. Sci. *1337*, 26–31.
7. Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., and Schroeder, C.E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. *94*, 1904–1911.
8. Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. *15*, 511–7.
9. Peelle, J.E., and Davis, M.H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. Front. Psychol. *3*, 320.
10. Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. *2*, 130.
11. Ding, N., and Simon, J.Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. *8*, 311.
12. Zoefel, B., and VanRullen, R. (2015). The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. Front. Hum. Neurosci. *9*, 651.
13. Kösem, A., and van Wassenhove, V. (2016). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Lang. Cogn. Neurosci., 1–9.
14. Obleser, J., Herrmann, B., and Henry, M.J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. Front. Hum. Neurosci. *6*, 250.
15. Zoefel, B., ten Oever, S., and Sack, A.T. (2018). The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. Front. Neurosci. *12*, 95.
16. Haegens, S., and Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. Neurosci. Biobehav. Rev. *86*, 150–165.
17. Peelle, J.E., Gross, J., and Davis, M.H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex *23*, 1378–87.
18. Ding, N., and Simon, J.Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. *33*, 5728–35.
19. Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M.M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci. U. S. A. *98*, 13367–72.
20. Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., *et al.* (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party." Neuron *77*, 980–991.
21. Dilley, L.C., and Pitt, M.A. (2010). Altering context speech rate can cause words to appear or disappear. Psychol. Sci. *21*, 1664–70.
22. Reinisch, E., Jesse, A., and McQueen, J. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. J. Exp. Psychol. Hum. Percept. Perform. *37*, 978.
23. Bosker, H.R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. Attention, Perception, Psychophys. *79*, 333–343.
24. Reinisch, E., and Sjerps, M.J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. J. Phon. *41*, 101–116.
25. Bosker, H.R., and Kösem, A. (2017). An entrained rhythm's frequency, not phase, influences temporal sampling of speech.
26. Kayser, S.J., Ince, R.A.A., Gross, J., and Kayser, C. (2015). Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. J. Neurosci. *35*, 14691–701.
27. Alagapan, S., Schmidt, S.L., Lefebvre, J., Hadar, E., Shin, H.W., and Fröhlich, F. (2016). Modulation of Cortical Oscillations by Low-Frequency Direct Cortical Stimulation Is State-Dependent. PLOS Biol. *14*,

e1002424.
28. Kösem, A., Gramfort, A., and van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. Neuroimage *92*, 274–284.
29. Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A.-L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. Elife *4*, e06213.
30. Kösem, A., Basirat, A., Azizi, L., and van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. J. Neurophysiol. *116*, 2497–2512.
31. Holt, L., and Lotto, A. (2002). Behavioral examinations of the level of auditory processing of speech context effects. Hear. Res. *167*, 156–169.
32. Baese-Berk, M.M., Heffner, C.C., Dilley, L.C., Pitt, M.A., Morrill, T.H., and McAuley, J.D. (2014). Long-Term Temporal Tracking of Speech Rate Affects Spoken-Word Recognition. Psychol. Sci. *25*, 1546–1553.
33. Maslowski, M., Meyer, A.S., and Bosker, H.R. (2018). How the tracking of habitual rate influences speech perception. J. Exp. Psychol. Learn. Mem. Cogn.
34. Ding, N., and Simon, J.Z. (2009). Neural Representations of Complex Temporal Modulations in the Human Auditory Cortex. J. Neurophysiol. *102*, 2731–2743.
35. Teng, X., Tian, X., Rowland, J., and Poeppel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. PLOS Biol. *15*, e2000812.
36. Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. Neurosci. Biobehav. Rev. *81*, 181–187.
37. Falk, S., Lanzilotti, C., and Schön, D. (2017). Tuning Neural Phase Entrainment to Speech. J. Cogn. Neurosci. *29*, 1378–1389.
38. Zoefel, B., Archer-Boyd, A., and Davis, M.H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. Curr. Biol. *28*, 401–408.e5.
39. Riecke, L., Formisano, E., Sorger, B., Başkent, D., and Gaudrain, E. (2018). Neural Entrainment to Speech Modulates Speech Intelligibility. Curr. Biol. *28*, 161–169.e5.
40. Wilsch, A., Neuling, T., and Herrmann, C.S. (2017). Envelope-tACS modulates intelligibility of speech in noise. bioRxiv.
41. Lafon, B., Henin, S., Huang, Y., Friedman, D., Melloni, L., Thesen, T., Doyle, W., Buzsáki, G., Devinsky, O., Parra, L.C., *et al.* (2017). Low frequency transcranial electrical stimulation does not entrain sleep rhythms measured by human intracranial recordings. Nat. Commun. *8*, 1199.
42. Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci. *8*, 389–95.
43. Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' Speech Commun. *41*, 245–255.
44. Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. *11*, e1001752.
45. Abrams, D.A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. J. Neurosci. *28*.
46. Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S.J., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron *56*, 1127–34.
47. Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., and Samson, Y. (1998). Lateralization of speech and auditory temporal processing. J. Cogn. Neurosci. *10*, 536–540.
48. Zatorre, R.J., and Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. Cereb. Cortex *11*, 946–953.
49. Bourguignon, M., De Tiège, X., De Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., and Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. *34*, 314–326.
50. Scott, S.K., and McGettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? Brain Lang. *127*, 36–45.
51. Di Liberto, G.M., O'Sullivan, J.A., and Lalor, E.C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr. Biol. *25*, 2457–2465.
52. Ten Oever, S., and Sack, A.T. (2015). Oscillatory phase shapes syllable perception. Proc. Natl. Acad. Sci. U. S. A. *112*, 15833–7.
53. Hickok, G., Farahbod, H., and Saberi, K. (2015). The Rhythm of Perception. Psychol. Sci. *26*, 1006–1013.
54. Herrmann, B., Henry, M.J., Haegens, S., and Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. Neuroimage *124*, 487–497.
55. Breska, A., and Deouell, L.Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. PLOS Biol. *15*, e2001665.

56. VanRullen, R. (2016). Perceptual Cycles. Trends Cogn. Sci. *20*, 723–735.
57. Doelling, K.B., Arnal, L.H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage *85 Pt 2*, 761–8.
58. Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. *19*, 158.
59. Park, H., Ince, R.A.A., Schyns, P.G., Thut, G., and Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. Curr. Biol.
60. Moulines, E., and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. *9*, 453–467.
61. Boersma, P., and Weenink, D. (2007). Praat ver. 4.06, software.
62. Bosker, H.R., Reinisch, E., and Sjerps, M.J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. J. Mem. Lang. *94*, 166–176.
63. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput. Intell. Neurosci. *2011*, 1–9.
64. Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A. *98*, 694–9.
65. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Anne Kösem (kosem.anne@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

33 native Dutch speakers took part in the experiment. All participants provided their informed consent in accordance with the declaration of Helsinki, and the local ethics committee (CMO region Arnhem-Nijmegen). Participants had normal hearing, no speech or language disorders, and were right handed. We excluded eight participants who presented strong bias in their perceptual reports (<20 % or >80 % long vowel reports throughout the experiment), one who reported explicit strategies during debriefing, one not passing the pretest, and two participants were excluded due to corrupted MEG data; leaving 21 participants (14 females; mean age: 22 years old) in the analysis (see Table S1 for % long vowel response per participant).

METHOD DETAILS

Stimuli

A female native speaker of Dutch was recorded at a comfortable speech rate producing five different sentences, each ending with "het woord [target]" (meaning: the word [target]). Recordings were divided into two temporal windows. The Carrier windows were composed of the first 12 syllables prior to "het" onset; the Target windows contained the ending "het woord [target]". Carrier sentences did not contain semantic information that could bias target perception and did not contain any /ɑ/ or /a:/ vowels (Table S2). Carriers were first set to the mean duration of the five carriers and then expanded (133% of original rate) and compressed (1/1.33 = 75% of original) using PSOLA [60] in Praat [61], manipulating temporal properties while leaving spectral characteristics intact (e.g., pitch, formants). The resulting Fast and Slow carriers had strong periodic components at 5.5 Hz and 3 Hz, respectively (Figure 1B). Please note that, in order to keep the stimuli as natural as possible we did not make the sentences artificially rhythmic, and we did not control for the phase of presentation of the syllables. Hence, the phase of syllables presentation was varying across carrier sentences while their syllabic rate was similar. The sentence-final Target window ("het woord [target]") was kept at the originally recorded speech rate (i.e., not compressed/expanded). As targets, the speaker produced 14 minimal Dutch word pairs that only differed in their vowel, e.g., "zag" (/zɑx/) - "zaag" (/za:x/), "tak" (/tɑk/) - "taak" (/ta:k/), etc… (Table S3). One long vowel /a:/ was selected for spectral and temporal manipulation, since the Dutch /ɑ/-/a:/ contrast is cued by both spectral and temporal characteristics [24,62]. Temporal manipulation involved compressing the vowel to have a duration of 140 ms using PSOLA in Praat. Spectral manipulations were based on Burg's LPC method in Praat, with the source and filter models estimated automatically from the selected vowel. The formant values in the filter models were adjusted to result in a constant F1 value (740 Hz, ambiguous between /ɑ/ and /a:/) and 13 F2 values (1100-1700 Hz in steps of 50 Hz). Then, the source and filter models were recombined and the new vowels were adjusted to have the same overall amplitude as the original vowel. Finally, the manipulated vowel tokens were combined with one consonantal frame for each of the 14 minimal pairs.

Procedure

Before MEG acquisition, participants were presented with a vowel categorization staircase procedure to estimate individual perceptual boundaries between /ɑ/ and /a:/. It involved the presentation of the target word "dat" (/dɑt/) - "daad" (/da:t/) in isolation (i.e., without preceding

speech) with varying F2 values (1100-1700 Hz), with participants indicating what word they heard. Based on this pretest, 3 F2 values were selected, corresponding to the individual 25%, 50%, and 75% long /a:/ categorization points. These values were used in the MEG experiment, where half of the target words contained an ambiguous vowel (F2 associated to 50% long /a:/ categorization point), a quarter of the target words with a vowel F2 associated to 25% long /a:/ responses, and a quarter target words with a vowel F2 corresponding to 75% long /a:/ responses. In the MEG experiment, stimuli included carrier sentences followed by target sequences. All participants heard the five carriers in both rate conditions in combination with all possible targets in a randomized order. Participants were asked to listen to the full sentences while fixating on a fixation cross on the screen, and to report what the target word was by button press once the response screen appeared (presented 700 ms after target offset, with the two response options presented left and right, e.g., "tak" or "taak"; position counter-balanced across participants). In total, 280 sentences were presented per Slow/ Fast speech rate condition, leading to a total of 560 trials. The experiment included 3 breaks and lasted about 75 min.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

For every participant, behavioral responses (i.e., whether the target word contained a short or a long vowel) were registered for both Fast and Slow conditions. The perceptual bias was calculated as the difference in the proportion of long vowel (/a:/) responses between the Fast and the Slow conditions. Statistical analysis was performed with Matlab R2015a. Repeated measures ANOVA were performed using the proportion of long vowel reports and the perceptual bias as dependent variables and factors of Speech rate (Fast, Slow) and second formant frequency F2 (25%, 50%, 75% long vowel reports F2s).

MEG analysis

MEG recordings were collected using a 275-channel axial gradiometer CTF MEG system at a sampling rate of 1.2 kHz. For source reconstruction analysis, structural magnetic resonance imaging (MRI) scans were obtained from all subjects using either a 1.5 T Siemens Magnetom Avanto system or a 3 T Siemens Skyra system. MEG data was analyzed using the Fieldtrip software [63]. MEG recordings were epoched at two distinct windows (Carrier and Target). Epochs for the Carrier window comprised the MEG recordings at the start of the sentence up to the change in speech rate (fixed 3.55 s duration for the Slow rate condition, 2.0 s for the Fast rate condition). Epochs in the Target window started after the change in speech rate and comprised the MEG recordings during the presentation of the last three words of the sentence ("Het woordje [target word]") up to 500 ms before the response screen (the window was of 1.3s duration for both Fast and Slow conditions). Noisy channels and trials with muscle artifacts were excluded after visual inspection. On average, 13% of trials and 0.7% of channels were discarded. An independent component analysis was performed to remove cardiac and eye movement artifacts.

The sources of the observed 3 Hz and 5.5 Hz activity were computed using beamforming analysis with the dynamic imaging of coherent sources (DICS) technique [64] to the power data. Power was used as a sensitive measure of entrainment in our study as we used natural stimuli as carrier sentences. All carrier sentences had strong periodic components at 3 Hz or 5.5 Hz while being spoken naturally. Because they were not artificially rendered rhythmic, the phase of syllable presentation was not fully controlled. Hence, the experiment was specifically designed to test the effect of speech rate on sustained entrainment frequency power – and not phase, as average time course or phase-locking across trials/subjects would potentially cancel out the effect due to non-phase-locked nature of the stimuli. To do so, the cross-spectral density data structure was computed using Fast Fourier transform (FFT) with Hanning tapering

performed at 3 Hz and at 5.5 Hz for both Carrier and Target windows. Epochs in the target and carrier windows were analyzed separately to prevent leakage of oscillatory components from the carrier to the target window. For the Carrier window, the first 500 ms of the epochs were removed to exclude the evoked response to the onset of the sentence and ensure the measure of the entrainment regime. The data was zero-padded up to 4.0 s for both conditions to match in FFT resolution. During the target window, the data was zero-padded up to 2.0 s so as to obtain more accurate amplitude estimates of the resolvable 3 Hz and 5.5 Hz signals components. The co-registration of MEG data with the individual anatomical MRI was performed via the realignment of the fiducial points (nasion, left and right pre-auricular points). Lead fields were constructed using a single shell head model based on the individual anatomical MRI. Each brain volume was divided into grid points of 1 cm voxel resolution, and warped to a template MNI brain. For each grid point the lead field matrix was calculated. Source reconstruction was then performed using a common spatial filter obtained from beaming data from both Slow and Fast speech rate conditions. The Entrainment Index (EI) was calculated based on the source reconstructed power for each grid point according to the formula:

$$EI = \frac{Power_{Slow}(3\ Hz)}{Power_{Fast}(3Hz)} \cdot \frac{Power_{Fast}(5.5\ Hz)}{Power_{Slow}(5.5Hz)}$$

Sources with significant EI > 1 were estimated using cluster-based permutation statistics across subjects [65]. First, a "null hypothesis" source dataset was generated by setting the EI values to 1. Pairwise t-tests were then computed for each grid point between the experimental EI source data to the generated "null hypothesis" source dataset. Grid points with a p-value associated to the t-test of 5% or lower were selected as cluster candidates. The sum of the t-values within a cluster was used as the cluster-level statistic. The reference distribution for cluster-level statistics was computed by performing 1,000 permutations of the EI and the generated null hypothesis source data. Clusters were considered significant if the probability of observing a cluster test statistic of that size in the reference distribution was 0.05 or lower.
The same procedure was performed for the time-resolved analyses of the Entrainment Index, except that a shorter time window of 700 ms was used to compute the EI at each time point. The 700ms-long time windows were sliding every 100 ms and were centered from 1650 ms up to 350 ms prior the change in speech rate (Carrier window), and from 350 ms up to 950 ms after the change in speech rate (Target window). The data was zero-padded up to 2.0 s so as to obtain accurate amplitude estimates of the resolvable 3 Hz and 5.5 Hz signals components. The analyses were performed at two representative grid points, corresponding to the maximum EI values in Carrier window (MNI coordinates: 60, -20, -10, located in right Superior Temporal Cortex) and Target window (MNI coordinates: 50, -40, -10, located in Right Middle Temporal Cortex). For illustration purposes in Figures 2 C-D and 3 C-D, we computed the power spectra in the Carrier and Target windows from 2 Hz to 10 Hz in 0.5 Hz steps, using the same parameters as for the Fourier transform at the frequencies of interest (3Hz and 5.5 Hz). Normalization was performed for visualization and consisted in dividing the individual power spectra by the average power across conditions.

The inter-individual correlation between brain data and perceptual bias was performed within the most strongly activated grid point (grid point with highest *t*-value) located within the significant observed cluster. Single-trial power analysis was computed at this grid point to estimate the inter-trials effects of sustained entrainment on the Perceptual Bias. Single-trial time series were first computed using a Linearly constrained minimum-variance (LCMV) beamformer spatial filter. Data were projected onto the dipole direction that explained most variance using SVD. The power at 3 Hz and 5.5 Hz was estimated for each trial using the same

parameters as for the first analysis. The trials were sorted in two groups based on the strength of the oscillatory component corresponding to the initial speech rate (3 Hz for Slow rate condition, 5.5 Hz for Fast rate condition). The % long vowel responses were then contrasted between the two groups using a two-way repeated measure ANOVA with Speech rate (Fast, Slow) and Sustained Entrainment Strength (Low, High) as factors.

DATA AND SOFTWARE AVAILABILITY

Data is available from the Donders Repository at
http://hdl.handle.net/11633/di.dccn.DSC_3011050.03_094

SUPPLEMENTAL INFORMATION

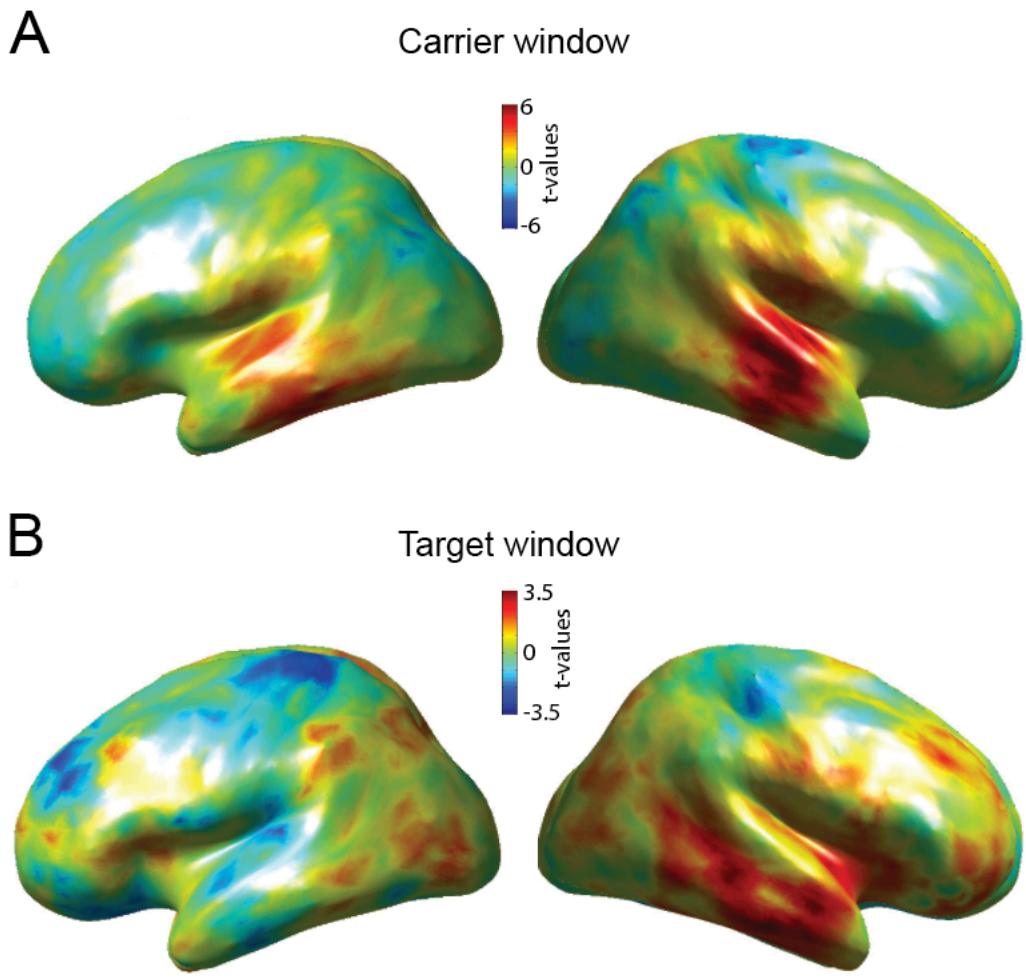

Figure S1: Unthresholded *t*-maps of the EI vs. 1 (null hypothesis). Related to Figure 2 and Figure 3. A) during the Carrier window. B) in the Target window.

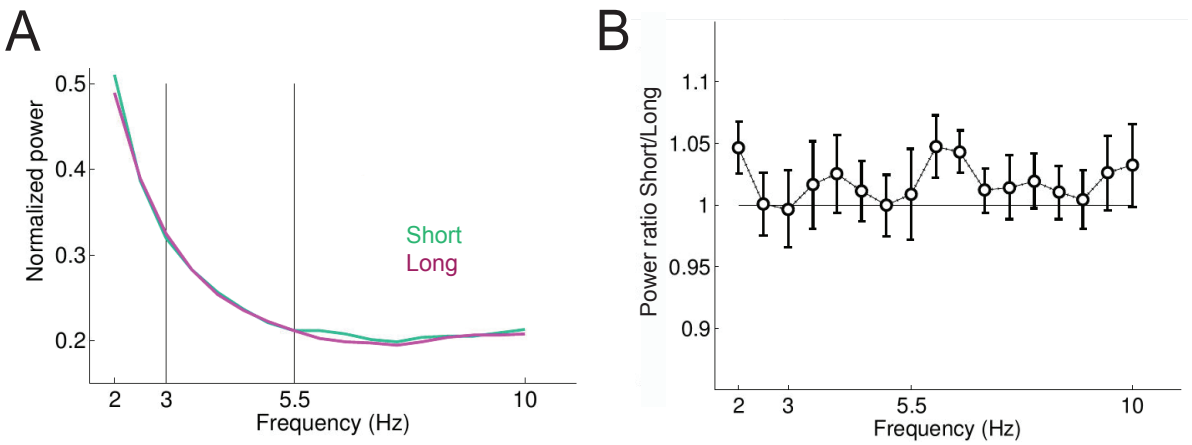

Figure S2: Power analysis in Target window as a function of vowel percept. Related to Figure 3.
A) Power spectra of neural activity for the Short vowel percept (green) and Long vowel percept (pink) conditions within the rMTC grid point (MNI coordinates: 50, -40, -10). B) Contrast in power between the two vowel percept conditions (using the power ratio Short/Long) within rMTC grid point. Ratios higher than 1 reflect stronger power in Short vowel percept condition, ratios lower than 1 reflect stronger power in Long vowel percept condition. Error bars denote s.e.m.

PARTICIPANT #	% LONG VOWEL RESPONSE	INCLUDED/REJECTED - REASON
1	12	rejected - bias in perceptual report
2	1	rejected - bias in perceptual report
3	63	included
4	1	rejected - bias in perceptual report
5	34	included
6	50	included
7	60	included
8	57	included
9	41	included
10	42	included
11	/	rejected- did not pass pretest
12	52	included
13	57	included
14	7	rejected - bias in perceptual report
15	66	included
16	54	rejected- corrupted MEG data
17	35	included
18	62	rejected- corrupted MEG data
19	27	included
20	52	included
21	7	rejected - bias in perceptual report
22	36	included
23	11	rejected - bias in perceptual report
24	47	included
25	48	rejected - reported using strategies during debriefing
26	6	rejected - bias in perceptual report
27	28	included
28	39	included
29	84	rejected - bias in perceptual report
30	54	included
31	66	included
32	59	included
33	30	included

Table S1: Percentage of long vowel response for every participant, and cause of inclusion/ rejection. Related to STAR Methods.

SENTENCE #	DUTCH	ENGLISH PARAPHRASE
1	*Freek ging het hok eerst in en toen weer uit en zei dus…*	"Freek first went in to the shack and then out again and said…"
2	*Job sprong toch eerst nog een keer op de kist en koos toen…*	"Job first jumped up unto the box another time and chose…"
3	*Teun zocht de poes eerst links en toen nog rechts en riep steeds…*	"Teun looked for the cat first left and then right and repeatedly called out…"
4	*De boot met het zeil voer snel weg en plots klonk toen nog…*	"The boat with the sail sailed off quickly and suddenly there sounded…"
5	*Kees vond de soep niet zuur en ook niet zoet dus koos hij…*	"Kees found the soup not sour and also not sweet so he chose…"

Table S2: List of Dutch carrier sentences, with English paraphrases. Related to Audio S1, Audio S2, and STAR Methods. In the experiment, each carrier sentence was followed by a target sequence: "the word [target]".

TARGET #	DUTCH	ENGLISH TRANSLATION
1	*dag - daag*	"day" - "sue" (1 sg. pres.)
2	*dan - Daan*	"then" - "Daan" (proper name)
3	*dat - daad*	"that" - "deed"
4	*sap - Saab*	"juice" - "Saab" (brand)
5	*stad - staat*	"city" - "state"
6	*staf - staaf*	"staff" - "bar"
7	*stak - staak*	"stab" (1 sg. past) - "stalk"
8	*Stan - staan*	"Stan" (proper name) - "stand"
9	*stand - staand*	"stance" - "standing"
10	*star - staar*	"strict" - "stare"
11	*tak - taak*	"branch" - "task"
12	*zag - zaag*	"saw" (1 sg. past) - "saw" (noun)
13	*zak - zaak*	"bag" - "case"
14	*zat - zaad*	"sat" (3 sg. past) - "seed"

Table S3: List of Dutch target word pairs, with English translations. Related to Audio S1, Audio S2, and STAR Methods.

Audio S1: Exemplar speech sentence, slow speech rate condition. Related to Figure 1 and STAR Methods

Audio S2: Exemplar speech sentence, fast speech rate condition. Related to Figure 1 and STAR Methods