Audiotactile Stimulation Can Improve Syllable Discrimination through Multisensory Integration in the Theta Frequency Band

Pierre Guilleminot[1], Cosima Graef[1], Emilia Butters[2], and Tobias Reichenbach[3]

Abstract

■ Syllables are an essential building block of speech. We recently showed that tactile stimuli linked to the perceptual centers of syllables in continuous speech can improve speech comprehension. The rate of syllables lies in the theta frequency range, between 4 and 8 Hz, and the behavioral effect appears linked to multisensory integration in this frequency band. Because this neural activity may be oscillatory, we hypothesized that a behavioral effect may also occur not only while but also after this activity has been evoked or entrained through vibrotactile pulses. Here, we show that audiotactile integration regarding the perception of single syllables, both on the neural and on the behavioral level, is consistent with this hypothesis. We first stimulated participants with a series of vibrotactile pulses and then presented them with a syllable in background noise. We show that, at a delay of 200 msec after the last vibrotactile pulse, audiotactile integration still occurred in the theta band and syllable discrimination was enhanced. Moreover, the dependence of both the neural multisensory integration as well as of the behavioral discrimination on the delay of the audio signal with respect to the last tactile pulse was consistent with a damped oscillation. In addition, the multisensory gain is correlated with the syllable discrimination score. Our results therefore evidence the role of the theta band in audiotactile integration and provide evidence that these effects may involve oscillatory activity that still persists after the tactile stimulation. ■

INTRODUCTION

Speech is built from units of increasing duration and complexity, from phonemes to syllables, words, and sentences. Extraction of the linguistic information in the brain and its subsequent processing requires some segmentation of the acoustic signal into these units (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). This segmentation presumably relies on cortical activity that tracks the rhythms set by the different speech units (Giraud & Poeppel, 2012).

The theta frequency range of cortical activity, between 4 Hz and 8 Hz, appears especially important as it corresponds to the rhythm of syllables and has been demonstrated to reflect several speech-related processes (Di Liberto, O'Sullivan, & Lalor, 2015; Ding & Simon, 2014; Doelling, Arnal, Ghitza, & Poeppel, 2014). Moreover, influencing the cortical tracking in the theta frequency range through transcranial alternating current stimulation (tACS) has been found to modulate speech comprehension (Keshavarzi, Kegler, Kadir, & Reichenbach, 2020). The delta frequency range, between 1 Hz and 4 Hz,

is important as well as it corresponds to the rhythm of words and has been linked to syntactic and semantic information (Weissbart, Kandylaki, & Reichenbach, 2020; Ding, Melloni, Zhang, Tian, & Poeppel, 2016). However, it should be noted that influencing the cortical tracking in that range using taCS did not modulate speech comprehension (Keshavarzi et al., 2020).

The neural activity in the theta band is also thought to support multisensory integration for speech comprehension as well as for attention to speech (Keil & Senkowski, 2018; Lakatos et al., 2009). Because the cortical activity in frequency bands such as the theta band is often assumed to be oscillatory, the multisensory integration has been hypothesized to work based on a phase reset mechanism. Under this hypothesis, the ongoing oscillations in the theta frequency band are reset through a stimulus, yielding and enhancing the observed cortical tracking of amplitude changes in a continuous stimulus such as speech (van Atteveldt, Murray, Thut, & Schroeder, 2014; Mercier et al., 2013). This hypothesis is supported by the observation of resets of activity in the auditory cortex through visual inputs (Kayser, 2009). Regarding speech processing, the neural activity linked to speech rhythms may accordingly be reset by stimuli in other sensory modalities, possibly impeding or enhancing speech comprehension (Crosse, Di Liberto, & Lalor, 2016; Crosse,

[1]Imperial College London, United Kingdom, [2]University of Cambridge, United Kingdom, [3]Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Butler, & Lalor, 2015). In contrast to this bottom–up mechanisms, multisensory effects are also modeled as a top–down attentional modulation. In that regard, alpha band activity has been shown to reflect such a multisensory effect (Misselhorn, Friese, & Engel, 2019).

Tactile stimuli have been found to elicit activity in the auditory cortex, and this neural activity becomes integrated with auditory activities at similar latencies as the processing of the auditory information itself (Fu et al., 2003; Foxe et al., 2002; Schroeder et al., 2001; Foxe et al., 2000). Somatosensory stimuli, in particular, have been shown to reset the phase of ongoing oscillatory activity in the auditory cortex (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007) as well as to modulate the cortical tracking of the amplitude fluctuations in speech (Riecke, Snipes, van Bree, Kaas, & Hausfeld, 2019).

Regarding behavior, studies have found that speech comprehension can be modulated using tactile stimuli derived from slow rhythms of speech, generally after training of the participants (Cieśla et al., 2022; Fletcher, Mills, & Goehring, 2018). We showed recently that sparse vibrotactile signals delivered at the rhythm of syllables can modulate both speech comprehension and certain neural responses to speech (Guilleminot & Reichenbach, 2022). In particular, the tactile stimuli can both enhance as well as reduce speech comprehension, depending on the delay between the vibrotactile pulses and the perceptual centers of the syllables. The dependence of the speech comprehension score on the delay follows a sinusoidal function with a frequency that lies in the theta range: Speech comprehension is maximal when the oscillations reach their peak. This behavior can possibly be explained by a multisensory phase reset mechanisms. By resetting the phase of the auditory oscillations, the tactile stimulation indicates the incoming syllables and the phase is maximal when they should be integrated, thus improving syllable parsing. However, because this study employed continuous speech, it remains unclear whether the observed sinusoidal dependence reflects oscillatory brain activity or rather a regularity in the timing of the syllables (van Bree, Sohoglu, Davis, & Zoefel, 2021).

Here, we aim to further investigate the audiotactile integration. In particular, we wish to test whether tactile stimuli can affect the comprehension of a single syllable even after the tactile stimuli have ended, as would be expected if the stimuli entrained ongoing oscillatory brain activity. To do so, we limit speech comprehension to a syllable discrimination task. Indeed, by stimulating the activity in the auditory cortex through vibrotactile stimulation at a naturally occurring syllabic rate, we should observe an improvement of the syllable parsing process and therefore an improvement in syllable discrimination. We therefore aim to investigate both behaviors, through measuring the ability of participants to discriminate between two similarly sounding syllables, as well as electrophysiology, to elucidate neural correlates of audiotactile integration.

METHODS

Participants

Sixteen participants (aged 22 ± 2 years, 7 women) participated in the experiment. The number of participants was chosen based on previous related work (Guilleminot & Reichenbach, 2022; Keshavarzi et al., 2020; Fletcher et al., 2018; Crosse et al., 2016). All participants were native English speakers, right-handed, and had no history of neurological disorders or hearing impairment. No prior training regarding the tactile stimulation, with or without audio, was given beforehand. Because of equipment failure, one of the participants' EEG could not be recorded. Volunteers gave their written informed consent before the experiment. The research was approved by the Imperial College research ethics committee with reference: 19IC5388 NoA 1.

Experimental Design

Participants were first stimulated with five subsequent vibrotactile pulses that occurred at a rate of 5 Hz, that is, two successive pulses were spaced 200 msec apart (Figure 1). Background noise was presented continuously throughout the trial. After a delay of 150 msec to 300 msec from the time of the last pulses, participants heard a short syllable. They then had to choose which of two similarly sounding syllables was presented, a task for which they were allowed to take as much time as needed.

EEG responses were recorded throughout the experiment, starting before the delivery of the first vibrotactile pulse and ending after the presentation of the syllable.

Hardware

All signals were created digitally on a personal computer (Windows 7 operating system). They were then synchronized and converted into analogue signals through the RX8 multi-I/O processor device (Tucker-Davis Technologies [TDT]) at the sampling rate of 39,062.5 Hz. Participants listened to the acoustic stimuli through insert earphones (ER-2, Etymotic Research) placed in their ear canals. For tactile stimulation, participants were holding a vibrotactile motor (Tactuator MM3C-HF, TactileLabs) between the thumb and index finger of their right hand.

EEG signals were acquired using 64 active electrodes (actiCAP, BrainProducts) and a multichannel EEG amplifier (actiCHamp, BrainProducts). For a precise synchronization of the sensory stimuli and the EEG response, the delivered signals were recorded together with the EEG signals (StimTrak, BrainProducts, Germany), and each trial was preceded by a trigger signal sent to the TDT processor. The tactile and audio signals were sent using the TDT RX8, thus ensuring their synchronization together.
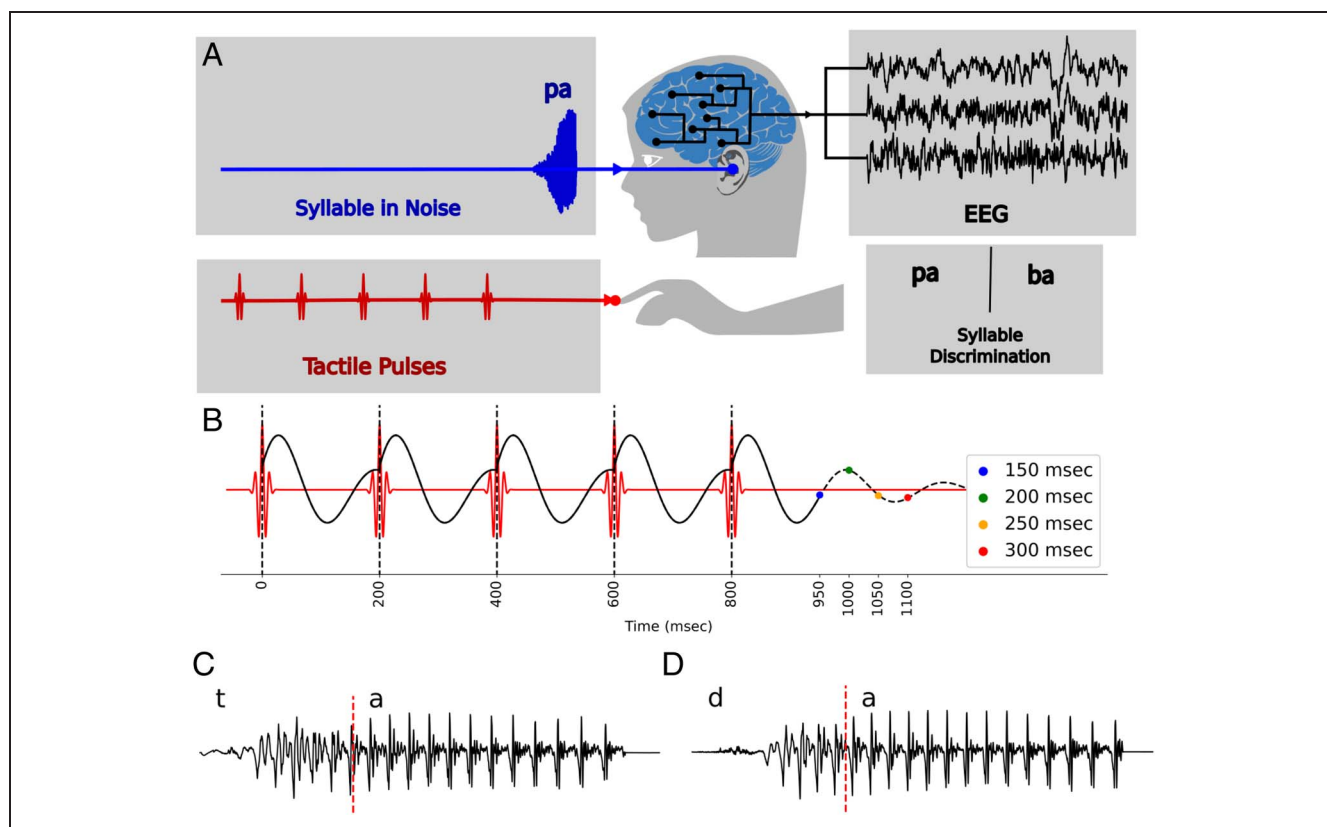
Figure 1. Experimental setup. (A) Each participant was first stimulated with vibrotactile pulses at their right index finger and subsequently heard a syllable in noise that they had to discriminate from another similarly sounding syllable. During the experiment, brain activity was recorded using EEG. (B) The stimuli consisted of five consecutive vibrotactile pulses (red line) at a rate of 5 Hz, followed by a syllable whose onset occurred 150 msec, 200 msec, 250 msec, or 300 msec (blue, green, yellow, and red disk) after the last vibrotactile pulse. Putative oscillatory neural activity in the theta range might be reset by the vibrotactile pulses and decay after the last pulse (black line). (C, D) The waveforms for one pair of syllables: ta and da. The red line marks 50 msec.

Acoustic Stimuli

In each trial of the experiment, participants first experienced five vibrotactile pulses, listened to a syllable in noise, and then selected the presented syllable from two possible choices. The two syllables from which the participant could choose in a given trial were displayed on the left and right side of a monitor, and participants were asked to indicate their choice by pressing the left or right key, respectively, on the keyboard. The syllables were always presented within one of the following three pairs: (pa, ba), (ta, da), and (ga, ka). These syllables all have a similar manner of articulation, and the two syllables of a given pair also possess a similar place of articulation (Blumstein & Stevens, 1979). Indeed, all of the syllables are made of a stops consonant, and the only difference between the two members of a same pair is whether they are voiced or unvoiced. Each member of the pairs can therefore be easily confused for the other, making them good candidates for a discrimination task.

The syllables were converted to both an artificial female and a male voice using the text-to-speech software TextAloud at a sampling rate of 44,100 Hz. The audio stimuli were manipulated using Praat to have the same duration and intensity (Boersma & Van Heuven, 2001). The syllables thus created were modified so that the transition of consonant ("d," "t," "p," "b," "g," "k") to vowel (always "a") would happen 50 msec after the start of the syllable, that is, the first change in amplitude, as represented in Figure 2C and D. The information that allowed for syllable discrimination was therefore contained within the first 50 msec. This duration was chosen because of the later defined audiotactile lags that are separated by 50 msec each. Containing the information into such a window therefore allows our sampling of different delays to be enough to detect a potential effect.

The signal was then mixed with speech-shaped noise at a signal-to-noise ratio of −10 dB. The speech-shaped noise was generated so that it covered the same spectral content as the target speech. The resulting audio signals were delivered to the participants at an intensity of 75 dB SPL, deemed comfortable by the volunteers. We refer to this signal as the audio stimuli or the syllables in the rest of the article.

Tactile Stimuli

The tactile stimuli consisted of five vibrotactile pulses that were presented in succession at a frequency of 5 Hz,
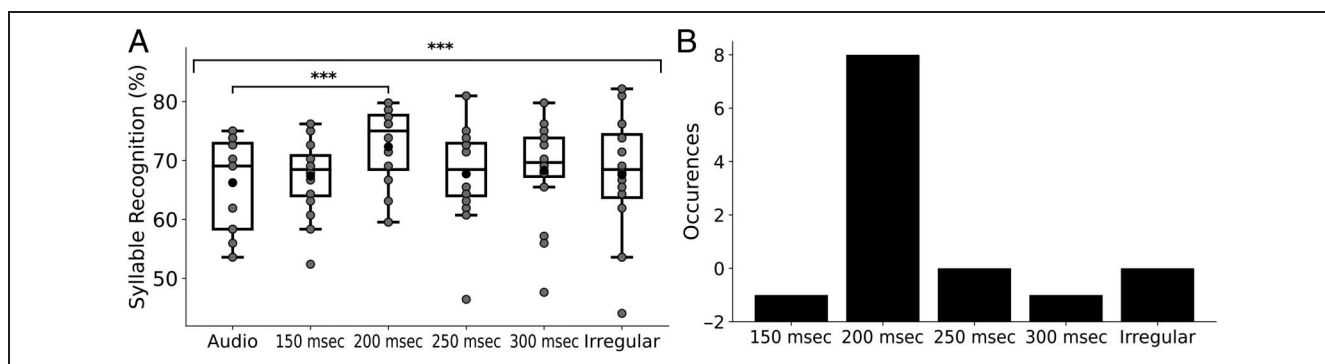
Figure 2. Syllable recognition scores for the different conditions. (A) The scores differed significantly between the conditions. In particular, syllable recognition at an audiotactile lag of 200 msec was significantly higher than in the audio-only condition. (B) A histogram of the number of occurrences at which a certain condition yielded the best syllable recognition per participant showed that the audiotactile lag of 200 msec gave, for most participants, the highest score. *** = $p < .001$.

resulting in a duration of 800 msec between the first and the last pulse. In addition to this regular train of vibrotactile pulses, we created irregular pulse trains. The time between two successive pulses was thereby drawn from a uniform distribution ranging from 50 msec to 300 msec. The total duration of the obtained irregular pulse trains ranged from 700 to 900 msec. The onset of the syllable was delivered 1000 msec after the first vibrotactile pulse and thus between 100 msec and 300 msec after the last vibrotactile pulse.

The temporal waveform $\psi(t)$ of the individual pulses was designed as real Morlet wavelets:

$$\psi(t) = \psi_0 sin(2\pi f t)e^{\frac{-t^2}{2\sigma^2}} \qquad (1)$$

in which t denotes time. The amplitude was set at $\psi_0 = 1.4$ V, the carrier frequency f at 80 Hz, and the width σ at 7.5 msec. We refer to this signal as the tactile stimuli or the vibrotactile pulses in the rest of the article.

Delays of the Sound Signal and Control Conditions

When presenting the syllable, we considered different time shifts between the latest vibrotactile pulse and the onset of the syllable, which we refer to as the audiotactile lags. We expected that the audiotactile integration would involve neural activity in the theta frequency range, in which the periods of oscillation range from 125 msec to 250 msec. Investigating a putative oscillatory component in the neural, and, respectively, in the behavioral responses to the syllables, therefore, required delays that would lie in the second period of the oscillations following the last vibrotactile pulse, but not later so that the oscillatory activity had not yet dissipated.

We accordingly chose lags of 150 msec, 200 msec, 250 msec, and 300 msec after the last vibrotactile pulse (Figure 1). This resulted in four audiotactile conditions. A fifth condition was obtained by starting with an irregular train of vibrotactile pulses followed by the presentation of a syllable that occurred randomly between 100 msec and

300 msec after the last pulse. Whether regular or not, conditions in which both tactile pulses and syllables are presented are referred to as audiotactile conditions. Finally, as a control, we used an audio-only condition where no vibrotactile pulses were presented. We therefore obtained six different conditions, one of which was irregular, to study if the simple addition of tactile stimulation, with no set rhythm, could elicit an effect. This was motivated by previous results that demonstrated a supra-additive effect on envelope tracking in the presence of mismatched tactile stimuli (Guilleminot & Reichenbach, 2022).

Data Collection

The data collection was carried out in a single session for each participant. In each trial, the participants were presented with a single syllable in one of the six conditions, that is, with prior tactile stimulation except in the audio-only condition. Before each trial, a window of text notified participants of whether the condition was going to be an audiotactile or the audio-only condition so they would not be surprised by the presence or absence of the vibrotactile pulses. This was done to avoid any effects of the participants' surprise on the EEG signal (Mars et al., 2008). Moreover, during the pilot study, participants tended to interpret the absence of tactile stimuli during the audio-only condition as a dysfunction, impeding their focus.

Each trial, audio or audiotactile, started with the presentation of speech-shaped noise for a random duration between 500 msec and 1 sec. This was done to get the participants used to the audio level at each trial as well as randomizing the time at which the syllables were presented after the start of the trial. Using a random duration avoided participants using the start of the noise presentation to infer the exact timing of the upcoming syllable. In the audiotactile conditions, the vibrotactile pulses were then sent, followed by the syllable. In the case of the

audio-only condition, no vibrotactile pulses were presented but noise was played for the same duration. Finally, after the syllable had been played, there was a 500-msec pause before the choice of two syllables was displayed on the monitor. Before this choice, the monitor continuously displayed a fixation cross that participants were asked to focus on.

Each unique combination of syllable (pa, ba, ka, ga, ta, da), condition (one audio-only, five audiotactile), and voice (female, male) was presented 7 times, resulting in 504 trials. During the whole duration of the experiment, we recorded the participant's EEG.

Preprocessing of EEG Recordings

We recorded the EEG data at a sampling rate of 1000 Hz using 64 electrodes in an extended 10–20 system referenced to the vertex (Cz). During the preparation of the EEG recordings for a particular participant, the maximal impedance was kept under 10 kΩ. The EEG data were then band-pass filtered between 0.1 Hz and 32 Hz (one pass, zero-phase, noncausal finite impulse response bandpass filter of order 33,000). For time–frequency analysis, the EEG data were resampled at 100 Hz.

Statistical Testing of the Behavioral Results

For the behavioral results on syllable discrimination, we first assessed if the results differed significantly between the six experimental conditions through a Friedman test. If an effect was detected, we then used multiple post hoc Wilcoxon tests between the audio-only condition that served as a control condition and all of the audiotactile conditions. The resulting p values were corrected for multiple comparisons using the conservative Bonferroni correction (Bland & Altman, 1995). We also computed the condition at which the maximum score was obtained for each participant. We then tested whether the obtained distribution could be obtained via a uniform distribution using a chi-squared test.

Statistical Tests for the EEG Data

To test an effect of the tactile stimulation on the neural response to the syllable, we investigated the EEG signal after the syllable had been presented. In particular, we chose a time window between 50 msec and 550 msec after the onset of the syllable. Typically, somatosensory evoked potentials range from 0 msec to 200 msec poststimuli (Nuwer, 1998; Desmedt & Robertson, 1977). Therefore, this choice of window allows to avoid late-evoked activity contamination from the tactile stimuli while still studying the entire neural response to the syllable.

In this time window, we studied the EEG power in different scalp areas and in different frequency bands. In particular, we focused on the power in the delta frequency band (1–4 Hz) and in the theta frequency band (4–8 Hz)

in the lateral temporal areas as they have been found to play important roles in speech processing (Etard & Reichenbach, 2019; Doelling et al., 2014; Poeppel, 2003). The theta band is especially relevant because of its role found in audiotactile integration of continuous speech (Guilleminot & Reichenbach, 2022). In addition, we considered the power in the alpha frequency band (8–12 Hz) in the frontal and central areas, as it accounts for top–down multisensory processes, that is, attentional modulation (Misselhorn et al., 2019).

The EEG signals in these scalp areas and frequency bands were compared between the audio-only control condition and each of the five audiotactile conditions. The comparison was obtained using a threshold-free cluster enhancement method using the Morlet transform (Roach & Mathalon, 2008; Maris & Oostenveld, 2007). For each combination of the five audiotactile conditions and the six space-frequency foci of interest, the clustering method yielded a collection of p values that were already corrected for multiple comparisons. Each point in each time–frequency domain therefore yields a p value. The p value indicates the chance that the point belongs to a cluster in the time–frequency domain that shows a statistically significant difference between an audiotactile condition and the audio-only condition. We selected the minimal p value for each of these 30 combinations, as this method indicates the best candidate and does not rely on the exact boundaries of the cluster, which are subject to uncertainty (Sassenhagen & Draschkow, 2019). We then corrected for multiple comparisons across the 30 combinations through the Bonferroni correction (Bland & Altman, 1995).

Although the threshold-free cluster enhancement method allows for interpretability of the obtained results, it is limited in terms of resolution and conclusions about the exact extent of latency, location, or frequency band (Sassenhagen & Draschkow, 2019; Mensen & Khatami, 2013; Smith & Nichols, 2009). However, the cluster enhancement method can reliably identify the frequencies at which the neural responses in the audiotactile condition and in the audio-only condition differ the most.

Damped Oscillator Model

We hypothesized that the tactile stimulation can reset the phase of intrinsic neural oscillations. After the last vibrotactile pulse, we assumed that the intrinsic oscillations would slowly dissipate, following damped oscillations (Figure 1B). We therefore modeled both the neural and the behavioral data obtained in the audiotactile conditions with different lags through damped oscillations:

$$A(t) = A_0 + A_1 cos(2\pi f t + \phi)e^{-\frac{t}{\tau}} \qquad (2)$$

in which t denotes the audiotactile lag. To reduce subject-specific differences, the data points were scaled so that they

would have a standard deviation of 1 and be centered around 0 for each participant, such that $A_0 = 0$. A_1 corresponds to the amplitude of the oscillation, whereas f represents the oscillation frequency and ϕ a phase shift. Finally, τ is the decay time of the oscillation.

This simple model is represented in Figure 1B, where neural oscillations are being reset by tactile stimulation. In this representation, the oscillation frequency f can be interpreted as an intrinsic rhythm, specific to speech segmentation, that we approximate as being the same for all participants. The decay time τ can be interpreted as the dissipation of the oscillations as time passes by with no stimulation. Our experiment assumes that the decay time is long enough that some oscillatory activity can be observed past one period. Finally, the phase shift ϕ corresponds to the phase of the oscillations following the reset. For example, Figure 1B represents a low negative phase shift $\phi \in [-\pi/2, 0]$, where the reset opens the window of temporal integration: We expect integration to be optimal shortly following the reset. For the frequency f, we considered the values that the cluster enhancement method identified as being most informative about the differences in the EEG signals between the audiotactile conditions and the audio-only condition. We made this choice because our analysis showed these frequencies to be related to audiotactile integration in the brain. We obtained 64 data points (4 audiotactile lags × 16 participants) for the behavioral measure and 60 data points (4 audiotactile lags × 15 participants) for the electrophysiological one. The fitting procedure was done 2 times, as per previously presented (Guilleminot & Reichenbach, 2022; Kadir, Kaza, Weissbart, & Reichenbach, 2020; Keshavarzi et al., 2020). First, we fitted the amplitude, phase, and decay time through least square optimization using the trust region reflective method and a linear loss. The parameters were fitted with the following bounds: $A_1 \in [0, 100]$, $\phi \in [-\pi, \pi]$, $\tau \in [0, 1000]$ msec. We later used the values of ϕ and τ found this way.

Finally, the statistical significance of the obtained fit was assessed through Hubert robust linear regression, as it accounts for possible outliers, using A_1 as the regression variable and a significance criterion of $\alpha = .05$ (Rousseeuw & Hubert, 2011).

Correlation between Electrophysiology and Behavior

Finally, we wanted to relate the behavioral results to the electrophysiological one, more specifically capture whether there was a common within-individual effect. To that end, we used repeated-measures correlation (Bakdash & Marusich, 2017). We therefore considered each of the different audiotactile condition, including the irregular one, as a measure for each of participants, resulting in 75 data points (15 participants × 5 audiotactile conditions).

RESULTS

Behavioral Assessment

We first determined the average syllable discrimination score across all participants and all conditions. On average, participants recognized 70 ± 1% of the syllables correctly (mean and standard error of the mean). We then studied the variation of this score between the different conditions. A Friedman test revealed a significant difference between the scores obtained in the various experimental conditions ($p < .0005$). We therefore tested the syllable discrimination score in each audiotactile condition against the score obtained in the control audio-only control condition using post hoc pairwise tests. These tests revealed that the syllable discrimination score in audiotactile condition with an audiotactile lag of 200 msec was significantly higher than the score in the audio-only condition ($p = .0006$, Bonferroni correction for multiple comparisons; Figure 2A). On average, participants scored 68 ± 1% in the control condition, whereas they achieved an average score of 74 ± 1% in the audiotactile condition with an audiotactile lag of 200 msec. No significant difference emerged for any of the other comparisons ($p > .1$). To further analyze the difference between the scores obtained with the audiotactile lag of 200 msec and the ones in the audio-only condition, we computed the difference between these scores for each syllable separately and compared them. A Friedman test did, however, not find a significant variation in the score differences between the different syllables ($p = .49$). Finally, as an additional measure of audiotactile integration, we computed the audiotactile condition for each participant at which they obtained the best score. The obtained distribution was significantly non-uniform ($p = .001$; Figure 2B). The condition that yielded the highest syllable recognition score for most participants was the one with the audiotactile lag of 200 msec.

Electrophysiological Responses

To study the neural responses to the sensory stimuli, we first investigated the responses to the five vibrotactile pulses. We obtained five consecutive peaks, spaced 200 msec apart (Figure 3A). Each peak occurred at a delay of 48 msec after the corresponding pulse. The topographic plots showed a response in the left hemisphere, with a dipole bipolar pattern that suggested a source in the left somatosensory cortex. This spatiotemporal pattern was confirmed in the average response across the five pulses (Figure 3B). To assess the amplitudes of the five peaks, we computed the global field power (GFP) of the EEG signal in a ROI (FC1, FC3, F3, F1, CP5, CP3, P5, P3; Figure 3C). Finally, we assessed the EEG response to the syllable in the audio-only condition (Figure 3D). We found two peaks, at delays of 93 msec and 207 msec after the onset of the syllable. The first peak showed a larger amplitude with a symmetric
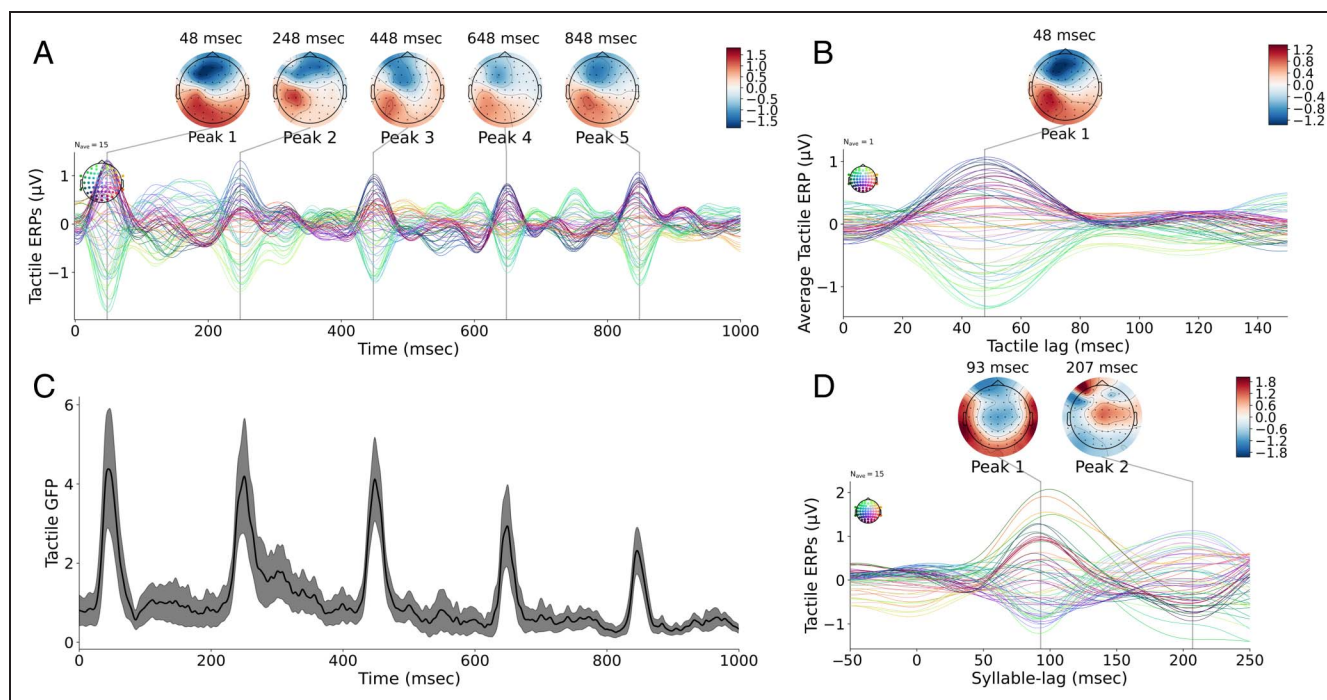
Figure 3. EEG responses to the audiotactile stimulation. (A) The EEG responses evoked by the vibrotactile pulses suggest a source in the left somatosensory cortex (average over all audiotactile conditions, time is relative to the first pulse). (B) The EEG signal in a response to a single vibrotactile pulse. This was obtained by averaging the individual responses to each of the five pulses together and showed a clear left-lateralized response in the somatosensory area. (C) The global field power (GFP), that is, the average power in the EEG signal across all electrodes, evidenced a decay in the amplitude between the first and the last pulse (black line, mean; black shading, standard error of the mean). (D) The EEG signal in response to the syllable in the audio-only condition, with time measured relative to the syllable onset, showed two peaks, a first at 93 msec and a second at 207 msec.

topography whereas the second peak displayed a left-lateralized topography.

Audiotactile Gain in the EEG Response

To assess whether there was an audiotactile gain in the EEG response, we considered the EEG signal from before the onset of the syllable to a certain period afterward. We computed the space–time–frequency response of each audiotactile condition, and subtracted the response of the audio-only condition. This yielded an estimate of the audiotactile gain provided by the prior presentation of the vibrotactile pulses.

We focused on specific ROIs in the channel-frequency space (Figure 4A). First, we were interested in the delta and theta frequency band in the left and right lateral temporal areas. Second, we also investigated the responses in the alpha band in the frontal and central areas.

We found only one space–time–frequency cluster at which a statistically significant audiotactile gain emerged. This cluster was located on the right lateral area, contained frequencies in the theta range, and occurred at a delay of 200 msec after the syllable onset ($p = .036$, Bonferroni correction for multiple comparisons; Figure 4B, C). The cluster corresponding to this particular set of frequency (theta range), condition (audiotactile with a latency of 200 msec), and location (right lateral are) was represented

in Figure 4. No other cluster proved statistically significant after multiple test correction ($p > .13$).

The center frequency of the cluster at which the audiotactile gain emerged was 6 Hz. We computed the GFP at this frequency in each audiotactile condition and subtracted the corresponding power in the audio-only condition (Figure 5D). Because the audiotactile gain occurred around 200 msec after the syllable onset, we extracted the power difference at this latency. Because this quantity represents a marker of audiotactile integration, we refer to it as the *neural audiotactile gain*.

Time Dependency of the Syllable Recognition Score and the Neural Audiotactile Gain

Having identified a behavioral quantity that informs on the audiotactile integration, the syllable recognition score, and an electrophysiological one, the neural audiotactile gain, we wondered how both depended on the delay at which the syllable occurred after the last vibrotactile pulse. Because we hypothesized that speech processing employs neural oscillations, the phase of which is reset by the tactile stimuli, we assumed that these oscillations would disperse and hence decay in amplitude after the last vibrotactile pulse.

We therefore described the time dependence of both the behavioral as well as of the electrophysiological

Figure 4. EEG markers of audiotactile integration. (A) A map of the different scalp areas and associated frequency bands for which we assessed potential differences in the neural response in the audiotactile conditions to that in the audio-only condition. (B) A threshold-free cluster enhancement method applied to EEG data in the theta band and in the right temporal area as a function of time and frequency showed a significant signal around a lag of 200 msec and around a frequency of 6 Hz. The lag was measured with respect to the onset of the syllable. Insignificant *p* values are shown in black. (C) Results from the threshold-free cluster enhancement for the theta-band in the right temporal area as a function of EEG channels at the maximum given by the previous graph (B). (D) The power in the EEG signals at

6 Hz in the right temporal area as a function of the postsyllable lag for different audiotactile lags. A maximum occurred at a postsyllable lag of 200 msec. The value of the EEG power at this lag, indicated by the colored disks, was subsequently used as an electrophysiological marker of audiotactile integration.
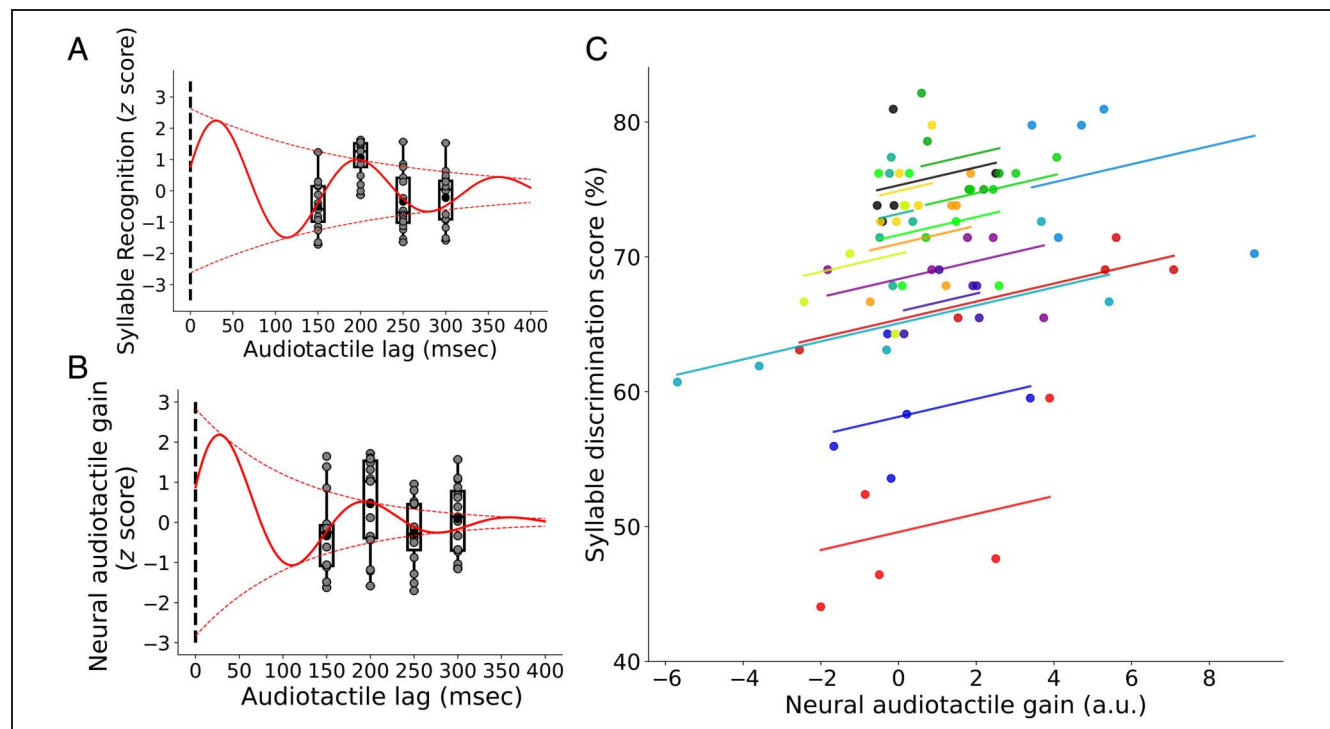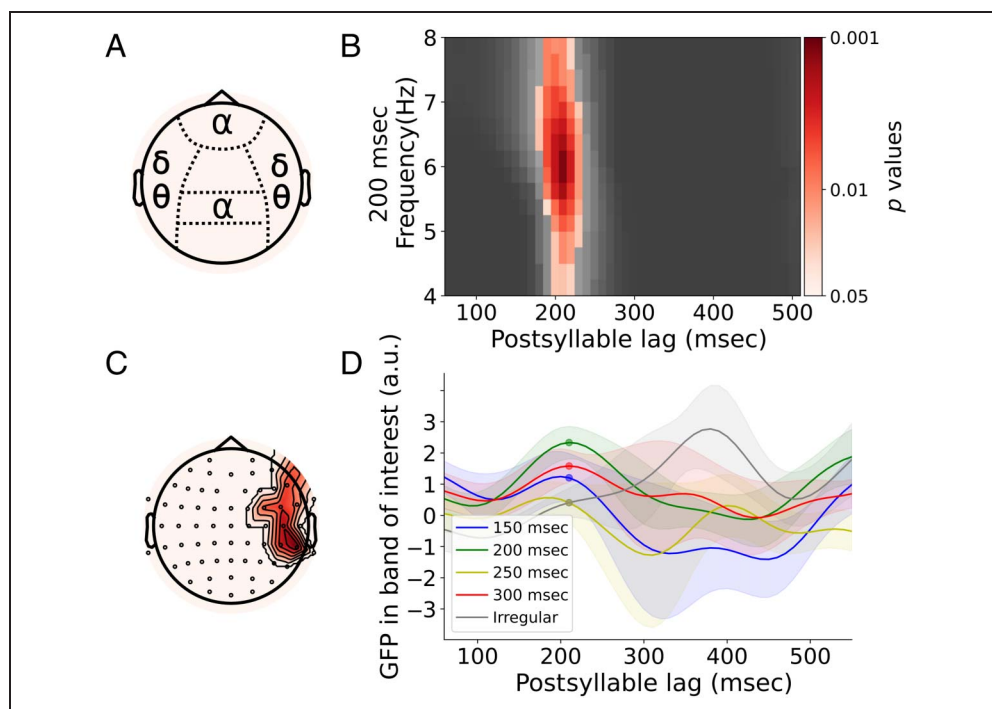

Figure 5. Dependence of behavioral and electrophysiological results on the audiotactile lag and on each other. (A) The dependence of the syllable discrimination scores (box plots, disks show results from individual participants) on the audiotactile lag could be modeled as a damped oscillation (red line). (B) Similarly, the dependence of the electrophysiological marker of audiotactile integration, the neural audiotactile gain (box plots, disks show results from individual participants), on the audiotactile lag could be fitted by a damped oscillation (red line). (C) The behavioral and the neural measure were correlated with each other across the different participants and experimental conditions (red line). The average value for each audiotactile lag is shown as a colored disk.

quantity through a damped oscillation. We chose an oscillation at a frequency of 6 Hz, the one at which the neural response was observed, to display audiotactile gain. The other parameters of the damped oscillations were fitted to the data. We obtained a significant fit with a phase shift of -1.27 rad and a decay time of 166 msec ($p < 1e - 6$, $R^2 = .358$; Figure 5A). The first and second peaks of the fitted curve occur at 31 msec and 197 msec after the last vibrotactile pulse, respectively.

We analogously fitted the time dependency of the neural audiotactile gain through a dampened oscillation with a frequency of 6 Hz. We obtained a phase shift of -1.26 rad, and a decay time of 116 msec ($p = .027$, $R^2 = .078$; Figure 5B). The first and second peaks of the obtained curve emerged at delays of 27 msec and 193 msec after the last pulse, very similar to those obtained for the syllable recognition score.

Correlation between Behavior and Electrophysiology

Because of the similar time dependencies of the behavioral and the neural measure, we wondered if both measures were correlated. We found that the quantities displayed a significant correlation ($r = .33$, $p = .0085$, $power = 0.75$; Figure 5C).

DISCUSSION

We have shown that rhythmic tactile stimulation can significantly improve the recognition of a subsequently presented syllable. The improvement occurred when the syllable started 200 msec after the last vibrotactile pulse. Moreover, we identified an electrophysiological marker of the audiotactile integration, namely, the power in the theta frequency band, around 6 Hz, in the right lateral temporal channels and at a delay of 200 msec after the onset of the syllable.

We could furthermore model the time dependency of the syllable recognition score as well as of the neural marker of audiotactile integration through a damped oscillation. Both model fits showed a similar time course, and we found indeed that the behavioral and the neural measure were directly correlated with each other. Given the absence of an auditory rhythm by speech, the oscillatory activity in response to the tactile rhythm can only be attributed to intrinsic brain oscillations. These results thus help in disentangling the intrinsic brain oscillatory activity from the regularity of speech stimuli.

Regarding this intrinsic rhythm, our working hypothesis was that the brain acted similarly to an oscillator with a characteristic frequency. Therefore, although we used a tactile stimulation at 5 Hz, any frequency close to the characteristic frequency of the oscillator it should in principle also entrain the network.

Syllable Discrimination Score

The syllable discrimination score in the audiotactile conditions was statistically different from that in the audio-only condition at a single audiotactile lag, namely, 200 msec. At this delay, the tactile stimulation resulted in a significant improvement in the syllable recognition.

This finding contrasts with previous studies on audiotactile integration regarding syllable recognition, where the effect was found to be restricted to a 50-msec temporal window when the tactile stimulation preceded the audio signal (Gick, Ikegami, & Derrick, 2010). However, it should be pointed out that this experiment focused on tactile stimuli through a puff of air and how it influenced the perception of syllables with different timing of aspiration ("ba" vs. "pa"). In our experiment, we used vibrotactile pulses that did not carry such information and therefore focused on modulating a rhythm of time windows at which speech processing was more or less effective.

It should also be noted that in related previous articles, the puff of air had an asymmetrical effect on the discrimination between two similar syllables (Gick et al., 2010; Gick & Derrick, 2009). In our case, we did not observe that the recognition of syllables differed between the different considered syllables, in either of the experimental conditions. Moreover, no condition significantly decreased the syllable discrimination scores. The lack of an adverse effect of the tactile stimulation may aid the application of tactile information in multisensory auditory prosthetics.

Because the delay of 150 msec between the onset of the syllable and the last vibrotactile pulse did not yield a behavioral effect, it appears that the tactile stimulation indeed works in an oscillatory manner, setting up temporal windows where behavior is improved and others where it remains unaffected. A full description of these windows will, in the future, require the consideration of additional audiotactile lags, both shorter and longer than the ones we have investigated here.

EEG Responses

To characterize the stimuli we used, we computed the EEG responses to the tactile stimuli as well as to the syllable. The response to the vibrotactile pulses agreed with that found in previous work using the same pulse waveform for somatosensory stimulation, with a strong early response at 48 msec (Guilleminot & Reichenbach, 2022). For repeated vibrotactile events, the EEG response decreased in amplitude, which might be caused by short-term synaptic depression (Benita, Guillamon, Deco, & Sanchez-Vives, 2012). Similarly to previous work on tactile ERPs, we found that the somatosensory-evoked EEG response occurred on the left hemisphere, contralaterally to the location of the right hand to which the tactile stimuli were presented (Yang, Jin, Lee, Jeong, & An, 2018).

The EEG response to the syllable presentation exhibited a strong early peak at a delay of 93 msec and a later one at a delay of 207 msec with respect to the onset of the syllable. Given its latency, we can associate the first peak to early acoustic representation (Salmelin, 2007). The topography of the first peak was symmetrical between the left and right hemispheres and suggested an equal contribution from both parts, which is coherent with the asymmetric sampling in time hypothesis (Poeppel, 2003). Given its latency, the second peak can in principle either correspond to semantic or phonological information (Salmelin, 2007). However, because of the absence of semantic content of the stimuli and the noise-induced latency we expect to observe in auditory ERPs (Bidelman, Bush, & Boudreaux, 2020), this second peak presumably reflects the phonological information in the stimulus. Moreover, its topography was in contrast to that of the first peak, asymmetrical, which is still in line with the asymmetric sampling in time hypothesis, and indicated more right-lateralized activity. This experiment did not measure speech comprehension per se, but rather syllable discrimination. Such a syllable needs, to be properly processed, to first be parsed properly. In natural speech, this parsing process relies on the theta rhythm (Giraud & Poeppel, 2012), which is more prominent in the right auditory cortex (Morillon, Liégeois-Chauvel, Arnal, Bénar, & Giraud, 2012; Poeppel, 2003). Although here we have only used single isolated syllables, the brain may still carry out a similar parsing process. It should be noted that an alternative explanation for the lateralization of the EEG response lies in the stimuli we were using. Indeed, the syllables we were using only differed by whether their consonants were voiced or unvoiced. This ability to discriminate voicing-cue is typically attributed to the right hemisphere (Simos, Molfese, & Brenden, 1997).

Neural Marker of Audiotactile Integration

Using a cluster method, we uncovered a neural marker of audiotactile integration that was also correlated to the behavioral measure. Clustering methods allow for powerful analysis but the interpretation of the obtained results is not straightforward (Sassenhagen & Draschkow, 2019). Indeed, a lot of variability has been observed in the boundaries of clusters, making exact inferences about the" shape" (location, latency, and frequency range) of the cluster unreliable. Therefore, instead of computing all clusters over the whole sets of parameters, we restricted the analysis to certain EEG channels, frequency bands, and time windows from certain ROIs. By doing so, our conclusions do not rely on the exact boundaries of the output cluster but rather on predefined sets established by our working hypotheses. Once a cluster has been established in one of those predefined space–time–frequency regions, we simply reduced the cluster to a maximum point, thus getting rid of the notion of cluster boundaries of the cluster and reducing the result space to a single point for later

analysis. As a downside of this approach, we had to perform a relatively high number of comparisons to explore several hypotheses, requiring adequate correction for multiple comparisons.

We nonetheless obtained a significant response for audiotactile integration that was located in the right lateral area and contained the power in the theta band. This finding accords with previous work on the theta population in the right auditory cortex, confirming that our task of syllable recognition relied on syllable tracking and that tactile stimulation did indeed affect this process (Poeppel, 2003). None of the other hypotheses yielded significant results. In particular, the absence of effect of the delta band are coherent with the roles typically associated with it, as no element of syntax or semantic was involved. Moreover, it has been observed that motor delta activity played a role in speech perception (Morillon, Arnal, Schroeder, & Keitel, 2019). The absence of effect therefore hints at the fact that either this process does not come into play for syllable discrimination or the motor activity in the delta band has not been modulated by our vibrotactile pulses. As for the alpha band, the absence of audiotactile response would indicate that the behavioral results we have observed was not because of an attentional modulation effect, hinting instead at a purely bottom–up effect.

The neural marker of audiotactile integration occurred at a delay of 200 msec after the syllable onset, which coincides with the second peak of the auditory ERP, thus indicating a direct effect on phonetic information processing. We also note that only one audiotactile condition, the one with a delay of 200 msec, yielded a significant neural gain. This delay is in line with our hypothesis that the vibrotactile pulses reset the phase of ongoing theta oscillations, in a similar way as tACS may entrain endogenous brain oscillations (van Bree et al., 2021). The audiotactile gain in theta activity therefore depended on the alignment between the tactile and the audio signals, as we observed before in related work (Guilleminot & Reichenbach, 2022). This finding might reflect the phenomenon of supra-additivity, which may occur when theta oscillations become entrained by the tactile stimulation and are in phase with the syllable (Stevenson et al., 2014). Moreover, it should be noted that no multisensory activity has been detected in the case of the lowest audiotactile lag, thus discarding the possibility that the observed effect is because of contamination by the somatosensory ERP. In addition, although the GFP of the irregular condition (Figure 4D) presents a later maximum at 400 msec, which was not present for the other condition, the negative results from the clustering prevents us from commenting more on this result.

Damped Oscillation

We were able to describe the time dependency of the behavioral measure and the neural marker of audiotactile integration through damped oscillations. This result hints at the tactile phase reset we hypothesized. Moreover, both

modeled dependencies exhibited similar temporal dynamics. They both presented a phase shift of about -1.16 rad, that is, with a first maximum occurring shortly after the last vibrotactile pulse. This suggests that the largest enhancement of syllable recognition would occur if the syllable started just slightly after the last vibrotactile pulse, in accordance with past work on the participant (Guilleminot & Reichenbach, 2022). This suggests that the window of enhanced syllable processing does not coincide exactly with the tactile event. Rather, the vibrotactile pulse may open up a series of windows of attention for upcoming events, defined by a theta oscillation. In addition, the decay times for both behavioral and electrophysiological model lied were on the same scale as a single oscillation. This hints at an actually observable dissipation, as opposed to a much longer decay time, that would hint at nondampened oscillations.

In addition to the similarity of the obtained model fits, we found that the syllable discrimination scores were indeed correlated to the electrophysiological marker. This reinforces the growing body of evidence that tactile stimuli may modulate speech comprehension through phase reset of endogenous sustained oscillations (van Bree et al., 2021).

Conclusion

Our previous work on audiotactile speech revealed that comprehension of ongoing speech could be modulated through simultaneous tactile signals linked to the syllable rhythm. The modulation depended in a sinusoidal manner on the time difference between the vibrotactile pulses and the perceptual centers of the syllables (Guilleminot & Reichenbach, 2022). However, it was unclear whether the effect of the tactile stimuli was caused by a reset of ongoing neural oscillations or simply reflected the rhythm set by the successive syllables in the speech stream. Here, we limited the task to syllable discrimination rather than speech comprehension. Therefore, there was a single syllable and therefore no rhythm established by preceding speech, thus isolating the effect of the tactile stimulation. We demonstrated that establishing a tactile rhythm before presenting a single syllable affected the recognition of that syllable, in a way that could be described by a damped oscillation. A neural marker of the audiotactile integration was found in the theta frequency range, linked to the rate of syllables in continuous speech, and displayed a similar temporal behavior. Together, these results highlight the importance of theta rhythm in syllable perception and are consistent with the hypothesis that tactile signals aid comprehension through resetting the phase of ongoing cortical oscillations in the theta range.

Data and Code Availability

The data are accessible from a public database (https://zenodo.org/record/7544913; Guilleminot, 2023), and code is accessible on Github (https://github.com/phg17/ATacSy).

Author Contributions

Pierre Guilleminot: Conceptualization; Data curation; Investigation; Methodology; Software; Validation; Writing—Original draft. Cosima Graef: Data curation; Investigation; Methodology; Software. Emilia Butters: Investigation; Methodology; Software. Tobias Reichenbach: Conceptualization; Funding acquisition; Methodology; Project administration; Writing—Review & editing.

Funding Information

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456, PubMed: 28439244

Benita, J. M., Guillamon, A., Deco, G., & Sanchez-Vives, M. V. (2012). Synaptic depression and slow oscillatory activity in a biophysical network model of the cerebral cortex. *Frontiers in Computational Neuroscience*, *6*, 64. https://doi.org/10.3389/fncom.2012.00064, PubMed: 22973221

Bidelman, G. M., Bush, L. C., & Boudreaux, A. M. (2020). Effects of noise on the behavioral and neural categorization of speech. *Frontiers in Neuroscience*, *14*, 153. https://doi.org/10.3389/fnins.2020.00153, PubMed: 32180700

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, *310*, 170. https://doi.org/10.1136/bmj.310.6973.170, PubMed: 7833759

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1001–1017. https://doi.org/10.1121/1.383319, PubMed: 512211

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, 5, 341–347.

Cieśla, K., Wolak, T., Lorens, A., Mentzel, M., Skarżyński, H., & Amedi, A. (2022). Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding. *Scientific Reports*, 12, 3206. https://doi.org/10.1038/s41598-022-06855-8, PubMed: 35217676

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, 35, 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015, PubMed: 26490860

Crosse, M. J., Di Liberto, G. M., & Lalor, E. C. (2016). Eye can hear clearly now: Inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *Journal of Neuroscience*, 36, 9888–9895. https://doi.org/10.1523/JNEUROSCI.1396-16.2016, PubMed: 27656026

Desmedt, J. E., & Robertson, D. (1977). Differential enhancement of early and late components of the cerebral somatosensory evoked potentials during forced-paced cognitive tasks in man. *Journal of Physiology*, 271, 761–782. https://doi.org/10.1113/jphysiol.1977.sp012025, PubMed: 926022

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030, PubMed: 26412129

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, 8, 311. https://doi.org/10.3389/fnhum.2014.00311, PubMed: 24904354

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19, 158–164. https://doi.org/10.1038/nn.4186, PubMed: 26642090

Doelling, K. B., Arnal, 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, PubMed: 23791839

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, 39, 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019, PubMed: 31109963

Fletcher, M. D., Mills, S. R., & Goehring, T. (2018). Vibro-tactile enhancement of speech intelligibility in multi-talker noise for simulated cochlear implant listening. *Trends in Hearing*, 22, 2331216518797838. https://doi.org/10.1177/2331216518797838, PubMed: 30222089

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Research. Cognitive Brain Research*, 10, 77–83. https://doi.org/10.1016/s0926-6410(00)00024-0

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: An fMRI study. *Journal of Neurophysiology*, 88, 540–543. https://doi.org/10.1152/jn.2002.88.1.540, PubMed: 12091578

Fu, K.-M., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., et al. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*, 23, 7510–7515. https://doi.org/10.1523/JNEUROSCI.23-20-07510.2003, PubMed: 12930789

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504. https://doi.org/10.1038/nature08572, PubMed: 19940925

Gick, B., Ikegami, Y., & Derrick, D. (2010). The temporal window of audio-tactile integration in speech perception. *Journal of the Acoustical Society of America*, 128, EL342–EL346. https://doi.org/10.1121/1.3505759, PubMed: 21110549

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517. https://doi.org/10.1038/nn.3063, PubMed: 22426255

Guilleminot, P. (2023). *Audio-tactile syllables EEG dataset [data set]*. Zenodo. https://doi.org/10.5281/zenodo.7544913

Guilleminot, P., & Reichenbach, T. (2022). Enhancement of speech-in-noise comprehension through vibrotactile stimulation at the syllabic rate. *Proceedings of the National Academy of Sciences, U.S.A.*, 119, e2117000119. https://doi.org/10.1073/pnas.2117000119, PubMed: 35312362

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. https://doi.org/10.1038/nrn2113, PubMed: 17431404

Kadir, S., Kaza, C., Weissbart, H., & Reichenbach, T. (2020). Modulation of speech-in-noise comprehension through transcranial current stimulation with the phase-shifted speech envelope. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28, 23–31. https://doi.org/10.1109/TNSRE.2019.2939671, PubMed: 31751277

Kayser, C. (2009). Phase resetting as a mechanism for supramodal attentional control. *Neuron*, 64, 300–302. https://doi.org/10.1016/j.neuron.2009.10.022, PubMed: 19914178

Keil, J., & Senkowski, D. (2018). Neural oscillations orchestrate multisensory processing. *Neuroscientist*, 24, 609–626. https://doi.org/10.1177/1073858418755352, PubMed: 29424265

Keshavarzi, M., Kegler, M., Kadir, S., & Reichenbach, T. (2020). Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. *Neuroimage*, 210, 116557. https://doi.org/10.1016/j.neuroimage.2020.116557, PubMed: 31968233

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53, 279–292. https://doi.org/10.1016/j.neuron.2006.12.011, PubMed: 17224408

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: Supramodal control of neurophysiological context by attention. *Neuron*, 64, 419–430. https://doi.org/10.1016/j.neuron.2009.10.014, PubMed: 19914189

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024, PubMed: 17517438

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., et al. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience*, 28, 12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008, PubMed: 19020046

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage*, 67, 111–118. https://doi.org/10.1016/j.neuroimage.2012.10.027, PubMed: 23123297

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage*, *79*, 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060, PubMed: 23624493

Misselhorn, J., Friese, U., & Engel, A. K. (2019). Frontal and parietal alpha oscillations reflect attentional modulation of cross-modal matching. *Scientific Reports*, *9*, 5030. https://doi.org/10.1038/s41598-019-41636-w, PubMed: 30903012

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, *3*, 248. https://doi.org/10.3389/fpsyg.2012.00248, PubMed: 22833730

Morillon, B., Arnal, L. H., Schroeder, C. E., & Keitel, A. (2019). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neuroscience and Biobehavioral Reviews*, *107*, 136–142. https://doi.org/10.1016/j.neubiorev.2019.09.012, PubMed: 31518638

Nuwer, M. R. (1998). Fundamentals of evoked potentials and common clinical applications today. *Electroencephalography and Clinical Neurophysiology*, *106*, 142–148. https://doi.org/10.1016/s0013-4694(97)00117-x, PubMed: 9741775

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*, 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Riecke, L., Snipes, S., van Bree, S., Kaas, A., & Hausfeld, L. (2019). Audio-tactile enhancement of cortical speech-envelope tracking. *Neuroimage*, *202*, 116134. https://doi.org/10.1016/j.neuroimage.2019.116134, PubMed: 31470124

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time–frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, *34*, 907–926. https://doi.org/10.1093/schbul/sbn093, PubMed: 18684772

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*, 73–79. https://doi.org/10.1002/widm.2

Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach. *Clinical Neurophysiology*, *118*, 237–254. https://doi.org/10.1016/j.clinph.2006.07.316, PubMed: 17008126

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335. https://doi.org/10.1111/psyp.13335, PubMed: 30657176

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, *85*, 1322–1327. https://doi.org/10.1152/jn.2001.85.3.1322, PubMed: 11248001

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*, 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1, PubMed: 12536133

Simos, P. G., Molfese, D. L., & Brenden, R. A. (1997). Behavioral and electrophysiological indices of voicing-cue discrimination: Laterality patterns and development. *Brain and Language*, *57*, 122–150. https://doi.org/10.1006/brln.1997.1836, PubMed: 9126410

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061, PubMed: 18501637

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., et al. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography*, *27*, 707–730. https://doi.org/10.1007/s10548-014-0365-7, PubMed: 24722880

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron*, *81*, 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044, PubMed: 24656248

van Bree, S., Sohoglu, E., Davis, M. H., & Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biology*, *19*, e3001142. https://doi.org/10.1371/journal.pbio.3001142, PubMed: 33635855

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience*, *32*, 155–166. https://doi.org/10.1162/jocn_a_01467, PubMed: 31479349

Yang, S. T., Jin, S. H., Lee, G., Jeong, S. Y., & An, J. (2018). Dominant and subdominant hand exhibit different cortical activation patterns during tactile stimulation: An fnirs study. In *2018 6th international conference on brain computer interface (BCI)* (pp. 1–3). IEEE. https://doi.org/10.1109/IWW-BCI.2018.8311502