Alpha-Band Activity During Sustained Speech Production and Playback: EEG Evidence for Continuous Monitoring

Human Neuroscience / Faculty of Medicine
Master's thesis

Author(s):
Fatemeh Naderipalangi

13.05.2026
Turku

Master's thesis
Subject: Human Neuroscience
Author(s): Fatemeh Naderipalangi
Title: Alpha-Band Activity During Sustained Speech Production and Playback: EEG Evidence for Continuous Monitoring
Supervisor(s): Assistant Professor, PhD. Henry Railo
Number of pages: 42 pages
Date: 13.05.2026

Speech production is unquestionably the most vital element in communication, and it depends heavily on a dynamic interaction between motor commands and sensory feedback. To produce speech, auditory feedback, in particular, plays an important role in detecting and refining mismatches between expected and actual vocal output through rapid corrections (Behroozmand & Larson, 2011).

A widely utilized approach for examining this feedback control system is the pitch-shift paradigm, wherein perturbations (unexpected changes, such as shifts in voice pitch during vocalization) in the auditory feedback signal elicit compensatory vocal and neural reactions (Behroozmand et al., 2011). This feedback control mechanism has been traditionally studied using Event-Related Potentials (ERPs), particularly N1 and P2 components (Y. Liu et al., 2018). However, these paradigms primarily capture transient, event-locked responses, leaving the neural dynamics underlying sustained, unperturbed speech production less well understood. In particular, frequency-specific neural activity during continuous speech has received comparatively limited attention.

In EEG data from previous experiments (Suchý et al., 2023), a recurring pattern of oscillatory activity was observed both before and after ERP responses, but only during vocalization, not during passive listening. The origin of this pattern remains unclear, and it is uncertain whether it reflects a functionally meaningful neural process or arises from methodological factors such as phase-locked averaging or the mixture of periodic and aperiodic components in the power spectrum. The dataset includes 64-channel EEG recordings of 20 participants producing sustained vocalizations of the vowel /a/ while receiving pitch-shifted auditory feedback through headphones. As the auditory stimuli are physically identical across conditions, the difference between self-produced and externally presented stimuli can be attributed to any neurological variations. This design enables a direct comparison between self-produced vocalization and passive listening under controlled conditions.

This study, therefore, focuses on alpha-band activity (7 – 12 Hz) during sustained speech production and playback. Spectral analysis using Fourier-based methods was employed to quantify spectral power within a post-onset time window. The results indicate that alpha-band activity is present in both conditions but does not differ significantly between vocalization and playback.

These findings suggest that alpha-band activity during sustained speech may not reflect a process uniquely associated with self-generated vocal monitoring. Instead, the previously observed rhythmic patterns may arise from more general neural dynamics or from the combined contribution of oscillatory and aperiodic components. Future work using approaches that separate these components may help clarify the functional role of frequency-specific activity in speech monitoring.

Keywords: auditory feedback, EEG, alpha-band activity, speech production, spectral power

Table of contents

1. Introduction

1.1 Background: Speech Production

1.1.1 Speech as a complex sensorimotor system

Speech production is a highly complicated sensorimotor act that involves coordinating muscles across the respiratory, laryngeal, and articulatory systems to produce sounds (Bradshaw et al., 2025). Precise changes in the activation of these muscles are required to produce more than 10 phonemes per second in normal conversation (Franken, Eisner, et al., 2018).

The complexity of the act of speech is further amplified by the precise orchestration of more than 100 muscles (Franken, Eisner, et al., 2018; Zhang et al., 2025a). These muscles must work in harmony to regulate respiratory airflow, the vibration of the vocal folds, and the specific positioning of articulatory structures (Franken, Eisner, et al., 2018). As a result, speech production is not just the sequential execution of motor commands, but a dynamic sensorimotor process that depends on continuous sensory monitoring and ongoing adjustment of vocal output (Franken, Eisner, et al., 2018; Houde & Nagarajan, 2011).

1.1.2 Speech as Goal-directed Behavior

Speech can be conceptualized as a goal-directed motor behavior because its primary objective is to obtain specific internal sensory targets that correspond to an intended message (Bradshaw et al., 2025; Hickok et al., 2011). Rather than being limited to the execution of motor sequences, the speech production system is specifically tuned toward generating a desired perceptual result (Bradshaw et al., 2025; Hickok et al., 2011; Weerathunge et al., 2022). This framework suggests that speech production is organized around achieving internally defined auditory and somatosensory goals rather than merely executing isolated motor movements (Bradshaw et al., 2025; Hickok et al., 2011; Smith et al., 2020; Weerathunge et al., 2022).

These sensory benchmarks are acquired through developmental experience and are archived as internal representations of specific speech sounds (Bradshaw et al., 2025; Hickok et al., 2011; Tourville & Guenther, 2011). During speech production, the brain identifies and triggers the appropriate motor programs necessary to hit these targets, ensuring that the resulting acoustic output remains both phonetically and perceptually precise (Bradshaw et al., 2025; Hickok et al., 2011). This target-driven approach underscores how expected sensory consequences govern motor behavior and highlights the fundamental role of internal models in speech production (Franken, Eisner, et al., 2018; Hickok et al., 2011).

1.1.3 Sensory Targets in Speech Production

The process of speech production is fundamentally guided by two primary types of sensory targets: auditory and somatosensory targets (Bradshaw et al., 2025; Smith et al., 2020). The auditory target represents the sound the speaker intends to produce, such as specific format frequencies, and the somatosensory target represents the intended physical configuration of the vocal tract (Smith et al., 2020). These targets are not static but are continuously refined through experience and learning (Bradshaw et al., 2025; Hickok et al., 2011; Tourville & Guenther, 2011). Motor commands are tuned to ensure the articulators move in a way that minimizes the error between these desired goals and actual sensory feedback generated during speaking (Hickok et al., 2011; Smith et al., 2020). The selection of motor programs for specific phonemes or syllables is driven by internal representation of speech sounds, and this neural process coordinates the articulatory system to produce acoustically precise output (Weerathunge et al., 2022).

1.1.4 Feedforward and Feedback Control Mechanisms

The regulation of speech production is dependent on the integration of top-down motor predictions, internally generated sensory targets that represent the expected auditory and somatosensory consequences of a specific movement, with bottom-up sensory input, the actual auditory and somatosensory feedback generated by the sound of our own voice and the physical sensation of our articulators during the act of speaking, from our own voices (Bradshaw et al., 2025).

This process is governed by two primary control systems: a feedforward control, which uses internally stored motor programs to drive the vocal tract, and a feedback control, which monitors auditory and somatosensory signals to detect and correct errors in real-time (Smith et al., 2020; Tang, 2024; Weerathunge et al., 2022).

These two mechanisms operate in parallel, with the feedforward control providing speed and efficiency, and the feedback control ensuring precision and adaptability (Houde & Nagarajan, 2011; Smith et al., 2020; Zhang et al., 2025a).

1.1.5 Prediction and Error Minimization

A fundamental feature of speech production is the use of predictive neural mechanisms that anticipate the sensory outcomes of vocal motor commands (Bradshaw et al., 2025; Zhang et al., 2025a). The central nervous system utilizes internal forward models to generate specific expectations regarding the auditory and somatosensory consequences of an intended movement (Franken, Eisner, et al., 2018; Suchý et al., 2023; Tourville & Guenther, 2011). This process allows the brain to compare internally predicted sensory consequences of speech movements, generated from copies of outgoing motor commands known as efference copies or corollary discharges, with the actual sensory feedback received during speaking (Aliu et al., 2009; Hickok et al., 2011; Houde & Nagarajan, 2011). The resulting discrepancy between the anticipated and actual feedback, often referred to as a prediction error, serves as a vital computational signal used to update motor commands instantly and refine the accuracy of future articulatory performance (Bradshaw et al., 2025; Hickok et al., 2011; Houde & Nagarajan, 2011; Tourville & Guenther, 2011).

This predictive framework allows the speech system to compensate for sensory delays and noise that would otherwise limit the speed and stability of vocal control (Hickok et al., 2011; Houde & Nagarajan, 2011; Parrell & Houde, 2019). By relying on internal predictions, the brain can detect and correct errors more rapidly than would be possible through purely reactive feedback mechanisms (Hickok et al., 2011; Houde & Nagarajan, 2011). Because purely reactive mechanisms are often too slow to maintain stability during the high-velocity, sequential movements required for speech, the brain relies on these internal predictions to detect and counteract errors far more rapidly than would be possible through generic feedback loops (Houde & Nagarajan, 2011). Over time, this persistent process of error minimization is what enables the maintenance of stable, fluent, and phonetically accurate speech production (Franken, Eisner, et al., 2018; Hickok et al., 2011; Houde & Nagarajan, 2011).

1.2 Auditory Feedback in Speech Production

1.2.1 The Role of Sensory Feedback in Speech

Producing accurate speech depends on more than the execution of motor programs; it requires sensory feedback to monitor whether vocal output matches the intended speech target (Behroozmand & Larson, 2011; Bradshaw et al., 2025; Zhang et al., 2025b). Auditory feedback is especially important because it provides information about the acoustic consequences of speech, including whether the produced sound corresponds to the intended vocal output (Bradshaw et al., 2025; Tang, 2024).

When a mismatch occurs between expected and actual auditory feedback, the speech system can initiate rapid corrective adjustments (Bradshaw et al., 2025; Tang, 2024). For example, unexpected shifts in auditory feedback can trigger compensatory responses within approximately 100 – 150 milliseconds, often without conscious awareness (Franken, Eisner, et al., 2018; Guenther, 2006). Thus, auditory feedback provides a key source of information for maintaining vocal accuracy during ongoing speech.

1.2.2 Feedforward and Feedback Integration

Speech production relies on the interaction between feedforward and feedback control systems (Scheerer et al., 2013). Feedforward mechanisms support the rapid and fluent execution of speech movements, whereas feedback mechanisms monitor auditory input to ensure that produced speech remains accurate and stable (Guenther, 2006; Hickok, 2012; Scheerer & Jones, 2018). When discrepancies between expected and actual sensory feedback are detected, corrective adjustments can be initiated to refine ongoing motor output (Bradshaw et al., 2025; Franken, Eisner, et al., 2018). Together, these complementary systems combine the speed of feedforward execution with the adaptive precision of feedback-based monitoring (Bouchard & Chang, 2014; Hickok et al., 2011).

1.2.3 Forward Models and Efference Copy

Forward prediction during speech production is supported by internal models that anticipate the sensory consequences of motor commands (Bradshaw et al., 2025). When speech movements are initiated, a copy of the outgoing motor command, known as an efference copy or corollary discharge, is transmitted to sensory regions of the brain (Aliu et al., 2009; Behroozmand & Larson, 2011; Hickok et al., 2011; Jenson et al., 2015). This mechanism enables the system to generate predictions about the expected auditory and somatosensory consequences of speech before the arrival of actual sensory feedback (Aliu et al., 2009; Behroozmand & Larson, 2011; Franken, Eisner, et al., 2018).

Predicted sensory outcomes are continuously compared with incoming feedback (Franken, Eisner, et al., 2018). If the predicted and actual signals match, the system confirms that speech production is proceeding as intended (Aliu et al., 2009). However, when discrepancies are detected, prediction error signals can initiate rapid corrective adjustments to ongoing motor output (Bradshaw et al., 2025; Franken, Acheson, et al., 2018; Jenson et al., 2015). This predictive framework allows speech production to operate more efficiently than purely reactive control systems by compensating for inherent sensory delays during vocalization (Hickok, 2012; Hickok et al., 2011; Scheerer et al., 2013).

1.2.4 Neural Circuits Supporting Auditory Feedback

Forward prediction during speech production relies on a distributed neural network connecting frontal motor regions with temporal and parietal sensory (Houde & Nagarajan, 2011; Tourville & Guenther, 2011). Premotor and posterior inferior frontal regions are involved in the planning and initiation of speech motor commands, whereas the primary motor cortex executes the physical movements required for speech production while simultaneously transmitting efference copies to sensory systems (Houde & Nagarajan, 2011; Parrell & Houde, 2019; Tourville & Guenther, 2011).

Auditory feedback is processed primarily within the temporal cortex, particularly the superior temporal gyrus, which contributes to the spectro-temporal analysis of incoming auditory signals (Behroozmand & Larson, 2011; Hawco et al., 2009; Jenson et al., 2015). In addition, sensorimotor integration regions such as area Spt, located near the temporal-parietal junction, facilitate the transformation between motor commands and sensory representations, enabling comparison between predicted and actual auditory feedback (Hickok, 2012; Hickok et al., 2011).

Cerebellar and sensorimotor forward modelling systems further contribute to the generation and refinement of predicted auditory and somatosensory outcomes, which are represented in higher-order sensory regions, including posterior temporal and supramarginal cortices (Guenther, 2006; Hickok, 2012; Hickok et al., 2011; Tourville & Guenther, 2011).

1.2.5 Continuous Monitoring and Error Detection

Auditory feedback functions as a continuous monitoring system that enables the brain to compare predicted and actual speech output during ongoing (Bradshaw et al., 2025; Hawco et al., 2009; Houde & Nagarajan, 2011; Zhang et al., 2025b). This process helps compensate for motor variability, sensory delays, and other sources of biological noise that could otherwise disrupt stable speech production (Hickok et al., 2011; Houde & Nagarajan, 2011; Parrell & Houde, 2019). Through the rapid detection and correction of mismatches between expected and actual auditory feedback, the speech system is able to maintain accurate and stable vocal output (Behroozmand & Larson, 2011; Bradshaw et al., 2025).

The dynamic interaction between prediction and sensory feedback raises important questions regarding how these processes are implemented neurally (Franken, Eisner, et al., 2018; Houde & Nagarajan, 2011; Zhang et al., 2025b). To investigate these mechanisms, researchers have increasingly relied on electroencephalography (EEG) and magnetoencephalography (MEG), which provide the temporal resolution necessary to capture rapid neural responses during speech production (Guo et al., 2021; Hawco et al., 2009; Jenson et al., 2015).

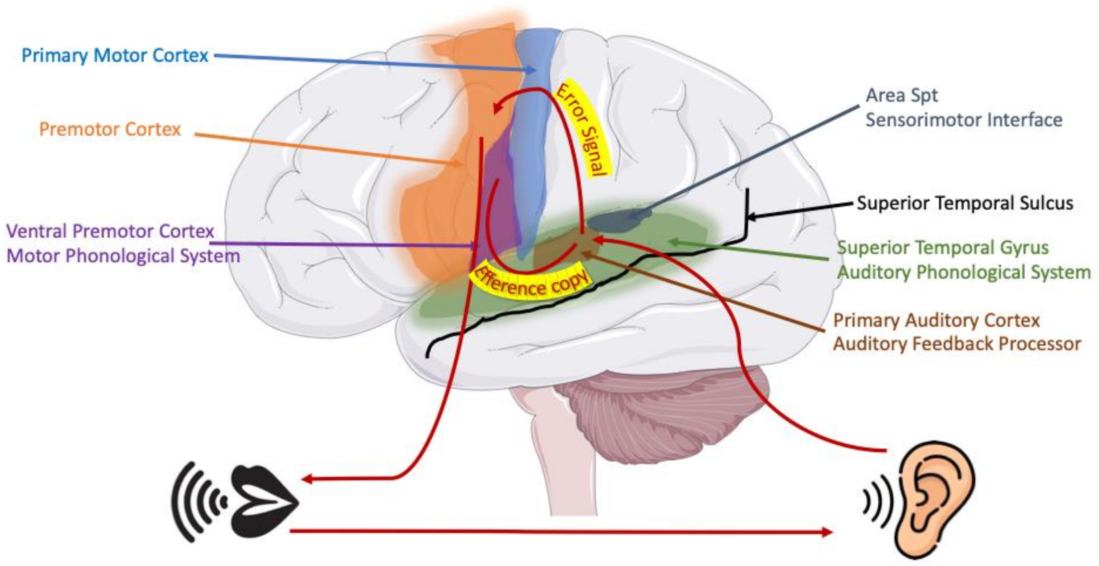


Figure 1: Speech production involves continuous comparison between predicted and actual sensory feedback (Bradshaw et al., 2025). Motor commands generate efference copies that are evaluated in auditory and sensorimotor regions (e.g., superior temporal gyrus, area Spt) (Aliu et al., 2009). Detected mismatches (prediction errors) rapidly drive corrective signals to maintain accurate and stable speech (~100–150 ms), forming an “always-on” monitoring system (Guenther, 2006; Hickok et al., 2011).

1.3 Speech Motor Control Models

1.3.1 Continuous Monitoring in Theoretical Models

Theoretical frameworks of speech motor control define monitoring as a persistent, automatic function rather than a reactive mechanism triggered only by external shifts (Franken, Eisner, et al., 2018; Hickok et al., 2011).

1.3.2 State Feedback Control (SFC) Model

In State Feedback Control (SFC) models, the brain maintains a dynamic internal estimate of the vocal tract's condition (Hickok et al., 2011). This model is constantly refined in near-real time by integrating forward predictions and incoming sensory feedback (Hickok et al., 2011). These systems are explicitly designed for continuous operation to counter the fundamental instabilities of the biological motor systems (Hickok et al., 2011).

1.3.3 Need for Continuous Control

Because sensory feedback is naturally delayed and noisy, and because articulatory precision is vulnerable to fatigue or physical drift, speech cannot be maintained through simple reactive corrections (Hickok et al., 2011; Houde & Nagarajan, 2011). Consequently, constant comparison between intended sensory goals and actual feedback is necessary to stop the gradual accumulation of error and deterioration of phonetic accuracy (Bradshaw et al., 2025; Hickok et al., 2011; Houde & Nagarajan, 2011).

1.3.4 DIVA Model: Tonic and Phasic Control

Neurophysiological models, such as the DIVA model, further distinguish between tonic control mechanisms that stabilize articulator positions and phasic error responses that signal the brain to deviations (Tourville & Guenther, 2011).

While experimental paradigms frequently isolate these phasic mismatches through artificial perturbations, the tonic neural dynamics required for stable, steady-state production remain less well-defined (Franken, Eisner, et al., 2018; Tourville & Guenther, 2011).

Researchers have noted that the emphasis on event-locked responses to transient triggers has left the continuous neural process of ongoing monitoring comparatively underexplored (Franken, Eisner, et al., 2018; Reznik & Mukamel, 2019; Tourville & Guenther, 2011). The functional roles of spectral activity within frequency bands such as theta (θ) and beta (β) during sustained speech and their contribution to regulating predictions and errors are yet to be studied systematically (Franken, Eisner, et al., 2018).

1.4 Electrophysiological Correlates of Auditory Feedback

1.4.1 EEG and MEG in Speech Monitoring

Speech monitoring has often been examined electrophysiologically using Magnetoencephalography (MEG) and Electroencephalography (EEG) measures of evoked responses and sensory suppression, which capture both transient and ongoing neural dynamics (Franken, Eisner, et al., 2018; Reznik & Mukamel, 2019). Event-related potentials (ERPs) and event-related fields (ERFs) provide millisecond resolution that allows researchers to track these fast cortical responses to prediction-error signals (Franken, Eisner, et al., 2018).

1.4.2 Speaking-Induced Suppression and Prediction

A well-established marker of speech monitoring is speaking-induced suppression (SIS) of early auditory components, specifically the EEG component N1 and its MEG counterpart M100, which are attenuated during speech production compared to passive listening (Houde & Nagarajan, 2011; Reznik & Mukamel, 2019).

This attenuation is interpreted as evidence of forward prediction, such that when the consequence of a motor command is accurately predicted by the corollary discharge, there is a dampening of auditory cortical responses (Franken, Eisner, et al., 2018; Reznik & Mukamel, 2019).

1.4.3 Perturbation Paradigms and Prediction Error

On the other hand, perturbations, artificial real-time alterations made to the auditory feedback that a person receives while they are actively speaking, such as shifts in pitch consistency, evoke larger N1 (in EEG) and M100 (in MEG) amplitudes than the suppressed response observed during normal, unperturbed speech (Aliu et al., 2009; Franken, Eisner, et al., 2018). Together, these findings have led to the hypothesis that speech is monitored via prediction mechanisms (Franken, Eisner, et al., 2018). Motor commands create efference copies that are compared with incoming sensory signals in the auditory cortex (Reznik & Mukamel, 2019).

When the predicted and actual feedback match, neural responses are suppressed; when they do not, a prediction error signal is generated, and compensation occurs (Franken, Eisner, et al., 2018).

1.4.4 Altered Auditory Feedback (AAF)

Most electrophysiological research on speech monitoring uses experimentally induced perturbations to briefly disrupt auditory feedback (Franken, Eisner, et al., 2018; Tang, 2024). Altered Auditory Feedback (AAF) paradigms, which typically shift pitch or formant frequencies, are designed to create specific mismatches between the brain's internal predictions and actual sensory outcomes, thereby eliciting quantifiable neural and motor corrections (Franken, Eisner, et al., 2018). These experiments generally use event-locked analysis, synchronizing neural data to the start of the perturbation and averaging across many trials to identify fast-acting, mismatch-related brain components (Franken, Eisner, et al., 2018; Max & Maffett, 2015).

1.4.5 Example Studies

An illustrative study by Franken et al. (2018) utilized magnetoencephalography (MEG) alongside real-time altered auditory feedback to investigate oscillatory dynamics during speech (Franken, Eisner, et al., 2018). In this experiment, participants produced continuous vowel sounds while listening to their own voices through headphones, and brief pitch perturbations were introduced into the auditory feedback (Franken, Eisner, et al., 2018). By analyzing the neural data through both time-frequency methods and event-related field (ERF) measures, the researchers were able to identify both brief, immediate responses and longer-lasting spectral power shifts that were precisely synchronized to the onset of the feedback errors (Franken, Eisner, et al., 2018). The study reported significant increases in theta-band activity (4 – 8 Hz) and modulation in lower beta power (12 – 16 Hz), which were localized to the sensorimotor and temporal regions of the brain (Franken, Eisner, et al., 2018). Crucially, these oscillatory responses were robustly present during active vocalization but were either significantly reduced or entirely absent when participants heard passive playback of the same acoustic recordings (Franken, Eisner, et al., 2018). These findings support the theory that such neural dynamics represent specialized speech-monitoring processes, specifically the active comparison between an internal forward model's prediction and actual sensory input, rather than a generic response to auditory stimulation (Franken, Eisner, et al., 2018).

Similar altered auditory feedback (AAF) paradigms have been frequently employed to explore the brain's speech-monitoring mechanisms (Tang, 2024). In these experiments, researchers artificially manipulate pitch or vowel resonance (formants) in real-time while a subject is speaking, enabling them to study how the brain identifies prediction errors and initiates compensatory motor movements (Bradshaw et al., 2025; Houde & Nagarajan, 2011; Smith et al., 2020; Tang, 2024). Such feedback shifts reliably trigger activity in a network of frontal and temporal sensorimotor areas, specifically the auditory cortex and ventral premotor regions, that are responsible for comparing internal predictions with actual sensory input (Franken, Eisner, et al., 2018; Houde & Nagarajan, 2011; Tourville & Guenther, 2011). Time-frequency analyses have indicated that neural activity within theta (4–8 Hz) and beta (12–20 Hz) frequency ranges may contribute to sensorimotor integration during speech for managing this sensorimotor integration (Franken, Eisner, et al., 2018). These neural dynamics are thought to facilitate active error monitoring, the real-time updating of motor programs, and the persistent quality control necessary to maintain phonetic stability and prevent speech from deteriorating over time (Franken, Eisner, et al., 2018; Hickok et al., 2011).

1.4.6 Limitations

While the event-locked analysis method efficiently identifies prediction error signals, it focuses primarily on sudden, external shifts rather than the mechanisms of steady, uninterrupted speech (Franken, Eisner, et al., 2018). Furthermore, even designs incorporating full sentences or longer-lasting changes remain highly controlled and do not fully mirror the complexity of natural conversation (Bradshaw et al., 2025).

Researchers have acknowledged that these experimental frameworks emphasize the detection of short-term errors, leaving the neural process of continuous monitoring during ongoing speech relatively under-investigated (Bradshaw et al., 2025).

1.5 Research Gap

Neural oscillations provide a unique perspective on speech monitoring, yet most research continues to rely on externally introduced perturbations to trigger measurable brain responses (Franken, Eisner, et al., 2018; Tang, 2024; Weerathunge et al., 2022). Although these paradigms have illuminated how the brain identifies and fixes prediction errors, they primarily index neural activity tied to discrete, transient mismatches rather than the continuous monitoring mechanisms utilized during stable speech production (Franken, Acheson, et al., 2018; Franken, Eisner, et al., 2018). In natural speech, the brain must perform persistent, real-time supervision to produce and regulate vocal output (Franken, Acheson, et al., 2018; Franken, Eisner, et al., 2018; Houde & Nagarajan, 2011). However, the neural dynamics that support this ongoing supervision, particularly during steady, unperturbed vocalization, remain largely underexplored (Franken, Eisner, et al., 2018). In particular, the extent to which frequent specific neural activity reflects continuous monitoring processes beyond transient, event-locked responses has not been systematically characterized.

Previous analyses of similar speech production data, (Suchý et al., 2023), revealed a prominent rhythmic pattern in the alpha frequency range (~10 Hz) during sustained vocalization. However, it remains unclear whether this activity reflects a functionally meaningful neural process related to speech monitoring or whether it may instead arise from methodological confounds. This ambiguity motivates further investigation into the role of alpha-band activity during continuous, unperturbed speech production.

1.6 Present study

To address this gap, the present study utilized electroencephalography (EEG) to analyze neural oscillations during continuous vocalization and passive listening. Participants completed two experimental conditions: an active vocalization state, where they produced sustained vowel sounds while hearing their voice through headphones, and a playback state, in which participants listened passively to recordings of the same auditory stimuli without producing speech. By ensuring the acoustic input was identical in both conditions, comparing neural responses between vocalization and playback allowed the isolation of neural processes specifically associated with self-generated speech monitoring rather than auditory perception alone (Aliu et al., 2009). In contrast to perturbation-driven paradigms that focus primarily on transient error responses, the present study examined spectral dynamics, capturing neural activity beyond the canonical event-related response window. The use of frequency analyses allowed for the characterization of neural dynamics beyond the standard event-related response window, providing an opportunity to investigate whether these persistent rhythmic patterns represent the brain's continuous monitoring process during speech production (Franken, Eisner, et al., 2018).

1.7 Research Questions

1. Is alpha-band activity present during sustained speech production in the absence of external perturbations?

2. Does alpha-band activity differ between active vocalization and passive listening when the auditory input is identical?

1.8 Hypotheses

Previous analyses of similar speech production data (Suchý et al., 2023) revealed prominent alpha-range activity during sustained vocalization, raising the possibility that alpha-band spectral power may be associated with processes involved in self-generated speech monitoring. In addition, previous studies have demonstrated differences in neural responses between active vocalization and passive listening conditions (H. Liu et al., 2010).

H1

Alpha-band spectral power will differ between active vocalization and passive playback conditions despite identical auditory input.

2. Methods

2.1 Participants

This study includes 20 healthy adult participants (Mean age: 24, range: 18 – 31, 16 females, 4 males) with normal hearing and without any diagnosed neurological or speech disorders. All participants provided written consent before the experiment. The study has received approval from the ethics committee of the Wellbeing Services County in Southwest Finland.

2.2 Experimental Design

This study employed a pitch-shift auditory paradigm to examine brain responses during both speech production and passive listening. Participants engage in two separate sessions on different days:

1. **Vocalization session (S1):** participants are asked to produce a continuous vocal sound of the vowel /a/ for 4 seconds per trial while hearing their voice in real time through headphones. Each trial is divided into two intervals, A and B, which are also displayed on the screen in front of the participant. In one of the intervals, chosen in a random order, with an equal number of trials assigned to each interval, a short upward pitch shift is introduced to the auditory feedback, while the other stays unaltered. The perturbation could occur in either the A or B interval of the trial. In A-interval trials, the pitch shift was introduced during the first designated interval, whereas in B-interval trials, it was introduced during the second designated interval. The onset timing within the selected interval was further varied using a randomized jitter of 0 to 0.4 seconds. Half of the trials contained perturbations in interval A and half in interval B, presented in randomized order. Each perturbation had a total duration of 0.22 seconds, a figure that accounts for the time needed to transition to and from the target frequency. This window consisted of a 0.01-second ramp up to the target pitch, a 0.2-second steady state at the target pitch, and a 0.01-second ramp down back to the participant's original voice pitch. Participants have to complete a two-interval forced-choice (2IFC) task to determine which interval contains the pitch shift, and subsequently rate how well they heard the perturbation using a three-point Perceptual Awareness Scale (PAS), rating "did not hear", "weakly", or "clearly". The pitch change includes two types: 1. "calibrated" trials, in which the magnitude of the change is based on each participant's perceptual

threshold using the QUEST staircase method, and 2. “preset” trials, which involve a +50-cent pitch shift.

2. **Playback session (S2):** The playback session followed the same overall experimental structure as the vocalization session (S1), with the key difference that participants did not produce any vocal output. Instead, they listened through headphones to recordings of their own sustained /a/ vocalizations that had been collected during S1. The same auditory feedback manipulations used in S1 were presented during playback. Each trial contained two visually indicated intervals (A and B), and a brief upward pitch perturbation was introduced in one interval while the other remained unaltered. The assignment of perturbations to intervals A and B followed the same randomized order as in S1. Perturbation timing, duration, and pitch-shift characteristics were identical to those used during the vocalization session. After each trial, participants completed the same two-interval forced-choice (2IFC) task and Perceptual Awareness Scale (PAS) rating procedure as in S1.

This setup facilitates a direct comparison between self-generated auditory feedback and externally provided sounds.

Before the main experimental blocks, the trials counted for the experiment were divided into two phases: practice trials and calibration trials. There were 30 practice trials in total. The first set included 10 trials with a 100-cent pitch shift (no pink noise). The second set consisted of 10 trials where the pitch shift was gradually lowered from 100 to 10 cents with feedback, followed by another 10 trials without feedback. A 52-trial calibration procedure was used to determine each participant’s individual perceptual threshold for pitch shifts using the QUEST algorithm. The main phase of each session consisted of 282 trials in total. The first 22 trials of the main phase were used to monitor the participant’s accuracy to ensure it remained between 55% and 85%; if it did not, the calibration was repeated. After these initial trials, the experiment continued with 5 blocks, each containing 52 trials, for a total of 260 additional trials. Within each 52-trial block, there were 42 calibrated trials, where the pitch shift magnitude was based on the individual’s threshold, and 10 preset trials, where the pitch shift was always a clearly audible +50 cents. Across all trials, the pitch shift was introduced in the interval A for half of the trials and in the interval B for the other half, with the order randomized. Participants completed these tasks across two sessions, which shared the same trial information.

The recording environment for this study was designed to carefully control auditory feedback and monitor neural activity during both active speech and passive listening. Participants vocalized into a head-mounted microphone (AKG C544 L) positioned 5 cm from the mouth at a 45-degree angle, connected to a Motu MicroBook IIc audio interface. They heard their voice in real time through sound-isolating in-ear earphones (Neuroscan 10 Ω). To minimize the impact of bone conduction and any sound leaking through the earphones, pink noise (set at 75 dB) was mixed into the audio feedback. The intensity of the voice feedback was calibrated to be 5 dB louder than the participant's actual voice intensity. The Sound Pressure Level (SPL) of the participants' voices was monitored to ensure the participants remained between 76 and 90 dB; if a participant deviated from this range for three consecutive trials, they were prompted to adjust their volume. Participants faced a computer screen that displayed the intervals for the task (labeled A and B). The screen also provided prompts, such as "Start!" to begin vocalizing and "Stop!" to end the trial.

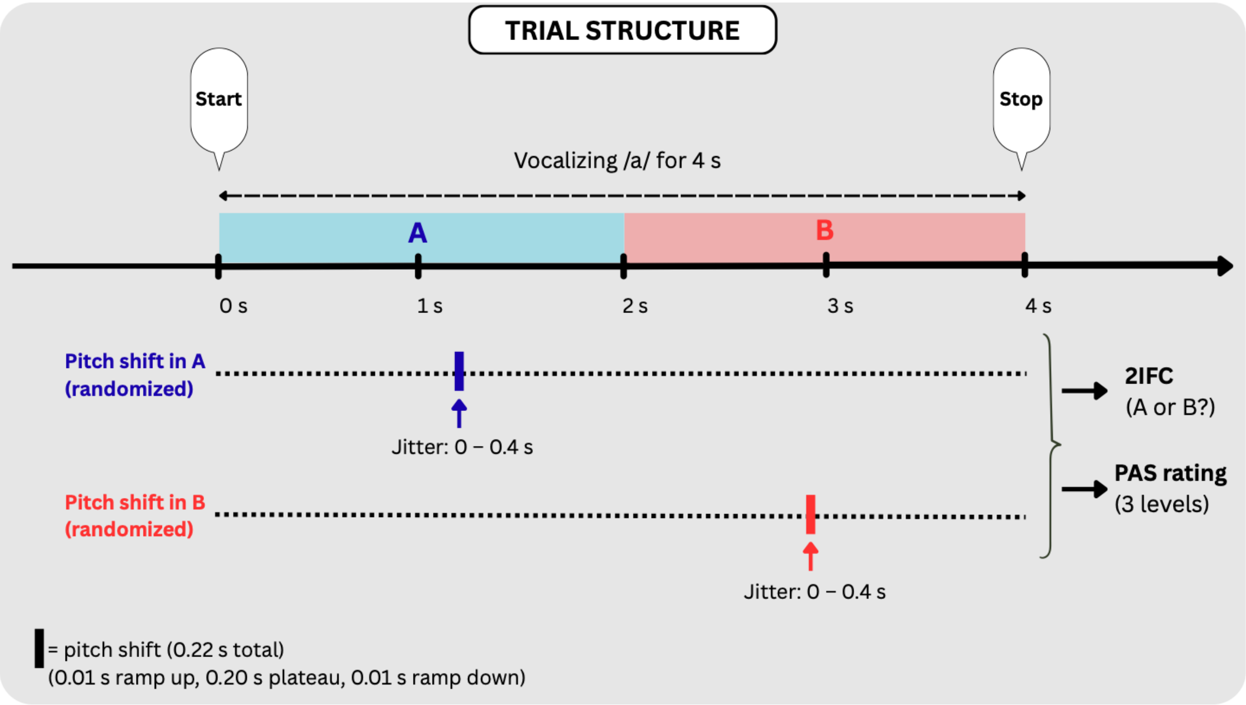


Figure 2: Trial structure of the pitch-shift paradigm. Each 4-s trial was divided into two intervals (A, B), with a brief pitch perturbation (0.22 s) occurring in one interval (randomized, jittered 0–0.4 s). Participants performed a 2IFC task and rated perceptual awareness (PAS) after each trial.

2.3 EEG Recording

Brain activity was monitored throughout both vocalization and playback sessions using a 66-channel EEG system, which incorporated electrooculography (EOG) channels to account for eye movements. The resulting data were resampled at 250 Hz and stored in EEGLAB format (consisting of .set and .fdt files). They included descriptive event markers for trial conditions, such as the interval of the pitch shift and the participant's level of awareness. The EEG signal was recorded using a NeuroOne Tesla amplifier (Bittium, 2022) running software version 1.5.0. An online passive reference electrode was used during the EEG recording and was positioned between the FC1 and FC2 electrodes of the standard 10 – 10 system. Additionally, an online passive ground electrode was placed between the AF3 and AF4 electrodes. The original sampling rate used to record the EEG signal was 500 Hz.

2.4 EEG Preprocessing

Preprocessing of the EEG data was conducted in MATLAB using EEGLAB. The preprocessing pipeline followed a sequence designed to remove noise and prepare the data for further analysis.

First, raw EEGLAB datasets were loaded and downsampled to 250 Hz. Channel locations were assigned using a custom electrode coordinate file (custom_channel_location.ced).

Next, noisy channels were identified using automated rejection procedures based on kurtosis and probability measures (threshold = 4 SD). Channels identified as noisy were removed and later restored using spherical interpolation to preserve the original channel montage. The data were then band-pass filtered between 0.1 and 40 Hz using an FIR filter and re-referenced to the common average reference.

Following these preprocessing steps, the data were segmented into epochs surrounding the events of interest and baseline corrected. Trials containing excessive artifacts were rejected using automated procedures.

To further separate neural signals from non-neural artifacts, Independent Component Analysis (ICA) was applied to the cleaned data. Independent components were classified using ICLabel, and components were retained only if they had a probability of >= 30% of being classified as "brain" or "other". Components identified as artifacts were removed before reconstructing the cleaned EEG data. The resulting dataset was saved for subsequent analyses.

2.5 Voice-onset ERP Analysis

To analyze the neural responses to speech production, Event-Related Potentials (ERPs) locked to the onset of the voice were calculated using datasets that had been previously refined through Independent Component Analysis (ICA) (*_ica_cont.set). The onset of vocalization was identified using event markers labelled in the format vocal_onset_bX_tY, where bX indicates the experimental block number, and tY indicates the trial number corresponding to each vocalization event. Data segments (epochs) were then extracted, spanning from 500 ms before speech started to 3000 ms after, relative to vocal onset and baseline-corrected using the -500 to 0 ms interval.

To create individual profiles, trials were averaged for each participant and session across all electrode channels and time points (channel x time). These individual results were then compiled into group-level matrices (66 x 875 x 20) for both the vocalization (S1) and playback (S2) sessions.

2.5.1 Grand-average ERP (S1 vs S2)

Finally, to visualize these responses, ERPs were first averaged across a predefined central ROI (FC1, FC2, C1, Cz, C2, CP1, CPz, CP2) for each participant, and then averaged across participants to obtain session-level grand-average waveforms for S1 and S2.

2.5.2 Interval-specific ERP analysis (A vs B)

To investigate whether Event-Related Potential (ERP) responses fluctuated based on when a perturbation occurred during a trial, trials were separated into A-interval and B-interval conditions based on behavioral markers. For both S1 and S2, ERPs were calculated at electrode Cz by averaging the trials within each interval type (A vs B). ERPs were time-locked to the vocal onset and analyzed over the -500 to 3000 ms window. This descriptive approach was intended to visualize potential variations in neural activity associated with perturbation timing.

2.6 Frequency Analysis

Following the visualization of ERPs, spectral power analyses were performed to investigate oscillatory brain activity, with a specific emphasis on the alpha band.

Power spectral density (PSD) was computed from ICA-cleaned datasets using the FFT-based EEGLAB's spectopo function. These analyses were conducted individually for every participant, across both the vocalization (S1) and the playback (S2) sessions, and for each experimental interval (A and B).

To facilitate trial selection, EEG data segments (epochs) were synchronized with behavioral information using event markers that contained specific block and trial identifiers. This matching process allowed the categorization of the EEG epochs into A-interval or B-interval for each participant. Ultimately, separate datasets were created for A-interval and B-interval trials within both S1 and S2.

2.6.1 Power Spectral Density estimation

To analyze the oscillatory activity within the EEG recordings, Power Spectral Density (PSD) was calculated. PSD estimation was performed using the EEGLAB spectopo function, which utilizes a Fast Fourier Transform (FFT) approach with segment averaging, similar to Welch's method, which improves spectral stability by averaging power spectra across multiple signal segments (Delorme & Makeig, 2004; Welch, 1967).

In the frequency domain, the PSD can be expressed as:

$$PSD(f) = \frac{|X(f)|^2}{N}$$

Where $X(f)$ represents the Fourier transform of the EEG signal and N is the number of samples used in the spectral estimate.

For every participant and experimental condition, the PSD was determined for each EEG channel across a frequency window of 0 to 40 Hz, with the resulting power values expressed in decibels (dB). Power spectra were first computed at the single-channel level and then averaged across a specific central region of interest (ROI). This ROI included a cluster of central electrodes: FC1, FC2, C1, Cz, C2, CP1, CPz, and CP2.

These PSD estimates were subsequently used for extracting alpha-band power and conducting statistical analyses to compare neural activity across the different experimental conditions. The Nyquist frequency was 125 Hz due to the 250 Hz sampling rate. This ensured that frequencies within the analyzed range (0 – 40 Hz), including the alpha band (7 – 12 Hz), could be accurately represented without aliasing.

2.6.2 Vocalization vs playback PSD (500 – 1000 ms window)

To evaluate the oscillatory activity during steady speech, Power Spectral Density (PSD) was additionally calculated within a specific time window that excluded pitch perturbations. Since the pitch-shift perturbations occurred later in the trial within designated A or B intervals, the analysis centered on the 500 – 1000 ms window following vocal onset, which occurs before the pitch-shift perturbations. This timing was also strategically selected to prevent the data from being skewed by the large transient ERP responses associated with the initial vocal onset and neural responses evoked by pitch perturbations.

For this comparison, PSD estimates were computed for each participant using EEGLAB's spectopo function and were then averaged across a central region of interest (ROI) consisting of electrodes FC1, FC2, C1, Cz, C2, CP1, CPz, and CP2. This electrode cluster covers the central sensorimotor region of the scalp, which is strongly associated with speech motor control and auditory-motor integration (Van Canneyt et al., 2021b).

For visualization purposes, group-level PSD curves were computed across participants to compare the vocalization (S1) and playback (S2) conditions during this pre-perturbation interval.

2.6.3 Alpha-band power extraction

To evaluate oscillatory activity within the alpha frequency range, alpha-band power was isolated from the power spectral density calculations between 7 and 12 Hz. For every participant and condition, the values in this specific range were averaged to create a single alpha power value, which was represented in decibels (dB), as returned by the EEGLAB spectopo function.

Two specific categories of alpha measures were elicited from the PSD analyses. First, power was recorded individually for interval A and interval B during each session, producing four alpha values per subject: S1–A alpha, S1–B alpha, S2–A alpha, and S2–B alpha. These values were compiled into a participant-level table together with the corresponding trial counts. Second, to facilitate the comparison between self-generated speech and hearing the recording, alpha power was measured from PSD estimates computed within the 500 – 1000 ms segment post-onset window at ROI electrodes, consisting of FC1, FC2, C1, Cz, C2, CP1, CPz, and CP2. This produced one alpha value per participant for each session (S1 and S2). These resulting alpha measurements provided the foundation for statistical comparisons across the different experimental conditions.

2.6.4 Statistical Analysis

Because the same participants completed both vocalization (S1) and playback (S2) sessions, statistical comparisons were performed using within-subject analyses. For each participant, alpha power values were obtained as described in section 6.3 and compiled into participant-level datasets.

Given the within-subject nature of the study, paired-sample t-tests were implemented to compare S1 and S2 conditions. Paired tests were conducted individually for interval A and interval B, as well as for the mean alpha power averaged across both time intervals.

To quantify the effect sizes, Cohen’s dz, was calculated, providing a standardized mean difference of the within-subject differences, and is appropriate for paired-sample designs. In addition, 95% confidence intervals were established for the mean discrepancies found between the two conditions.

Statistical significance was evaluated using a threshold of $\alpha = 0.05$, and the entire statistical analysis was executed within MATLAB (MathWorks, Natick, MA, USA).

3. Results

This section presents the results of the ERP and spectral analyses described in the Methods. Results are reported in the same order as the analytical procedures.

3.1 Voice-onset ERP results

Event-related potentials (ERPs) synchronized to the start of vocalization for every subject and session, utilizing epochs from -500 to 3000 ms. Baseline correction was applied using the -500 to 0 ms interval. Participant-level ERPs were obtained by averaging trials within each session.

For visualization, these signals were averaged across a central region of interest (ROI) composed of specific electrodes, including FC1, FC2, CP1, and Cz. These were then combined across all participants to generate a grand-average waveform.

The resulting grand averages demonstrated a robust neural response at the moment of vocal onset, which was followed by persistent activity that continued throughout the vocalization period. Across the entire analyzed window, the general shape and structure of the waveform were morphologically similar between the active vocalization (S1) and passive playback (S2) experimental states.

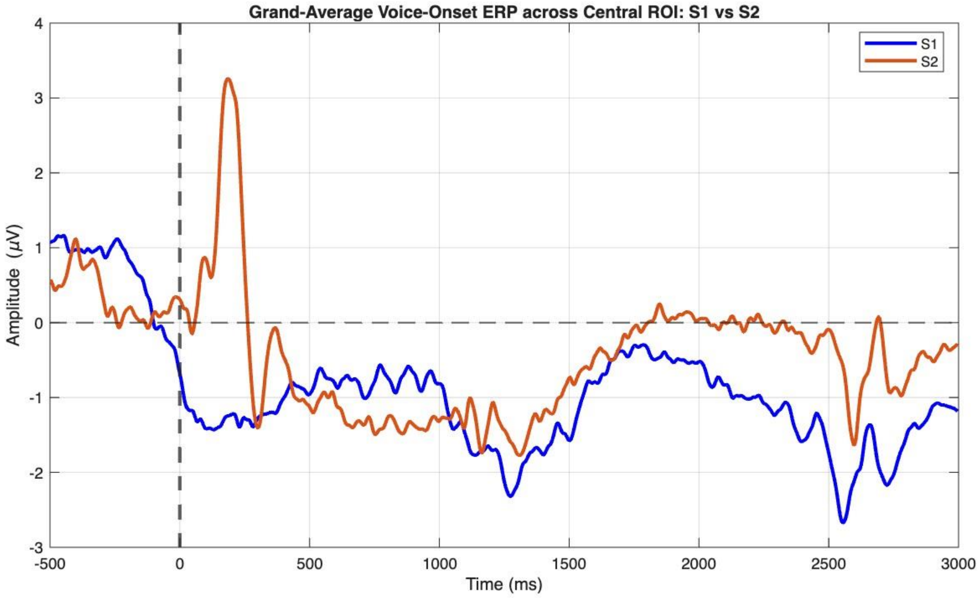
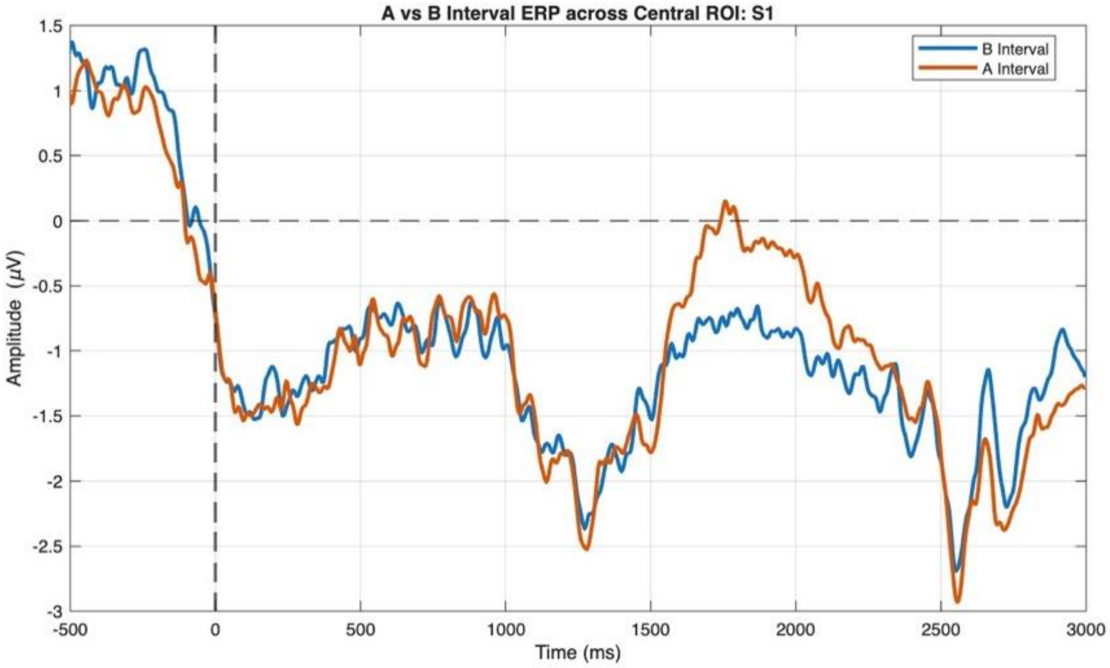


Figure 3: Grand-average ERPs time-locked to vocal onset (0 ms) across the central ROI for vocalization (S1) and playback (S2). Waveforms are baseline-corrected (–500 to 0 ms) and averaged across trials and participants. Both conditions show similar morphology, with a clear onset response followed by sustained activity.

3.1.1 Interval-specific analysis

To determine whether the timing of the pitch-shift perturbations influenced the neural responses, trials were further separated into A-interval and B-interval conditions and examined across the central ROI. In both S1 and S2 sessions, the ERP waveforms exhibited a broadly similar temporal structure across intervals. Following vocal onset (0 ms), the waveforms showed an initial large deflection that likely reflects neural activity associated with speech initiation and auditory processing of self-generated vocal output. Subsequent slower fluctuations and oscillatory-like patterns extended throughout the post-onset period, potentially reflecting ongoing sensorimotor and auditory feedback processes during the trial.

Although some amplitude differences between A- and B-interval trials were visible during later post-onset segments, particularly around the periods preceding the expected perturbation timing, the overall waveform morphology remained comparable between intervals. These observations were considered exploratory and were not subjected to formal statistical testing.



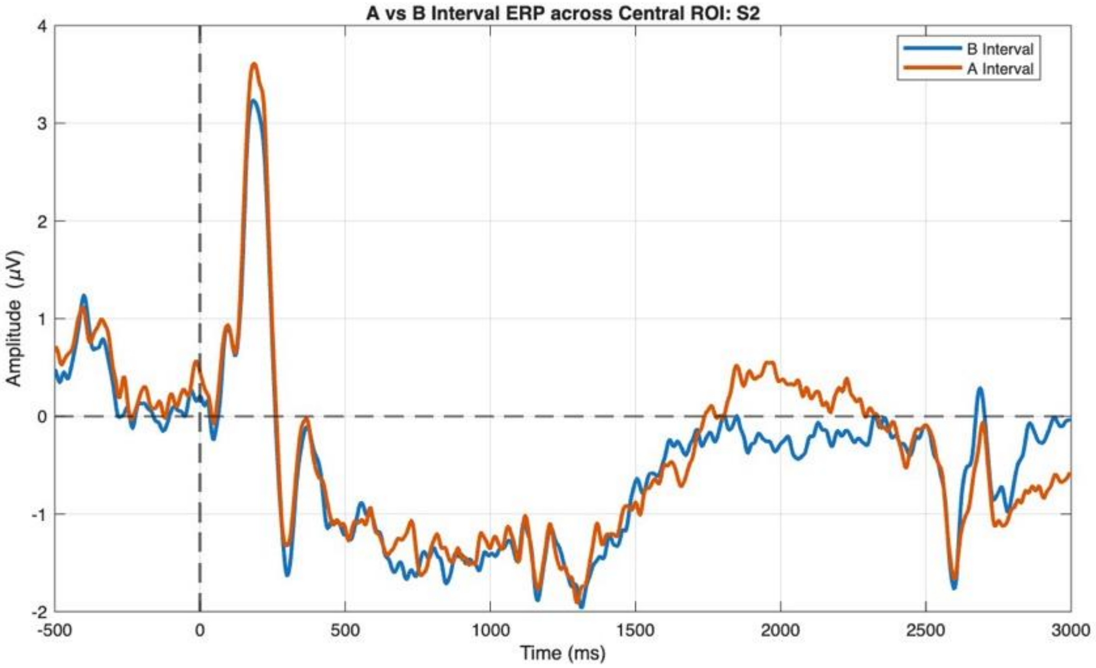
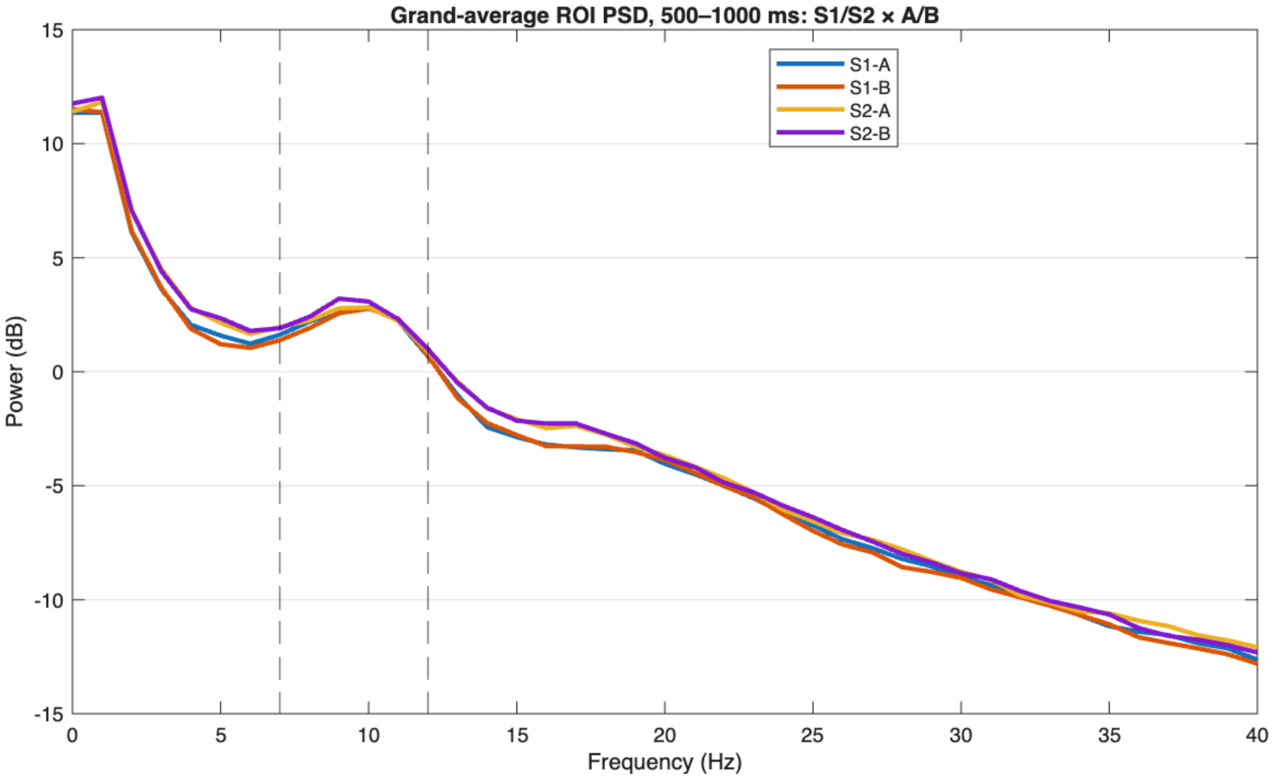


Figure 4: Grand-average ERPs across the central ROI for A- and B-interval trials in S1 (top) and S2 (bottom). Waveforms are time-locked to vocal onset (0 ms). Neural responses show similar morphology across intervals in both sessions, with only minor amplitude differences.

3.2 PSD results

Power spectral density (PSD) estimates were computed for each participant and condition using the procedures described in section 6. Spectral estimates were calculated from the 500–1000 ms post-vocal-onset window and averaged across participants as well as across the central ROI electrodes (FC1, FC2, C1, Cz, C2, CP1, CPz, and CP2).

Grand-average PSD curves demonstrated a clear spectral peak within the alpha frequency range (~9–10 Hz). This alpha-band feature was consistently observed across both vocalization (S1) and playback (S2) conditions, as well as across both A and B intervals. Visual inspection of the PSD curves indicated highly similar spectral profiles between conditions, with no pronounced differences in alpha-range power.



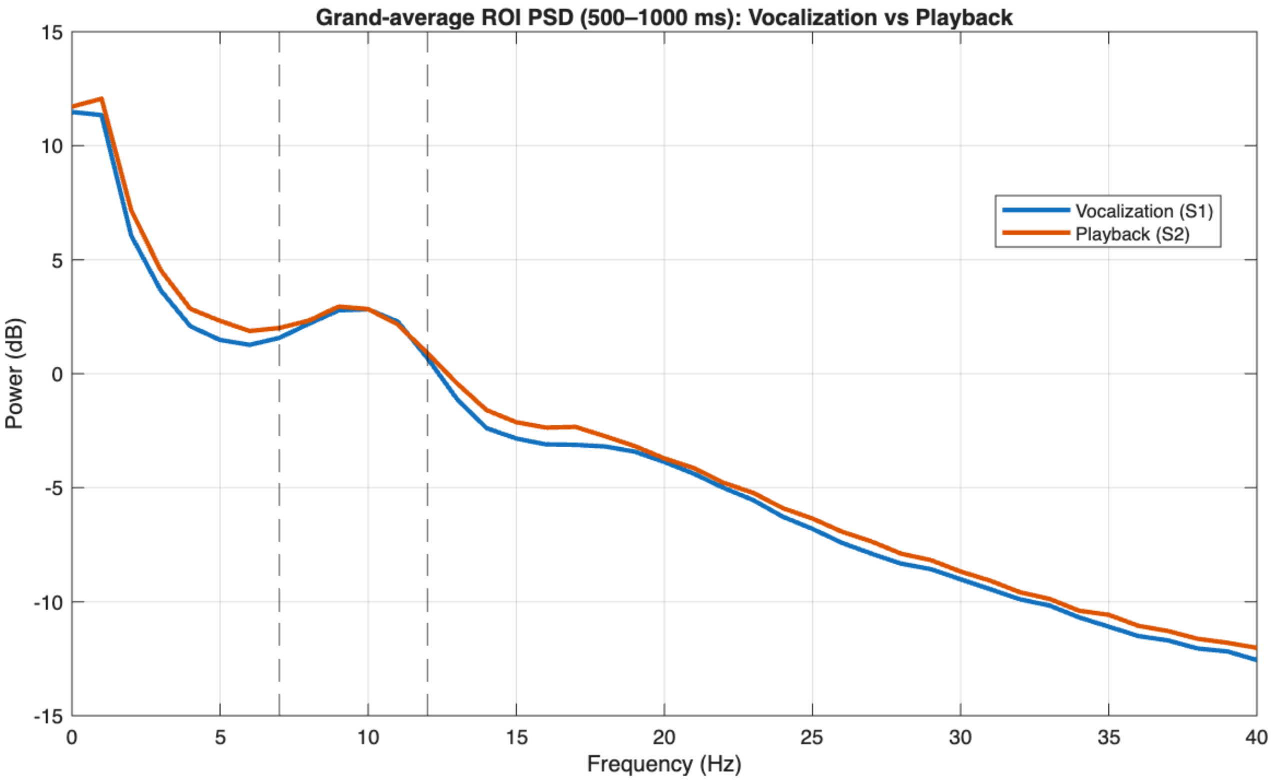


Figure 5: Figure 5: Grand-average power spectral density (PSD) across the central ROI during vocalization (S1) and playback (S2), calculated from the 500–1000 ms post-vocal-onset window. Both conditions show a clear alpha-band peak (~9–10 Hz; dashed lines), with overall similar spectral profiles and no pronounced difference in alpha-range power.

3.3 Alpha power statistical results

Alpha-band activity was quantified as the mean spectral power within the 7–12 Hz frequency range extracted from the 500–1000 ms post-vocal-onset window at the ROI electrodes. For each participant, alpha power values were computed separately for intervals A and B in both sessions, resulting in four condition-specific measures: S1-A, S1-B, S2-A, and S2-B.

Grand-average PSD curves showed a prominent alpha-band peak centered around 9–10 Hz across all conditions. Mean alpha power values derived from the grand-average spectra were comparable across conditions (S1-A = 2.03 dB, S1-B = 1.94 dB, S2-A = 2.15 dB, S2-B = 2.31 dB), indicating generally stable alpha activity across both vocalization and playback sessions.

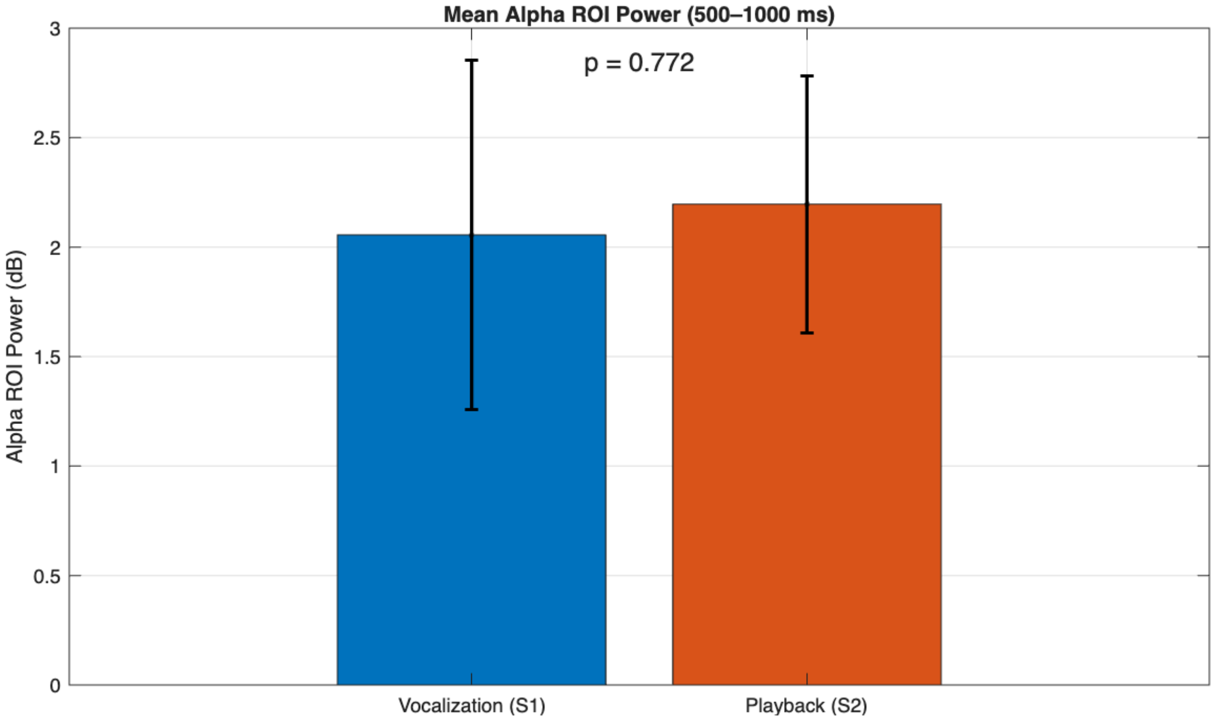


Figure 6: Mean alpha-band power (7–12 Hz) averaged across ROI electrodes for conditions S1 and S2. Error bars represent ± SEM across participants.

3.4 Statistical comparisons

Paired-sample t-tests were conducted to compare alpha-band power between vocalization (S1) and playback (S2) conditions within the examined ROI and 500–1000 ms post-vocal-onset window. Averaged across intervals, the comparison revealed no significant difference between conditions, $t(19) = -0.294$, $p = 0.772$, with a mean difference of -0.140 dB (95% CI [-1.140, 0.860]). The corresponding effect size was negligible (Cohen’s $d_z = -0.066$).

To further evaluate potential interval-related effects, a 2 × 2 repeated-measures ANOVA was performed with Session (S1 vs S2) and Interval (A vs B) as within-subject factors. The analysis revealed no significant main effect of Session, $F(1,19) = 0.285$, $p = 0.600$, no significant main effect of Interval, $F(1,19) = 0.363$, $p = 0.554$, and no significant Session × Interval interaction, $F(1,19) = 2.783$, $p = 0.112$.

Overall, these findings indicate that alpha-band activity remained highly comparable across vocalization and playback conditions, as well as across perturbation intervals, suggesting no substantial modulation of alpha-range spectral activity under the present experimental conditions.

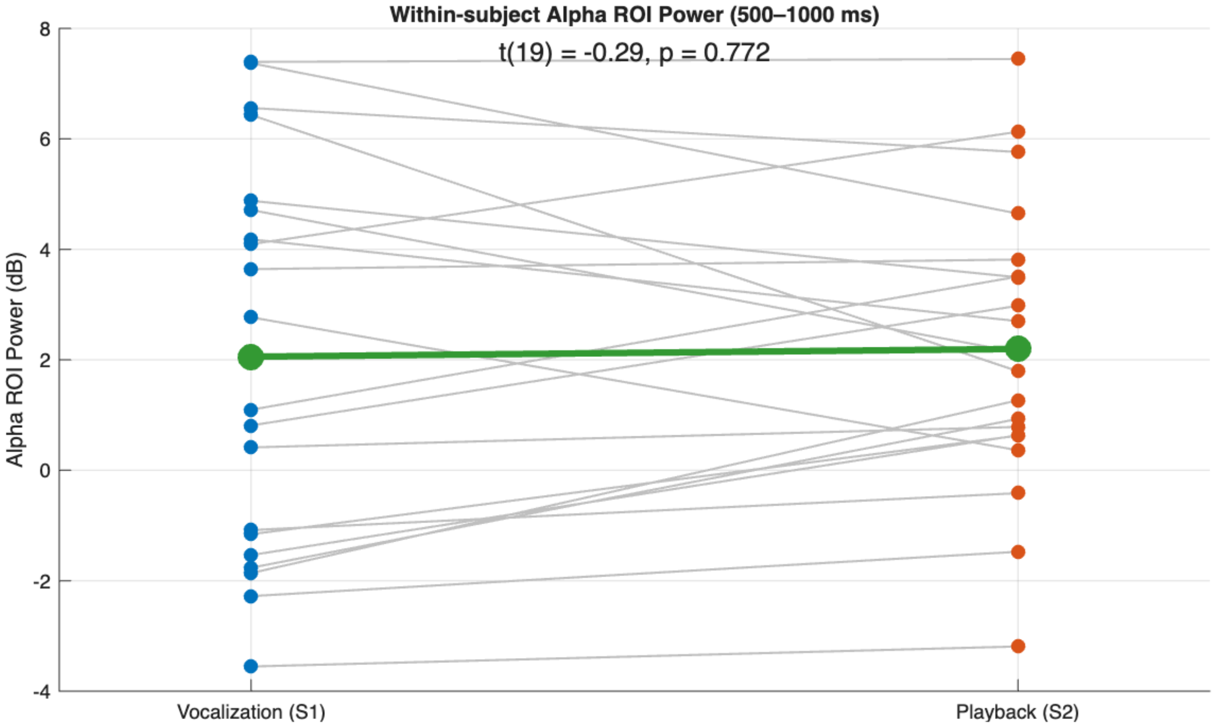


Figure 7: Individual and mean alpha-band ROI power (500–1000 ms post-vocal-onset window) during vocalization (S1) and playback (S2). Each gray line represents a participant, and the green line indicates the group mean. No significant difference in alpha power was observed between conditions, t(19) = −0.294, p = 0.772.

4. Discussion

4.1 Summary of main findings

The present study explored alpha-band activity during both sustained speech production and passive listening using EEG spectral power analysis. Neural signals were specifically monitored during a post-onset time window for active vocalization and playback trials, where the acoustic input was physically calibrated to be identical. The findings indicated that alpha-band oscillations (7 – 12 Hz) were prominent throughout both tasks; however, no statistically significant difference in alpha-band power was detected between the speaking and listening conditions. These results suggest that while alpha-band activity is a major feature of the neural signal during sustained speech processing, it does not appear to be selectively modulated by self-produced vocalization under these specific experimental parameters (Zhang et al., 2025b).

4.2 Interpretation of alpha-band activity

The observation that alpha-band activity remained relatively consistent across both active vocalization and passive playback conditions suggests that alpha-range spectral activity may reflect neural processes related to general auditory or cognitive processing rather than mechanisms uniquely associated with self-produced speech monitoring (Etard & Reichenbach, 2019; Viswanathan et al., 2023). Because no statistically significant difference in alpha-band power was observed between conditions, the present findings do not support the hypothesis that alpha activity serves as a selective neural marker of predictive speech-monitoring processes during sustained vocalization (Franken, Eisner, et al., 2018; Zhang et al., 2025b).

One initially surprising aspect of the findings was that previous analyses of similar data had visually suggested stronger alpha-range rhythmic activity during vocalization compared to playback (Suchý et al., 2023). However, when spectral power was analyzed across a broader central electrode cluster, alpha-band activity appeared highly similar between conditions. This suggests that earlier observations may have been influenced by single-electrode effects or by features of the averaged signal that were reduced when using a more spatially robust electrode-cluster approach.

Instead, alpha-range activity observed in the present study may reflect broader neural dynamics associated with sustained auditory engagement, attentional processes, or general cortical state regulation rather than speech-specific monitoring mechanisms (Ala et al., 2020; Billig et al., 2019). Importantly, the present findings do not necessarily indicate that continuous monitoring processes are absent during speech production. Rather, they suggest that such mechanisms may not be selectively reflected in alpha-band spectral power under the current experimental conditions (Etard & Reichenbach, 2019)

4.3 Methodological Considerations and Limitations

Several methodological limitations must be considered when interpreting the results of the current study. First, the investigation focused exclusively on alpha-band spectral power and did not conduct a formal statistical analysis of other frequency ranges. While visual inspection of the power spectra suggested potential differences in theta and beta-band activity between conditions, aligning with findings reported in previous altered auditory feedback studies (Franken, Eisner, et al., 2018), these frequency ranges were not formally analyzed in the present study.

Second, the current analyses utilized conventional power spectral density (PSD) estimates, which fail to separate the periodic (oscillatory) and aperiodic components of the EEG signal. Because these components were not explicitly distinguished, the observed alpha-range activity cannot be strictly interpreted as evidence of a distinct physiological oscillation (Donoghue et al., 2020). Utilizing spectral parameterization in future research would help determine if these findings represent true rhythmic oscillations or are instead driven by broader shifts in the aperiodic signal (Donoghue et al., 2020).

Third, the results were influenced by electrode selection and spatial averaging strategies. Initial observations of a stronger alpha rhythm during vocalization were attenuated when activity was averaged across a broader central electrode cluster. While this type of cluster-based averaging is effective for reducing noise and improving robustness, it may simultaneously make localized neural activity that is otherwise detectable at individual electrode sites (Jenson et al., 2015; Van Canneyt et al., 2021b).

Finally, although the laboratory setting allowed for precise control over auditory input, the task remained highly structured and distinct from natural communication. The reliance of sustained vowel vocalizations rather than spontaneous, continuous speech may limit the extent to which these findings can be generalized to the complex dynamics of everyday conversation (Bouchard & Chang, 2014).

4.4 Strengths and Future Directions

Despite these limitations, the present study also possesses several notable strengths. A major advantage of the experimental framework is the direct comparison between active vocalization and passive playback using physically identical auditory stimuli. This allowed differences between conditions to be interpreted primarily in relation to self-generated speech processes rather than acoustic variability.

In addition, the implementation of continuous vocalization paradigms allows for the investigation of neural dynamics that extend beyond the transient, event-locked responses typically emphasized in altered auditory feedback research (Bachmann et al., 2021; Van Canneyt et al., 2021a). By analysing sustained post-onset activity, this study aligns with the increasing scientific focus on the brain's "always-on" monitoring mechanisms, the persistent supervisory loops required to maintain speech stability in real-time (Bachmann et al., 2021).

Further investigations can build upon these results by exploring a broader range of frequency bands, particularly theta (4 – 8 Hz) and beta (13 – 30 Hz) activity, both of which have been identified as neural markers for sensorimotor integration and error monitoring during speech production (Franken, Eisner, et al., 2018; Zhang et al., 2025b).

Applying methods that separate periodic and aperiodic spectral activity may also provide a more detailed characterization of the neural dynamics underlying sustained speech production. Furthermore, combining spectral analyses with source localization or connectivity modelling (such as structural equation modelling or Granger causality) may help clarify how distributed sensorimotor networks contribute to continuous auditory feedback monitoring during speech (Behroozmand & Larson, 2011).

4.5 Conclusion

The present study investigated the alpha-band spectral activity during sustained vocalization and passive playback using EEG power spectral analyses. Although alpha-range activity was clearly present during both conditions, no statistically significant differences in alpha-band power were observed between self-produced and externally presented speech. These findings suggest that alpha activity may reflect broader auditory or cognitive processes, such as attentional allocation, functional inhibition, or global sensory engagement, rather than neural mechanisms uniquely associated with speech-specific monitoring (Billig et al., 2019). Overall, the study contributes to ongoing efforts to better understand the neural dynamics underlying continuous auditory feedback processing during speech production.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Henry Railo, for his guidance, patience, and continuous support throughout this thesis project. His feedback, scientific insight, and encouragement were invaluable during every stage of this work. I am sincerely thankful for the time and effort he dedicated to helping me improve both the quality of this thesis and my development as a researcher.

I would also like to extend my heartfelt thanks to Sampo Tanskanen for his immense help and support throughout this process. His willingness to assist with any challenges, answer countless questions, and provide guidance whenever needed made a significant difference in completing this work. I am deeply grateful for his kindness, patience, and generosity.

Completing this thesis was not only an academic challenge, but also a deeply personal one. During the course of writing this work, I experienced one of the most difficult periods of my life, including going through a divorce while witnessing the ongoing suffering and violence affecting my home country, Iran. Carrying the emotional weight of these experiences while trying to continue academic work was, at times, overwhelming. Yet, this thesis became a reminder that persistence can exist even during instability, uncertainty, and grief.

As an Iranian woman, I would also like to acknowledge the courage and resilience of those who continue to fight for dignity, equality, and freedom under the slogan “Woman, Life, Freedom” (“Zan, Zendegi, Azadi”). Their bravery and resistance became a source of inspiration and hope during difficult moments, and a reminder of the strength that can exist even in the face of fear and oppression.

I would especially like to thank my family for their unwavering support throughout my academic journey. To my mother, thank you for your endless love, sacrifice, and belief in me, even during the moments when I struggled to believe in myself. Your support has carried me through more than words can express.

I would also like to thank my friend, Nick, for the role he played during an important chapter of my life and for supporting parts of this journey in ways that helped me continue moving forward.

To my cat, Abigol, thank you for the comfort and companionship. Your presence made difficult days feel a little lighter.

Finally, I would like to thank my boyfriend, Elmo, for his patience, encouragement, and emotional support, and for standing beside me throughout the final stages of this journey. Your care and reassurance helped me keep going during moments when finishing this thesis felt impossible.

I would also like to express my gratitude to Albert and Jemina for their support, encouragement, and the work they have done in creating a supportive academic and student environment. Their kindness and dedication made a meaningful difference during my time in the program.

In addition, I would like to thank the Human Neuroscience program at the University of Turku for providing an inspiring and intellectually stimulating environment throughout my studies. I am grateful to my fellow students and friends for their support, encouragement, and companionship during both the stressful and rewarding moments of this journey.

This thesis is not only the result of academic effort but also a reflection of resilience, support, and the people who stood beside me throughout this journey.

Disclaimer: Grammarly was used for language corrections and some paraphrasing.
ChatGPT was used as an assisting tool for the coding process and thesis formatting.

References

Ala, T. S., Graversen, C., Wendt, D., Alickovic, E., Whitme, W. M., & Lunner, T. (2020). An exploratory study of EEG alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. *PLoS ONE*, *15*(7 July). https://doi.org/10.1371/journal.pone.0235782

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802. https://doi.org/10.1162/jocn.2009.21055

Bachmann, F. L., MacDonald, E. N., & Hjortkjær, J. (2021). Neural Measures of Pitch Processing in EEG Responses to Running Speech. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.738408

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*. https://doi.org/10.1186/1471-2202-12-54

Billig, A. J., Herrmann, B., Rhone, A. E., Gander, P. E., Nourski, K. V., Snoad, B. F., Kovach, C. K., Kawasaki, H., Howard, M. A., & Johnsrude, I. S. (2019). A sound-sensitive source of alpha oscillations in human non-primary auditory cortex. *Journal of Neuroscience*, *39*(44), 8679–8689. https://doi.org/10.1523/JNEUROSCI.0696-19.2019

Bouchard, K. E., & Chang, E. F. (2014). Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *Journal of Neuroscience*, *34*(38), 12662–12677. https://doi.org/10.1523/JNEUROSCI.1219-14.2014

Bradshaw, A. R., Wheeler, E. D., McGettigan, C., & Lametti, D. R. (2025). Sensorimotor learning during synchronous speech is modulated by the acoustics of the other voice. *Psychonomic Bulletin and Review*, *32*(1), 306–316. https://doi.org/10.3758/s13423-024-02536-x

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., Knight, R. T., Shestyuk, A., & Voytek, B. (2020).

Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665. https://doi.org/10.1038/s41593-020-00744-x

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. *The Journal of Neuroscience*, *39*(29), 5750. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin and Review*, *25*(4), 1458–1467. https://doi.org/10.3758/s13423-018-1494-x

Franken, M. K., Eisner, F., Acheson, D. J., McQueen, J. M., Hagoort, P., & Schoffelen, J. M. (2018). Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. *NeuroImage*, *179*, 326–336. https://doi.org/10.1016/j.neuroimage.2018.06.061

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guo, N., Si, X., Zhang, Y., Ding, Y., Zhou, W., Zhang, D., & Hong, B. (2021). Speech frequency-following response in human auditory cortex is more than a simple tracking. *NeuroImage*, *226*. https://doi.org/10.1016/j.neuroimage.2020.117545

Hawco, C. S., Jones, J. A., Ferretti, T. R., & Keough, D. (2009). ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*, *46*(6), 1216–1225. https://doi.org/10.1111/j.1469-8986.2009.00875.x

Hickok, G. (2012). Computational neuroanatomy of speech production. In *Nature Reviews Neuroscience* (Vol. 13, Number 2, pp. 135–145). https://doi.org/10.1038/nrn3158

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. In *Neuron* (Vol. 69, Number 3, pp. 407–422). https://doi.org/10.1016/j.neuron.2011.01.019

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. In *Frontiers in Human Neuroscience* (Number OCTOBER). Frontiers Media S. A. https://doi.org/10.3389/fnhum.2011.00082

Jenson, D., Harkrider, A. W., Thornton, D., Bowers, A. L., & Saltuklaroglu, T. (2015). Auditory cortical deactivation during speech production and following speech perception: An EEG investigation of the temporal dynamics of the auditory alpha rhythm. *Frontiers in Human Neuroscience*, *9*(OCT). https://doi.org/10.3389/fnhum.2015.00534

Liu, H., Behroozmand, R., & Larson, C. R. (2010). Enhanced neural responses to self-triggered voice pitch feedback perturbations. *NeuroReport*, *21*(7), 527–531. https://doi.org/10.1097/WNR.0b013e3283393a44

Liu, Y., Fan, H., Li, J., Jones, J. A., Liu, P., Zhang, B., & Liu, H. (2018). Auditory-Motor Control of Vocal Production during Divided Attention: Behavioral and ERP Correlates. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00113

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters*, *591*, 25–29. https://doi.org/10.1016/j.neulet.2015.02.012

Parrell, B., & Houde, J. (2019). Modeling the role of sensory feedback in speech motor control and learning. In *Journal of Speech, Language, and Hearing Research* (Vol. 62, Number 8S, pp. 2963–2985). American Speech-Language-Hearing Association. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Reichenbach, T. (2019). *Behavioral/Cognitive Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise*. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Reznik, D., & Mukamel, R. (2019). Motor output, neural states and auditory perception. In *Neuroscience and Biobehavioral Reviews* (Vol. 96, pp. 116–126). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2018.10.021

Scheerer, N. E., & Jones, J. A. (2018). The role of auditory feedback at vocalization onset and mid-utterance. *Frontiers in Psychology*, *9*(OCT). https://doi.org/10.3389/fpsyg.2018.02019

Scheerer, N. E., Liu, H., & Jones, J. A. (2013). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience*, *38*(8), 3189–3200. https://doi.org/10.1111/ejn.12301

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). *Contributions of Auditory and Somatosensory Feedback to Vocal Motor Control*. https://doi.org/10.23641/asha

Suchý, D., Behroozmand, R., & Railo, H. (2023). *Conscious and unconscious perception of pitch shifts in auditory feedback during vocalization: Behavioral functions and event-related potential correlates*. https://doi.org/10.1101/2023.10.13.562262

Tang, D. L. (2024). Using altered auditory feedback to study pitch compensation and adaptation in tonal language speakers. In *Frontiers in Human Neuroscience* (Vol. 18). Frontiers Media SA. https://doi.org/10.3389/fnhum.2024.1364803

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Van Canneyt, J., Wouters, J., & Francart, T. (2021a). Enhanced Neural Tracking of the Fundamental Frequency of the Voice. *IEEE Transactions on Biomedical Engineering*, *68*(12), 3612–3619. https://doi.org/10.1109/TBME.2021.3080123

Van Canneyt, J., Wouters, J., & Francart, T. (2021b). Neural tracking of the fundamental frequency of the voice: The effect of voice characteristics. *European Journal of Neuroscience*, *53*(11), 3640–3653. https://doi.org/10.1111/ejn.15229

Viswanathan, V., Bharadwaj, H. M., Heinz, M. G., & Shinn-Cunningham, B. G. (2023). Induced alpha and beta electroencephalographic rhythms covary with single-trial speech intelligibility in competition. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-37173-2

Weerathunge, H. R., Voon, T., Tardif, M., Cilento, D., & Stepp, C. E. (2022). Auditory and somatosensory feedback mechanisms of laryngeal and articulatory speech motor control. *Experimental Brain Research*, *240*(7–8), 2155–2173. https://doi.org/10.1007/s00221-022-06395-7

Welch, P. D. (1967). *The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Aver. aging Over Short, Modified Periodograms Ik(fn) =-I A h (%) [a k* (Vol. 15, Number 2).

Zhang, Y., Sarmukadam, K., Nejati, V., & Behroozmand, R. (2025a). Functional neural oscillatory activities reveal the impact of attentional instructions on speech auditory feedback control. *Neuropsychologia*, *216*. https://doi.org/10.1016/j.neuropsychologia.2025.109202