# Inhibition of articulation muscles during listening and reading: a matter of intention to speak out loud

Naama Zur[12] Zohar Eviatar[123] Avi Karni[34]

1 Institute for Information Processing and Decision Making, University of Haifa, Haifa, Israel 3498838

2 Department of Psychology, University of Haifa, Haifa, Israel 3498838

3 Edmund J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel 3498838

4 Sagol Department of Neurobiology, University of Haifa, Haifa, Israel 3498838

Corresponding Author:

Naama Zur

Institute for Information Processing and Decision Making

University of Haifa

Haifa, Israel 3498838

Email: naamazur@gmail.com

## Abstract

Articulation of a given speech sound is often contingent on the intention to produce subsequent sounds (co-articulation). Thus, intended acts affect the way current acts are executed. Here we show that the intention to subsequently repeat a short sentence, overtly or covertly, significantly modulated the articulatory musculature already during listening or reading (i.e., during the input phase). Young adults were instructed to read (whole sentences or word-by-word) or listen to recordings of sentences to be repeated afterwards. Surface electromyography (sEMG) recordings showed significant reductions in articulatory muscle activity, above the orbicularis-oris and the sternohyoid muscles, compared to baseline, during the input phase. These temporary reductions in EMG activity were contingent on the intention to subsequently repeat the input overtly or covertly, as well as on the input modality. Inhibition was stronger when an overt response was intended. When participants intended to repeat the sentences covertly, only auditory input but not reading resulted in significant reductions in sEMG activities during the input phase. Higher temporal resolution analyses also showed different patterns of activity modulations before the cue to respond, depending on the input modality and the intended response. Only when repetition was to be overt, a significant build-up of activity occurred after sentence presentation, before the cue to respond; this build-up was most pronounced when the sentence was heard or read word-by-word. Neurolinguistic models suggest that language perception and articulation interact; the current results suggest that the interaction begins already during the input phase, listening or reading, and reflects the intended responses.


Keywords: Surface electromyography (sEMG), articulatory muscles, inhibition, intention

**Statement of Significance**

Intended acts affect the way current acts are executed. Our results show that already during the perception of language input (listening to, or reading, short sentences), the state of muscles used in articulating speech sounds is affected by the intention of the individual to act in the future: repeat what was heard or read, either aloud or silently. There was stronger temporary inhibition of articulatory muscles before voiced than before silent repetition. Moreover, whether a sentence was heard or read was also reflected in the pattern of articulatory muscle inhibition. Neuro-linguistic models suggest that language perception and articulation interact; the current results suggest that the interaction begins already during input perception, listening or reading, and reflects our intended acts.

## 1. Introduction

Whether and how motor representations of speech are integral to speech perception has been debated for many decades. One perspective posits that the motor system plays a functional role both in speech production and in speech perception (e.g., Liberman, 1957; Liberman et al., 1967; Wilson, Saygin, Sereno, & Iacoboni, 2004; Pulvemuller and Fadiga, 2010; Yuen, Davis, Brysbaert, & Rastle, 2010; Skipper, Delvin, & Lametti, 2017). A related position assumes an interaction between the perception of speech and the motor articulatory system, although the actual activation of the motor system is not necessary for perception (e.g., Hickok, & Poeppel, 2007; Poeppel & Hickok, 2004; Scott, McGettigan, and Eisner, 2009). According to the State-Feedback-Control (SFC) model of speech production (Hickok, Houde, and Rong, 2011), there are pathways both for the activation of motor speech systems from sensory input (the forward prediction pathway) and for the mutual activation of auditory speech systems from motor activation (the feedback correction pathway). Thus, both views suggest that the state of articulatory musculature may be affected during speech perception (Galantucci, Fowler, & Turvey, 2006; Glenberg & Gallese, 2012; Hickok, Houde, and Rong, 2011). In the present study, the state of articulatory muscles, during speech perception, was recorded to test the conjecture that the articulatory musculature is affected by the mode of language input but also by the intended motor task, specifically the intention to subsequently produce covert vs. overt speech.

Covert speech can be defined as a subjective “experience of language in the absence of overt and audible articulation" (e.g., Alderson-Day & Fernyhough, 2015), conceptualized as a type of motor output, inner speech (Frith, 1992). In healthy individuals, the effortless differentiation between inner, self-generated, speech and the perception of one’s overt and others’ speech (Allen, Aleman, & Mcguire, 2007) but also between self-generated and elicited

inner speech (Hulburt, Alderson-Day, Kuhn, and Fernyhough, 2016) is considered to reflect an interaction of perceptual and motor speech systems.

Current notions of whether and how motor representations of speech contribute to speech perception do not differentiate between the effects of different language input modalities. The primary modality for language is auditory, however, in literate individuals the interpretation of visual written language becomes automatic (Logan, 1997; Augustinova and Ferrand, 2014). Thus, another goal of the present study was to explore the effects of the modality of the linguistic input, read or heard, on the state of the articulatory system. Input-modality specific effects would support the hypothesis of multiple modes of interaction between the perception and motor production of language.

Two questions were addressed in the current study: *i*. whether the intention to subsequently generate a covert vs. an overt motor output, would differentially affect the activation of articulatory muscles, during the language input phase, and *ii*. whether the sensory-input mode would affect the articulatory muscles before a speech response was initiated. To this end, a repetition task was developed, requiring the delayed production of either covert or overt speech in response to sentences presented as auditory or visual input.

We recorded surface electromyography (sEMG) over the Orbicularis Oris Inferior (OOI) and the Sternohyoid (STH) muscles throughout the repetition task, as shown in panel A of Figure 1. On each trial, participants were presented with the sentence to be repeated in either the visual or auditory modality. However, participants were instructed to repeat the target sentence only after a visual ‘go’ cue. Thus, each trial was divided into a baseline, an ‘input’ epoch, which was the sentence presentation part of the trial, and an ‘output’ epoch, after the cue to repeat. Articulatory muscle activation was calculated separately for each epoch; the current report

addresses the input, i.e., perceptual, epoch only (panel B of Figure 1). Previous studies have shown subtle muscle activity in speech musculature using sEMG during verbal mental imagery, silent reading, and silent recitation (Jacobson, 1931; Livesay, Liebke, Samaras, & Stanley, 1996; McGuigan & Dollins, 1989). Also, activity in the OOI and the STH muscles was significantly higher when patients with schizophrenia reported hallucinations (Rapin, Dohen, Polosan, Perrier, and Lœvenbruck, 2013). Lip muscles activation, as recorded in sEMG, was higher during covert speech, compared to activation in a silent visualization task (Livesay et al., 1996). However, studies that have examined EMG muscle activity of peripheral musculature in tasks unrelated to language (Conrad, Benecke, and Goehmann, 1983; Richartz, Lévénez, Boucart, & Duchateau, 2010; Kolářová, Krobo, Kolář, Hluštík, & Polehlová, 2016), using both real and imagined movements, reported a 'premotor silent period' that is, an inhibition relative to baseline. Thus, the approach to data analysis in the current study was to compare the relative activation of articulatory muscles before (baseline) and during stimulus presentation (up to the 'go' cue for snetence repetition) in the different input and intended response conditions.

## 2. Results

The panels in Figure 2 present examples of the raw sEMG recordings (in μV) from both the OOI and STH muscles recorded from one of the participants (a representative trial from each of the 6 study conditions). For each trial, we computed the average values for two dependent variables: *i.* average rectified value (ARV), and *ii*, the root mean square (RMS) of the sEMG activity (μV). The values were extracted from two segments (epochs) of each trial: 1) fixation (baseline) epoch, 2) the stimulus presentation epoch (sentence presentation up to the Go cue)

(Figure 1 Panel B) (Table 1). A relative stimulus presentation activity measure (r-StimPresA = $\mu V_{(\text{stimulus presentation segment})} - \mu V_{(\text{fixation segment})}$) was computed for each trial.

**ARV Analyses**

A 2-way repeated measures ANOVA (Greenhouse-Geisser corrected) with Input Type (A (auditory), V1 (visual whole sentence), V2 (visual word-by-word)) and the Intended Output (Covert and Overt) for the r-StimPresARV from the OOI muscle showed a significant main effect of the Intended Output ($F_{(1, 19)}$= 13.457, P=.002, $\eta_p^2$=.415). Post hoc tests showed larger inhibition of activity (more negative r-StimPresARV) during the stimulus presentation epochs when the intended response was to be overt compared to when the intended response was to be covert (mean difference: 1.466 µV, P=.002) (Figure 3A). There was no main effect of Input Type, and no significant interaction.

A similar analysis on the r-StimPresARV derived from the Sternohyoid (STH) also showed a significant main effect of the Intended Output ($F_{(1, 19)}$= 17.823, P<.001, $\eta_p^2$=.484) (Figure 3B). When an overt response was intended, stimulus presentation epochs were significantly more negative compared to stimulus presentation epochs when a covert response was intended (mean difference: .103 µV, P<.001). There was no main effect of the Input Type and no significant interaction.

**RMS Analyses**

The difference in RMS between the sentence presentation epochs and the corresponding fixation epochs (r-StimPresRMS) in the OOI and STH muscles are presented in Figure 3 panels C and D, respectively. A 2-way repeated measures ANOVA, with Input Type (A, V1, V2) and the Intended Output (Covert, Overt) for the r-StimPresRMS in the OOI muscle, showed a significant main effect of the Intended Output ($F_{(1, 19)}$= 8.685, P<.01, $\eta_p^2$=.314). This reflected

the relatively larger inhibition of activity (more negative r-StimPresRMS) during the stimulus presentation epochs when the intended response was overt, compared to when the intended response was covert (mean difference: 1.911 µV, P=.008) (Figure 3C). There was no main effect of Input Type, but there was a significant interaction of Intended Output and Input Type ($F_{(2, 38)}$= 3.516, P=.04, $\eta_p^2$=.156). Separate 1-way repeated measures ANOVAs, for each intended output, showed that when the intended output was covert there was a significant simple main effect of Input Type ($F_{(2, 38)}$= 3.406, P=.044, $\eta_p^2$=.152). Post hoc tests (Bonferroni corrected) indicated that OOI muscle activity was significantly more inhibited during the auditory presentation compared to V1 presentation (mean difference: .326 µV, p=0.01) (Figure 3C). There were no significant differences between r-StimPresRMS values during V2 presentation and the two other input types. In contrast, when the intended output was overt, there was only a trend towards a significant main effect of Input Type ($F_{(2, 38)}$= 2.753, P=.076, $\eta_p^2$=.127) with a somewhat larger inhibition of activity in the V2 condition (Figure 3C).

A similar analysis of r-StimPresRMS values derived from the STH muscle, showed a main effect of the Intended Output ($F_{(1, 19)}$= 14.685, P=.001, $\eta_p^2$=.436) (Figure 3D) with larger inhibition when an overt response was intended compared to when a covert response was intended (mean difference: .139 µV, P<.001) (Figure 3D). There was no main effect of Input Type, and no significant interaction.

Higher temporal resolution analysis

A higher resolution analysis of the fixation epoch was undertaken following the findings of Pei et al. (2011), who reported ECoG data during a single word repetition task. Of relevance here, is that in Pei et al. (2011) study there was a difference in the signal dynamics during return to baseline, between the overt and the covert conditions (see their Figure 8, p. 2968). We

therefore examined the dynamics within our baseline condition in both muscles. Figure 4 shows the changes in RMS activation values between the first and the last 200 ms of the fixation epoch (which lasted 2 sec). In both muscles, activation tended to decrease across the fixation epoch, and consistently there was higher activation, during fixation, when an overt response was intended. Given these dynamics within the fixation epoch, we used the value in the second fixation interval as the baseline in the following analyses.

The higher temporal resolution analyses of the stimulus presentation epoch was undertaken to explore the possibility that articulatory muscle inhibition is a type of 'preparatory inhibition' as was reported in studies of non-language related, simple hand movements (Doque, Greenhouse, Labruna, & Ivry, 2017). To this end, we compared muscle activity (relative to the second interval of the fixation epoch) in three intervals of 200 ms each (Figure 5A). These epochs included an interval related to the onset of sentence presentation, an interval beginning at the offset of the sentence presentation (beginning of the temporally jittered delay period), and the interval just before the Go signal (end of the delay period) (Figure 5A, intervals A, B, and C, respectively). The RMS values of the sEMG activity (μV) were computed for each of these three pre-repetition intervals, and a relative activity measure was computed by subtracting the RMS (μV) value of the second fixation epoch.

A 3-way repeated measures ANOVA of Intended Output (Covert and Overt), Input Type (A, V1, V2), and Interval (A, B, and C) showed a significant main effect of the Intended Output ($F_{(1, 19)}$= 16.859, P=.001, $\eta_p^2$=.470) for the OOI only. There was larger overall inhibition prior to overt responses. There was also a marginally significant main effect of interval (Greenhouse-Geisser corrected) ($F_{(1.469, 27.915)}$= 3.33, P=.064, $\eta_p^2$=.149) with the largest relative inhibition occurring during sentence presentation (interval A) across all input types. There was,

importantly, a significant 3-way interaction of Intended Output, Input Type, and Interval ($F_{(4, 76)}$= 4.022, P<.05, $\eta_p^2$=.175). Post hoc analyses showed that overall there was a clear trend for an *increase* in inhibition– from interval A to interval C, prior to covert responses, but a *decrease* in inhibition prior to overt responses (Figure 5B). These opposing trends, increased vs. decreased inhibition across the three intervals, prior to covert and overt responses respectively, were most pronounced when sentences were presented in the auditory modality (Figure 5B panels 1-4).

To examine the differential effects of modality, we tested the interaction of Input Type with Interval separately for the two conditions of Intended Output. In the Covert condition, there was no interaction: the three input types resulted in similar dynamics across the three time-points. In the Overt condition, the interaction was significant (F(3.253,61.807)=3.727, P=<.05, $\eta_p^2$=.164, Huynh-Feldt corrected) because of an interaction between Input type and Interval in time-points B and C (F(2,38)=5.22, P<.01, $\eta_p^2$=.22) (Figure 5B). Pairwise comparisons showed that both the Auditory and the V2 conditions differed from the pattern of change in the V1 condition (Auditory vs V1: F(1,19)=4.8, P<.05, $\eta_p^2$=.20; V2 vs V1: F(1,19)=4.27, P=.053, $\eta_p^2$=.18), whereas the patterns of change in Auditory and V2 did not differ from each other.

## 3. Discussion

The results of the current study show that articulatory muscle activity was inhibited during the stimulus presentation epoch, the language perception phase, but the magnitude of inhibition was contingent on the intended (future) response, i.e., more inhibition when an overt rather than a covert response was intended. The level of articulatory muscle inhibition was also dependent on the input type: the inhibition, during sentence presentation, was stronger in the auditory presentation condition, when covert speech was to follow; in the overt condition, inhibition was significantly stronger during visual word-by-word presentation.

The more fine-grained analysis of the sentence presentation epoch showed that the time-course of the articulatory muscle inhibition was also contingent on the intended response, overt or covert sentence repetition. Thus, prior to a covert response there was an increase in inhibition as the subjects were preparing to repeat the sentence, i.e., inhibition was at peak when the response initiation cue was nearing. In contrast, when overt responses were intended, the level of articulatory muscle inhibition tended to decrease (except in the V1 condition) as the cue to initiate the response approached; inhibition was at its peak during the presentation of the sentence rather than just before the response initiation cue. Indeed, when an overt response was intended, the articulatory muscle inhibition, that began very early in the sentence presentation epoch (within about 250 ms from sentence presentation onset), decreased regardless of whether the sentence was heard or read.

In the covert condition, there was significantly stronger inhibition during the auditory sentence presentation epoch compared to the visual sentence presentation epochs. Also, the differential trend, increased vs. decreased inhibition prior to the initiation of covert or overt speech responses, respectively, was the most pronounced when sentences were presented in the auditory modality (Figure 5B). One may conjecture that covert responses, where one will refrain from voicing to input, may reflect mechanisms that evolved to inhibit echolalia during a spoken-auditory language interaction (conversation) (Stiegler, 2015). Alternatively, it may be the case that the engagement of auditory-motor articulatory (feedforward and feedback) pathways, set up for spoken-heard language, is indirect and partial in reading. Reading may require extra steps in interacting with the motor-articulatory system, e.g., a transcoding of graphemes to speech sounds, and it may be the case that motor-speech networks are activated more strongly during listening to speech than in reading.

The current results may be construed as reflecting a type of preparatory inhibition in relation to elicited speech (Conrad, Benecke, and Goehmann, 1983; Richartz et al., 2010; Kolářová et al., 2016; Lebon et al., 2016; Doque et al., 2017). However, there is a difference in the time scales; in Lebon et al.'s study, inhibitory preparatory processes were found only 200 ms (but not 500 ms or 900 ms) prior to an imperative signal to move. Moreover, unlike the trend for increased inhibition closer to the Go signal, found by Lebon et al (2016), in the current study, inhibition tended to *decrease* across time-points A, B and C before the initiation of an overt speech response, and only slightly *increase* before covert sentence repetition was cued. A positive relationship between the magnitude of pre-movement depression of tonic muscle activity and the subsequent phasic innervation burst has been suggested for ballistic arm movements (Conrad, Benecke, and Goehmann, 1983). Our findings may be linked to the findings of Conrad et al. (1983), if one considers the sentence presentation (input) epoch in the current study as a predictive phase necessarily preceding the actual Go signal. With the caveat of a different time scale and resolution, the current results show that the intended speech action (overt or covert) was a major factor in determining the degree of inhibition. However, the extended time course of inhibition and the fact that the inhibition occurred from the very beginning of the stimulus presentation epoch, when an overt response was intended, suggest a pattern of inhibition that differs from the pattern of inhibition described in non-language manual responses.

The current results are in line with two electrocorticography (ECoG) studies (Pei, et al., 2011; Leuthardt et al., 2012) of neural activity during a repetition task, requiring either overt or covert responses. The stimuli were single words presented in the visual and auditory modalities. The intention to vocalize or not to vocalize, resulted in a differential buildup of cortical activity

before the 'Go' cue in a repetition task. Both studies also showed different patterns for the two modalities of word presentation. In the auditory presentation condition, left pre-motor and primary motor cortical activity were evident during the input presentation phase, prior to both covert and overt responses. Whereas in the visual presentation condition, there was much lower left premotor and motor activation during stimulus presentation. Moreover, Pei, et al., (2011) (see their Figure 8A) showed that visual and auditory presentations resulted in different time courses of left prefrontal and premotor activations during stimulus presentation epochs. More recently, Brumberg et al., (2016) reported ECoG recordings in participants instructed to read a familiar text aloud or silently. They found left fronto-motor activity at 440-240 ms prior to the initiation of repetition, in both the overt and covert speech production conditions. However, this activity was significantly higher when overt speech production was intended compared to the covert speech production condition. The findings of the current study raise the possibility that the left fronto-motor areas dynamics may relate, directly or indirectly, to the differential dynamics of inhibition of the articulatory musculature in the language perception phase. However, our measurements were purely motor; we have no way of knowing their origins in the central nervous system.

The higher temporal resolution analyses showed that in the overt condition, the pattern of inhibition across the input phase was similar in the auditory and the visual word-by-word conditions; the visual whole sentence condition differed from both. This difference may reflect the manner in which the perceptual information was accrued in the three conditions. In both the auditory and the visual word-by-word conditions, participants had to wait until the end of the stimulus presentation epoch in order to have the full stimulus content. In the visual whole

sentence condition, the simple three word sentences could be perceived at a glance, and the sentence content may have been available relatively earlier compared to the other two conditions.

The enhanced sEMG signal observed in the OOI muscle compared to the STH muscle may reflect a potential confound resulting from the engagement of the STH muscles both in breathing and speech. In the current experimental setup and analysis, these factors were not dissociated. Nevertheless, the sEMG signals for the STH muscle were inhibited as a function of the linguistic input and output manipulations of the task, in patterns of modulations similar to the patterns recorded in the OOI muscle. In future experimental setups, the monitoring of breathing may afford means to isolate muscle activation associated with breathing, to be used as a covariate in the analysis.

Conclusions

The current results suggest that, prior to repetition, the articulatory system is actively inhibited during speech perception and reading. This inhibition was modulated by the intention to subsequently vocalize the input or not. Thus, the intention to act in the future (in terms of voicing) was a significant factor in articulatory muscle activity during language perception. The data of the current study suggest that the role of intention in language production during reading and listening, is dynamic, continuous, and contextual.

In a more general context, the current findings are also consistent with previous findings showing that planned, intended, motor actions can significantly affect ongoing actions; e.g., co-articulation, the finding that skilled speakers generate the initial phonemes of a sequence differentially, depending on the final phonemes (Kühnert & Nolan, 1999). A recent study (Rozanov, Keren, & Karni, 2010) showed that the performance of the initial movements of a well-trained sequence of finger movements was compromised by the intention to subsequently

omit a movement. In the current study, anticipatory effects were evident at the stage of input, before articulation actually occurred.

## 4. Materials and Methods

### 4.1 Participants

Twenty undergraduate students (5 males) 21-35 years old, participated in the study. All were healthy native speakers of Hebrew, right handed, and had normal or corrected-to-normal vision and normal hearing. The participants received either course credit or monetary compensation for their time (70 NIS).

### 4.2 Stimuli

The experiment included a set of 60 sentences (3 words, subject-verb-object). The stimulus set contained. There were 3 sets of 20 sentences (each set containing 96 syllables of 100 possible syllables in Hebrew). The stimulus sets were counterbalanced across participants and the three input mode conditions (auditory, visual whole sentence, visual single word sentence). Visual sentence presentation was in Ariel font (size 20), black - on a gray background.

The duration of visually presented sentences was equated to the duration of the corresponding recorded sentences (Audacity(R) Version 2.0.0. (Audacity Team, 2012)); for word-by-word visual stimuli, each word was presented separately, at the center of the screen, for a duration proportionate to the duration of it’s voiced production.

### 4.3 Apparatus

Surface EMG data were recorded at a sampling rate of 1000 Hz and amplified with a multichannel BioNex 8-slot chassis (Mindware Technologies, Grahanna, OH) equipped with a

two BioNex 4-channel bio potential amplifier (Model 50-371102- 00) (Figure 1). Data were viewed using MindWare acquisition software BioLab 2.4.

The experiment was designed using E-Prime 2 professional software (Schneider, Eschman, & Zuccolotto, 2002) and run on a HP PC with manual responses recorded using a keyboard and vocal responses through a Lavlier microphone (Alto Professional wireless microphone system) and an e-prime compatible response box. A Dell UltraSharp U2412M color monitor and headphones (Sony Stereo-Headphones, MDR-XD100) were used for visual and auditory sentence presentation, respectively. Participants were video recorded continuously.

## 4.4 Experimental Design

Since stimulus recognition in the auditory and visual modality can be temporally distinct (e.g., Whiting, Shtyrov, & Marslen-Wilson, 2014a), two visual presentation conditions were used: a visual whole sentence presentation condition, and a visual word-by-word condition in which each word of the sentence was presented separately. In both conditions sentence presentation time was the equivalent amount of time in the corresponding auditory condition.

The experimental design was a 2 (Intended Output: Covert and Overt) X 3 (Input Type: Auditory (A), Visual Whole Sentence (V1), and Visual Word-by-Word (V2)) within subject comparison design. Input conditions (A, V1, V2) were mixed and counterbalanced within each block. Each target sentence was presented twice (for overt and for covert repetition) but never in the same input modality. There were separate blocks wherein either a covert or an overt repetition was required. The covert block always preceded the overt blocks (panel B of Figure 1).

## 4.5 Procedure

Before the presentation of each sentence, a fixation cross appeared at the center of the screen for 2 seconds. The sentence to be repeated followed in one of three modes: 1) auditory, via headphones (the fixation cross stayed on screen), 2) visual, sentence in full at the center of the screen (no audition), 3) visual, word by word presentation of the sentence at the center of the screen (no audition). In the first block, participants were instructed that repetition was to be covert (silent repetition); in the second block repetition was to be overt (repetition out loud). Immediately following sentence presentation offset, a fixation cross was presented at the center of the screen for a duration of 400-700 ms (jitter), until the onset of a red dot. The red dot cued response initiation (Go cue) and was presented at the center of the screen until the participant pressed the "space" key (end of response).

Response times (RT) were measured from the onset of the red dot to "space" key press. Six trials with responses over 5 seconds or shorter than 200 ms were excluded. Inaccuracy of production (overt repetition block) occurred in eight trials (out of 1200 responses); these trails were excluded.

**4.6 EMG data processing**

The preprocessing was done using MATLAB R2018b (MathWorks Inc.). The sEMG raw data were rectified by absolute value and fed into a 20–450 Hz Butterworth band-pass filter (Butter, filtfilt, MATLAB). Additional movement artifact removal was done by offline inspection of the video recording. Swallowing was determined using both the video recordings (for physical movement of STH electrodes) and irregular peaks in the signal.

About 4.7% of the trials (113/2400 trials) were excluded from analysis due to technical and movement artifacts.

In the initial analysis, average rectified value (ARV) and root mean square (RMS) of the sEMG activity were computed for two segments: i. fixation, and ii. stimulus presentation. Analyses of the output phase, sentence repetition segments, are not reported here. The fixation segment of the trial was considered a baseline, since no linguistic content was presented in this segment. Therefore, the values of the ARV and RMS of activation in the two muscles (OOI and STH) during the fixation segment were subtracted, respectively, from the ARV and RMS of activation during sentence presentation segment. We implemented tests for trial order or drift effects; these tests showed no significant trial order effects and no significant drifts in baseline activity (averaged for each participants, the linear slopes, <0.05%, were not significantly different from zero).

A higher temporal resolution analysis focused on RMS values. First, we addressed the dynamics of muscle activity during the fixation interval, by comparing two intervals of 200 ms each, one from the beginning of the fixation epoch (Fix1) and one from the end (Fix2) (Figure 4). In both conditions and in both muscles there was lower activation levels in Fix2 compared to Fix1; in the OOI, the difference between the two fixation intervals was greater in the Overt than in the Covert condition. We next compared muscle activity (relative to Fix2) in three intervals of 200 ms each (Figure 5B): an interval related to the onset of sentence presentation (interval A), an interval related to the offset of the sentence presentation (interval B), and an interval just before the repetition cue (interval C) (Figure 5B). RMS values of sEMG activity (μV) were computed for each of the intervals, and a relative activity measure was computed by subtracting the RMS (μV) value of Fix2.

# References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931.

Allen, P., Aleman, A., & Mcguire, P. K. (2007). Inner speech models of auditory verbal hallucinations: Evidence from behavioural and neuroimaging studies. *International Review of Psychiatry*, *19*(4), 407-415.

Augustinova, M., & Ferrand, L. (2014). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science*, *23*(5), 343-348.

Brodin, P., Türker, K. S., & Miles, T. S. (1993). Mechanoreceptors around the tooth evoke inhibitory and excitatory reflexes in the human masseter muscle. *The Journal of physiology*, *464*(1), 711-723.

Brumberg, Jonathan S., et al. "Spatio-Temporal Progression of Cortical Activity Related to Continuous Overt and Covert Speech Production in a Reading Task." *PloS one* 11.11 (2016): e0166872.

Conrad, B., Benecke, R., & Goehmann, M. (1983). Premovement silent period in fast movement initiation. *Experimental Brain Research*, *51*(2), 310-313.

Duque, J., Greenhouse, I., Labruna, L., & Ivry, R. B. (2017). Physiological markers of motor inhibition during human behavior. *Trends in neurosciences*, *40*(4), 219-236.

Fadiga, L., Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current opinion in neurobiology*, *15*(2), 213-218.

Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. Hove (UK): Lawrence Erlbaum Associates.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *cortex*, *48*(7), 905-922.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*(961), 940-944.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407-422.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.

Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the ecological validity of thinking on demand: neural correlates of elicited vs. spontaneously occurring inner speech. *PloS one*, *11*(2).

Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities. *American Journal of Physiology--Legacy Content*, *97*(1), 200-209.

Kolářová, B., Krobo, A., Kolář, P., Hluštík, P., & Polehlová, K. (2016). Muscle Activity Inhibition in Lower Limb Muscles during Gait Imagery Tasks. *Current Research in Motor Control V*, 123.

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. *Coarticulation: Theory, Data and Techniques*, , 7-30.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261-273.

Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776-785.

Lebon, F., Greenhouse, I., Labruna, L., Vanderschelden, B., Papaxanthis, C., & Ivry, R. B. (2015). Influence of delay period duration on inhibitory processes for response preparation. *Cerebral Cortex*, *26*(6), 2461-2470.

Leuthardt, E., Pei, X. M., Breshears, J., Gaona, C., Sharma, M., Freudenburg, Z., ... & Schalk, G. (2012). Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. *Frontiers in human neuroscience*, *6*, 99.

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America*, *29*(1), 117-123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perceptual and motor skills*, *83*(3 suppl), 1355-1362.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 123-146.

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *The Pavlovian journal of biological science*, *24*(1), 19-26.

Pei, X., Leuthardt, E. C., Gaona, C. M., Brunner, P., Wolpaw, J. R., & Schalk, G. (2011). Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *Neuroimage*, *54*(4), 2960-2972.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2).

Poeppel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, *92*(1), 1-12.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351-360.

Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Lœvenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, *56*(6).

Richartz, C., Lévénez, M., Boucart, J., & Duchateau, J. (2010). Initial conditions influence the characteristics of ballistic contractions in the ankle dorsiflexors. *European journal of applied physiology*, *110*(4), 805-814.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131-141.

Rozanov, S., Keren, O., & Karni, A. (2010). The specificity of memory for a highly trained finger movement sequence: Change the ending, change all. *Brain Research*, *1331*, 80-87.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295-302.

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain and language*, *164*, 77-105.

Stepp, C. E. (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation. *Journal of Speech, Language, and Hearing Research : JSLHR*, *55*(4), 1232-1246.

Stiegler, L. N. (2015). Examining the echolalia literature: where do speech-language pathologists stand? *American Journal of Speech-Language Pathology*, 24(4), 750-762.

Whiting, C., Shtyrov, Y., & Marslen-Wilson, W. (2014). Real-time functional architecture of visual word recognition. *Journal of Cognitive Neuroscience*.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature neuroscience*, *7*(7), 701-702.

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*(2), 592-597.

## Figure Legends

**Figure 1: Placement of the sEMG electrodes and the timeline of each trial. Panel A**: sEMG was recorded from the Orbicularis Oris Inferior (OOI) and Sternohyoid (STH) muscles (Stepp, 2012) with 4-mm miniature Beckman Ag/AgCl electrode pairs (gel filled). Skin sites were prepared with Nuprep (Weaver and Company, Aurora, CO) before electrode application. **Panel B**: The sequence of events on each trial – a fixation target (2 sec) preceded stimulus presentation (fixation segment); a stimulus presentation segment of variable duration (with randomly chosen time jitter wherein the fixation cross reappeared, total 2.2 – 2.8 sec); a Go signal (0.5 sec) and a response interval (ending with a space bar press).

**Figure 2 - sEMG activity separately for the covert and overt conditions of the task**. OOI – Orbicularis Oris Inferior, left panels; STH - Sternohyoid, right panels. Each panel presents a single trial. A,B,C: auditory, visual whole-sentence, and visual word-by-word presentation, respectively, for covert repetition; D, E, F: auditory, visual whole-sentence, and visual word-by-word presentation, for overt repetition. The vertical lines indicate (from left-to-right) the fixation onset, sentence presentation onset, the GO cue onset, repetition completion.

**Figure 3: Group average ARV (left panels) and RMS (right panels) difference values (vis-à-vis fixation epochs)**. Panels A, C - Orbicularis Oris Inferior (OOI) muscle; Panels B, D Sternohyoid (STH) muscle. A: ARV difference values were significantly lower during sentence presentation prior to covert compared to overt intended responses (mean difference: 1.466 µV, P=.002). B: ARV difference values were significantly lower during sentence presentation prior to overt compared to covert intended responses (mean difference: .103 µV, P<.001). C**:** RMS difference values were significantly lower during sentence presentation prior to overt compared to covert intended responses (mean difference: 1.911 µV, P=.008); the auditory sentence presentation condition resulted in significantly lower RMS difference values compared to the visual whole-sentence presentation condition (mean difference: .326 µV, p=0.01). D: RMS difference values were significantly lower during sentence presentation prior to overt compared

to covert intended responses (mean difference: .139 µV, P<.001). *P < 0.05, **P < 0.01, ***P < 0.001: difference from 0. Horizontal lines - main effect. Error bars - standard errors.

**<u>Figure 4</u>: Group average sEMG activity (RMS) at 2 intervals during the fixation epoch.** Data are shown for the 2 intended outputs, Overt (Ov) vs. Covert (Cov) in 2 intervals of the fixation epoch (Fix_1 – initial 200 ms ; Fix_2 – final 200ms)**.** In the OOI muscle (left panel), there was a significant interaction of Intended Output and Fixation interval, F(1,19 )=36.07, P<.0001, $\eta_p^{2=}$ .655) reflecting no difference between the 1st and 2nd fixation intervals in the Covert condition (P>.18) but a large, significant effect in the Overt condition (F(1,19)=34.14, P<.000, $\eta_p^2$=.642). In the STH muscle (right panel) there was a main effect of Intended Output ($F_{(1,19)}$=9.707, P<.01, $\eta_p^2$=.338), with more activation prior to overt speech (mean difference=.433 , P<.01). There was also a main effect of Fixation interval ($F_{(1,19)}$=13.589, P<.01, $\eta_p^2$=.417), with more activation in the 1st fixation interval compared to the 2nd fixation interval (mean difference=.149 , P<.01). * P < 0.05, ** P < 0.01, *** P < 0.00. Error bars - standard errors.

**<u>Figure 5</u> – Higher temporal resolution analysis. A**: The 5 intervals used for the higher resolution analyses are shown: 2 fixation intervals, and intervals A, B and C. The temporal jitter (the delay period) determined the interval between B and C. **B**: RMS difference values (vis-a-vis the 2nd fixation interval) are shown separately for each Intended Output condition and for each of the Input Type conditions (auditory presentation – A; visual whole sentence presentation - V1, visual word-by-word sentence presentation - V2). Oribicularis Oris Inferior (OOI) (panels 1 and 2); Sternohyoid (STH) muscle (panels 3 and 4). Panel **1**: intended covert response. There was a significant main effect of Interval ($F_{(2,38)}$=6.600, P=.003, $\eta_p^2$=.258) with activity in interval B significantly lower compared to interval A (mean difference=.195 µV, P=.021) and activity in interval C significantly lower compared to interval A (mean difference=.166 µV, P=.04). Panel **2**: intended overt response. There was a marginally significant main effect of Input Type ($F_{(2,38)}$=2.858, P=.07, $\eta_p^2$=.131), a significant main effect of Interval (Greenhouse-Geisser corrected) ($F_{(2,38)}$=5.695, P<.05, $\eta_p^2$=.231), and a significant interaction (Huynh-Feldt corrected) ($F_{(3.253, 61.807)}$=3.727, P<.05, $\eta_p^2$=.164). Post hoc Bonferroni analyses showed a decrease in

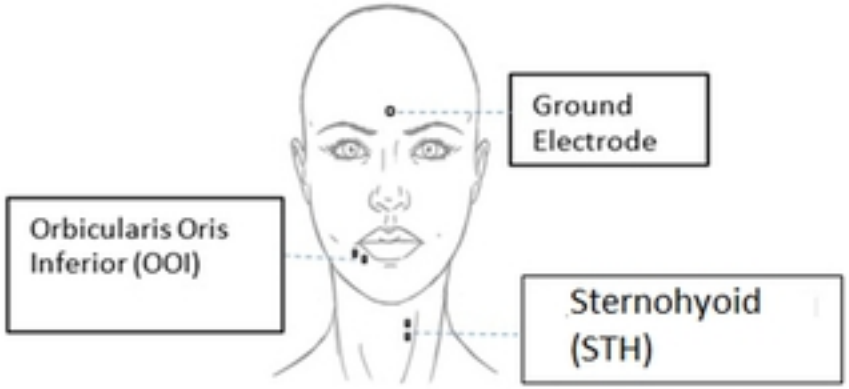
inhibition in interval C compared to interval A (mean difference= 1.230 µV, P=.035), and compared to interval B (mean difference=.764, µV, P=.027). Panel **3**: intended covert response. There was a significant main effect of Interval (Huynh-Feldt corrected) ($F_{(1.305,24.801)}$=4.302, P<.05, $\eta_p^2$=.185) with interval B marginally lower compared to interval A (mean difference=.034 µV, P=.097). Panel **4**: intended overt response. There were no significant main effects or interaction. * P < 0.05, ** P < 0.01, *** P < 0.001. Error bars - standard errors.

## Table Legends

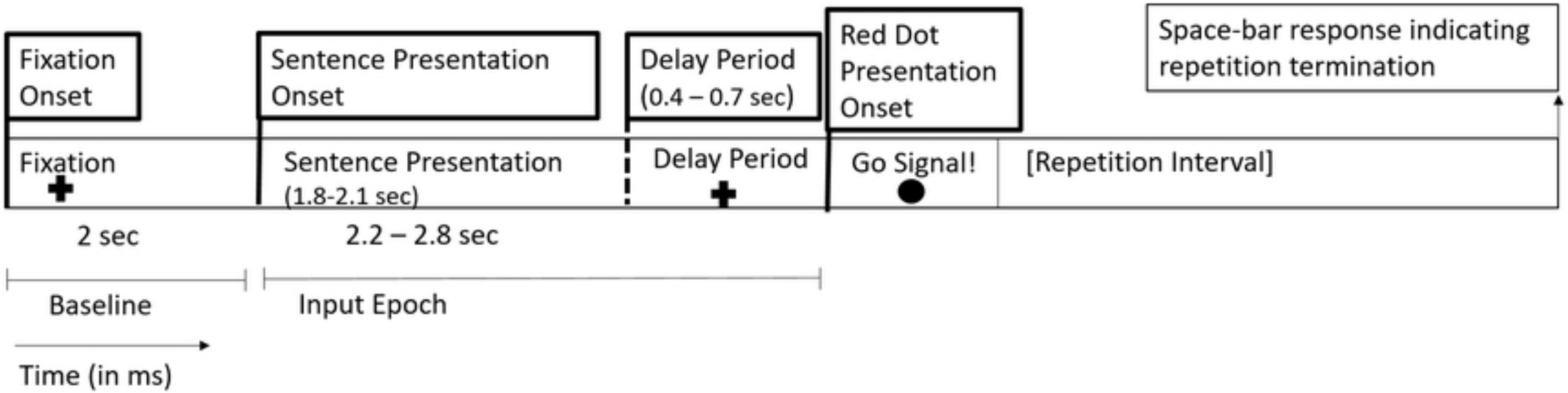
**Table 1 –** Mean and standard deviation for the average rectified value (ARV), and the root mean square (RMS) of the sEMG activity (µV) within each epoch (fixation and stimulus presentation), in each muscle (Orbicularis Oris Inferior – OOI, and Sternohyoid - STH). The ARV and RMS are shown for each of the six conditions (A_Cov = Auditory presentation, covert repetition intended; V1_Cov = visual whole-sentence presentation, covert repetition intended; V2_Cov = visual word-by-word sentence presentation, covert repetition intended; A_Ov = Auditory presentation, overt repetition intended; V1_Ov = visual whole-sentence presentation, overt repetition intended; V2_Ov = visual word-by-word sentence presentation, overt repetition intended).

**Figure 1**

**A**

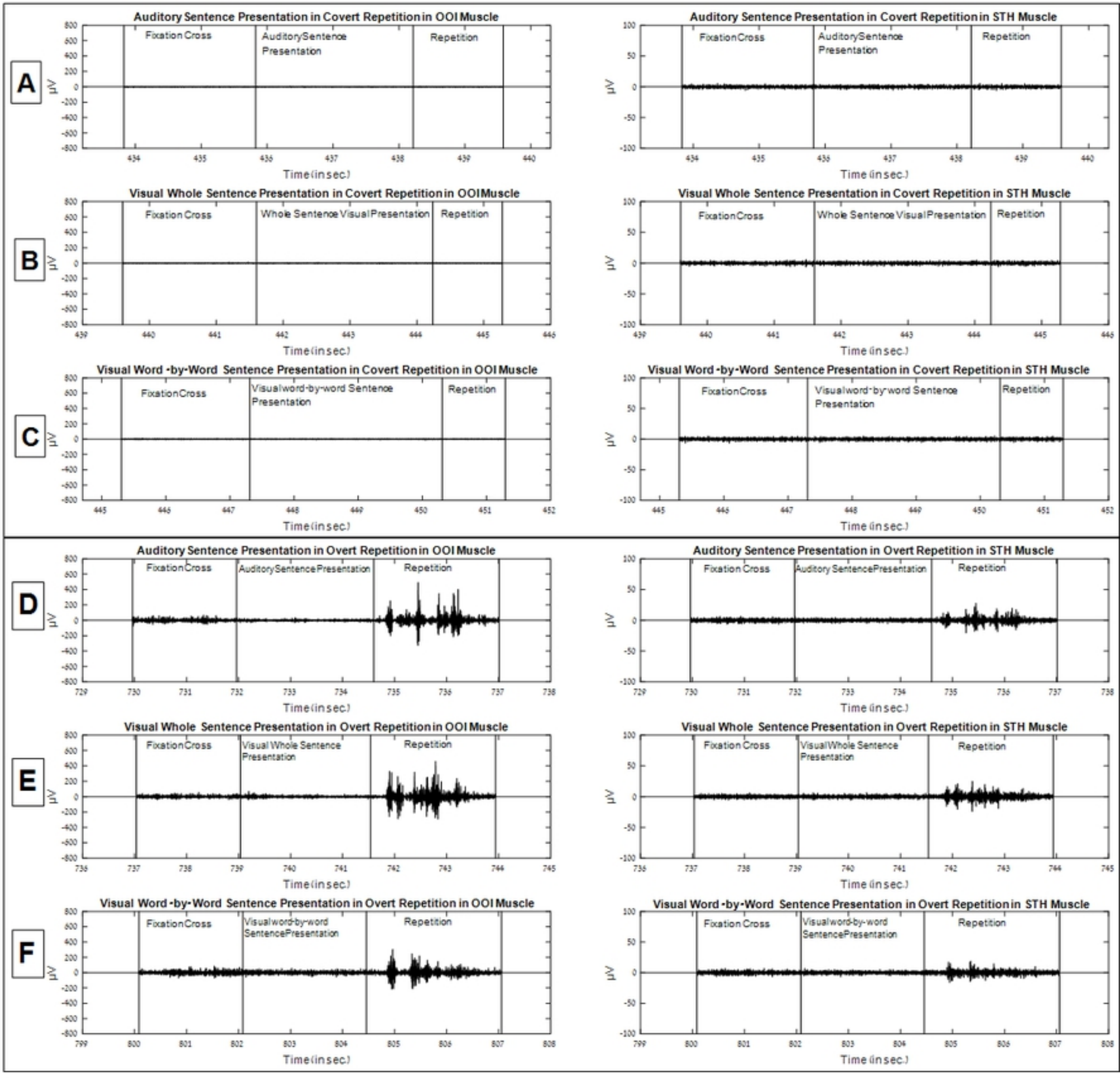



**B**



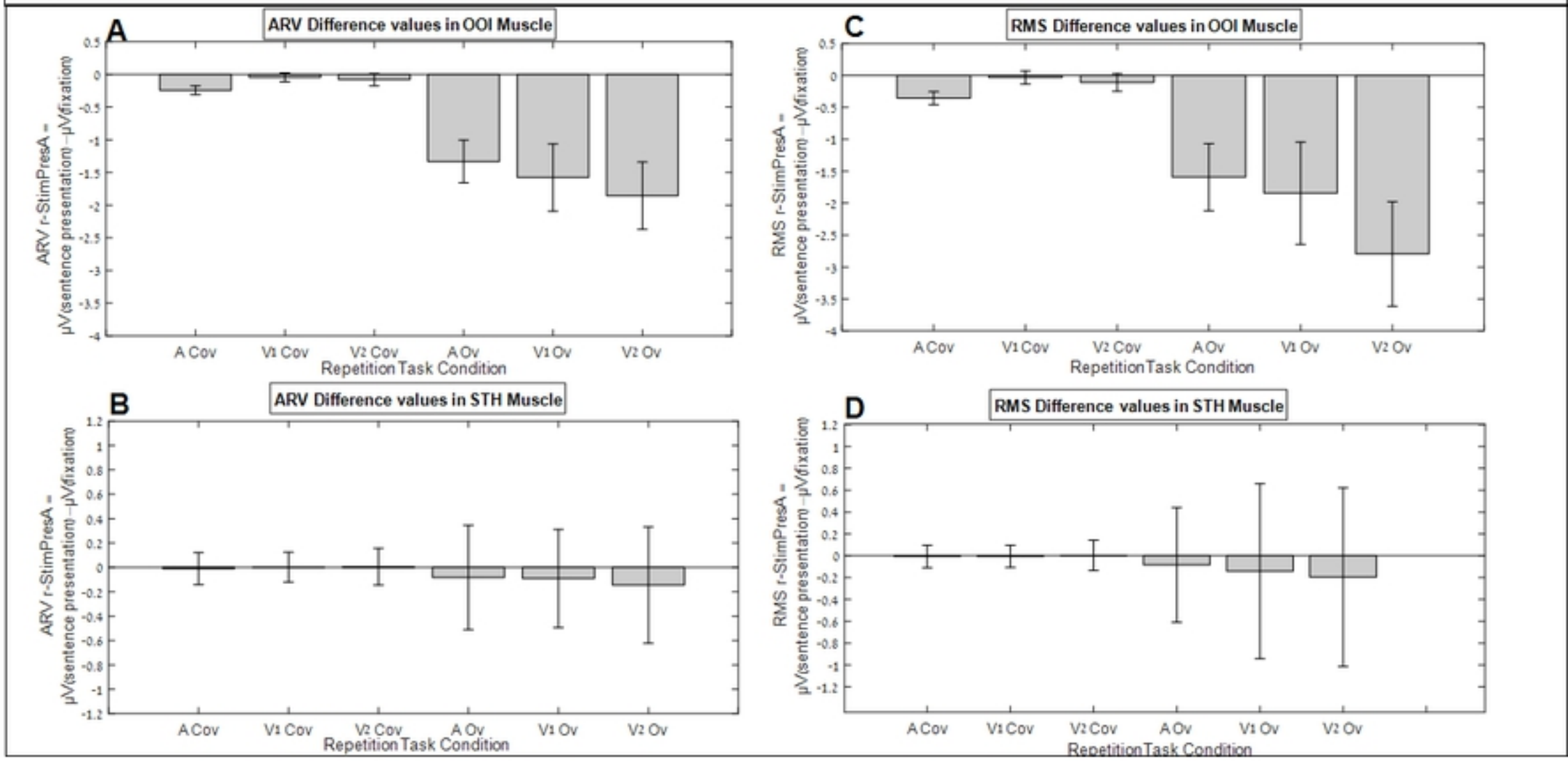

Figure

# Figure 2



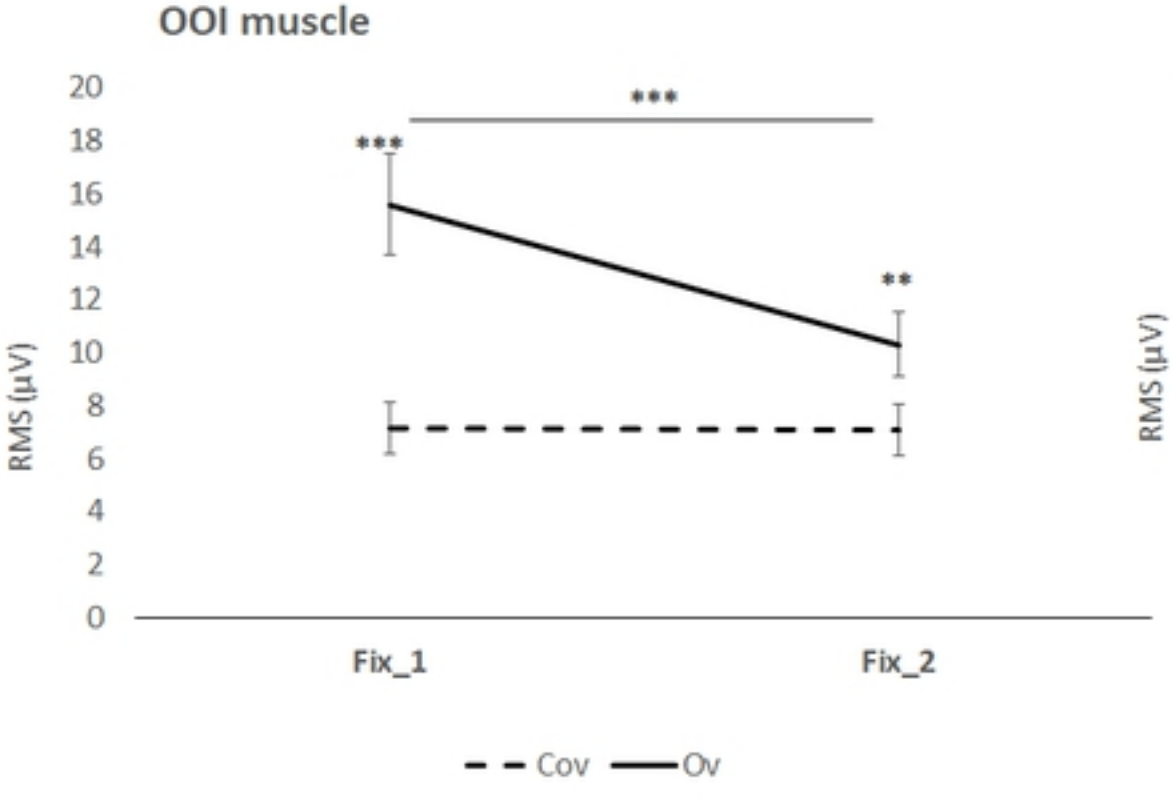

Figure

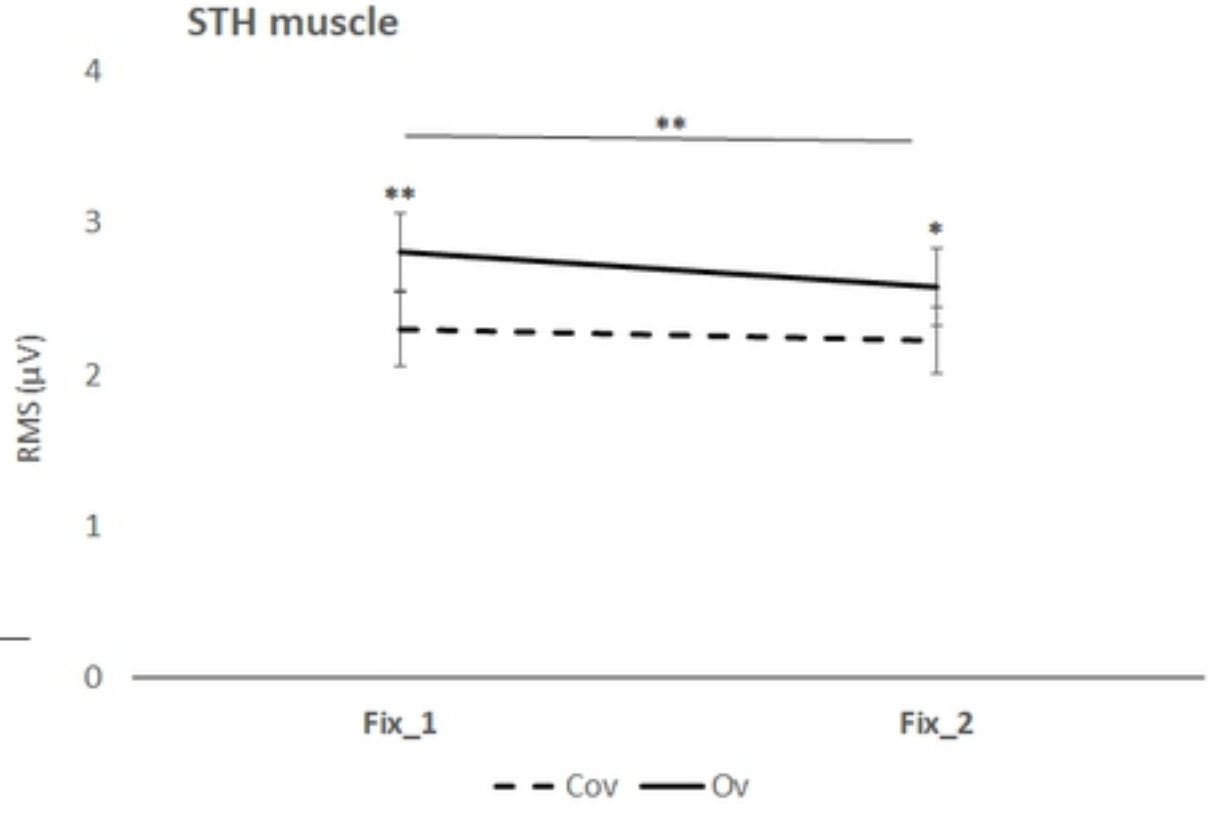
# Figure 3




Figure

**Figure 4**

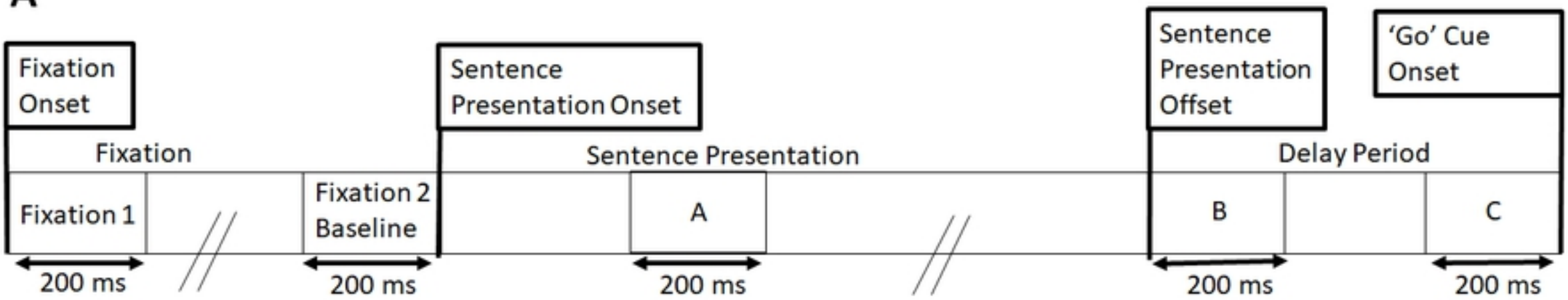


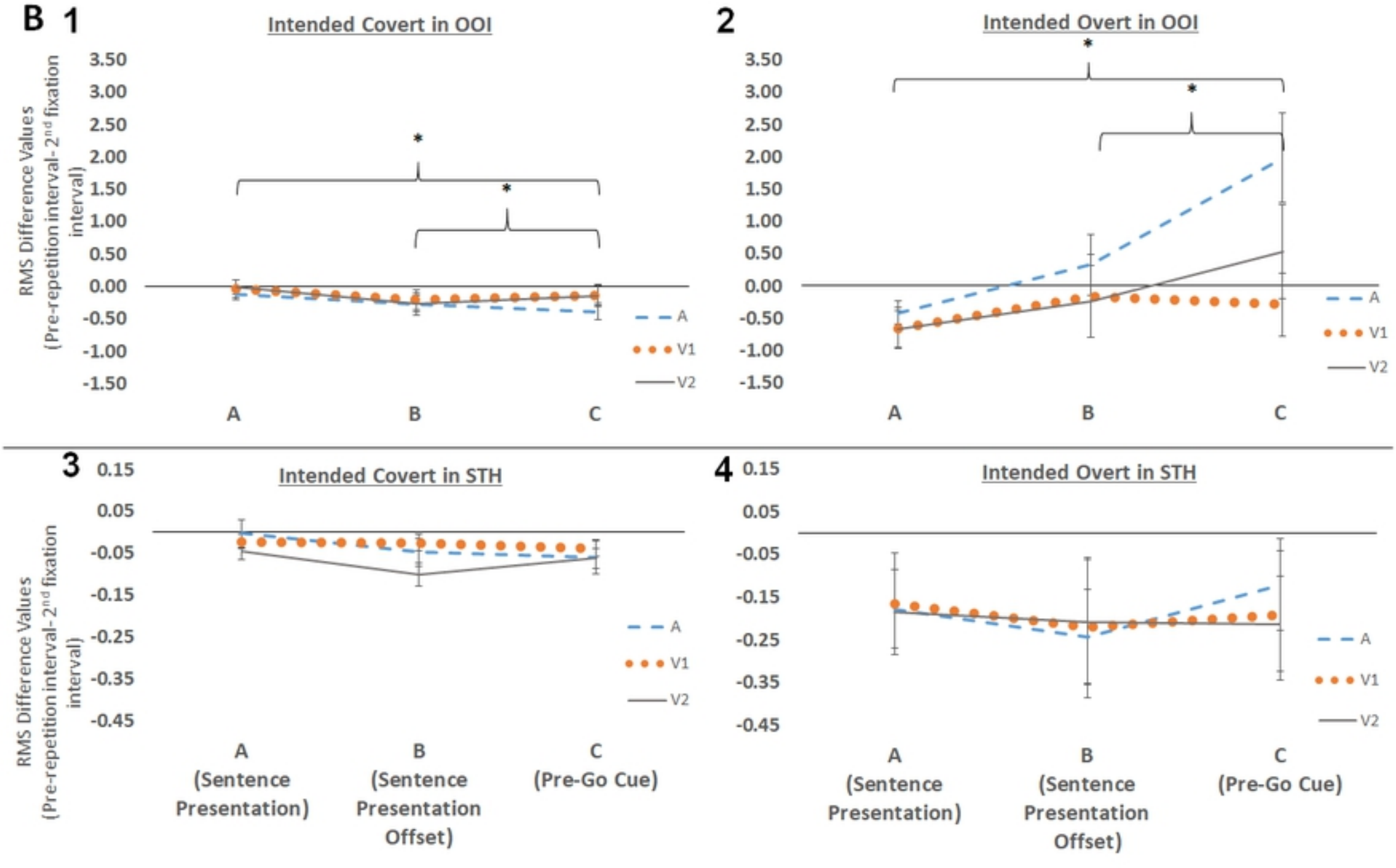

Figure

**Figure 5**

**A**




**B**




Figure

**Table 1**

| Trial Epochs | Condition | RMS (OOI) mean (s.d.) | ARV (OOI) mean (s.d.) | RMS ( STH ) mean (s.d.) | ARV (STH) mean (s.d.) |
|---|---|---|---|---|---|
| Fixation | A_Cov | 7.44 (5.09) | 5.70 (3.84) | 2.30 (1.10) | 1.77 (0.79) |
| | V1_Cov | 7.39 (5.09) | 5.63 (3.78) | 2.29 (1.06) | 1.77 (0.76) |
| | V2_Cov | 7.65 (5.22) | 5.76 (3.79) | 2.26 (1.08) | 1.76 (0.78) |
| | A_Ov | 13.31 (9.81) | 9.31 (6.36) | 2.83 (1.62) | 2.10 (0.98) |
| | V1_Ov | 14.02 (10.84) | 9.55 (6.59) | 2.80 (1.48) | 2.08 (0.91) |
| | V2_Ov | 14.32 (10.73) | 9.64 (6.46) | 2.82 (1.52) | 2.08 (0.89) |
| Sentence Presentation | A_Cov | 7.05 (5.05) | -0.23 (0.77) | 2.28 (1.14) | -0.01 (0.13) |
| | V1_Cov | 7.29 (5.14) | -0.04 (0.77) | 2.28 (1.06) | 0.00 (0.12) |
| | V2_Cov | 7.76 (6.22) | -0.08 (1.25) | 2.24 (1.08) | 0.01 (0.15) |
| | A_Ov | 11.52 (7.82) | -1.33 (3.74) | 2.64 (1.39) | -0.08 (0.43) |
| | V1_Ov | 11.98 (8.84) | -1.55 (4.48) | 2.68 (1.42) | -0.09 (0.40) |
| | V2_Ov | 11.26 (7.68) | -1.84 (4.14) | 2.61 (1.35) | -0.15 (0.48) |

Figure