Speaking-induced Middle Ear Muscle Reflex (MEMR): suppression of auditory feedback during self-vocalization

Hayo Terband[1], Caroline Cross[1,2], Joel Berger[3], Shawn Goodman[1]

[1]*Department of Communication Sciences and Disorders, University of Iowa, Iowa City IA, USA*
[2]*Hearing Associates, Mason City IA, USA*
[3]*Department of Neurosurgery, University of Iowa, Iowa City IA, USA*

hayo-terband@uiowa.edu, ccross@hearingassociatesmc.com, joel-berger@uiowa.edu, shawn-goodman@uiowa.edu

Abstract

The aim of the present study was to test a novel, non-invasive method to measure the middle ear muscle reflex (MEMR) during self-vocalization compared to when presented with external speech, to allow the quantification of its contribution to the reduced response in auditory cortex found in speaking-induced suppression (SIS). MEMR responses were measured utilizing the principle of change in impedance and thus reflectance characteristics of the eardrum, quantified by amplitudes of continuously presented low-level click sounds. Data from 15 healthy young adult speakers show 1) that MEMR was activated prior to the onset of speech, but not prior to the playback of the recorded utterances and 2) that MEMR is stronger for self-generated sound. As MEMR reduces low-frequency excitation in the cochlea, SIS may need to be corrected for MEMR.

Keywords: *speech production, speech synthesis*

1. Introduction

Corollary discharge (CD) is an umbrella term for brain functions that allow animals to differentiate external from self-generated sensory signals and encompasses both lower- and higher-order mechanisms, depending on their function (Crapse & Sommer 2008; Sperry 1950; Holst & Mittelstaedt 1950). Lower-order mechanisms concern the control of sensation by the Central Nervous System (CNS) and include sensory filtration and reflex inhibition; higher-order mechanisms concern the control of action and perception and include sensory analysis/stability and sensorimotor learning/planning (Crapse & Sommer 2008).

One higher order mechanism relevant to human speech production is speaking-induced suppression (SIS), the phenomenon of a reduced response in auditory cortex to self-produced compared to externally-produced speech (Numminen & Curio 1999; Houde et al. 2002; Greenlee et al. 2011). SIS is thought to be triggered by the efference copy from motor cortex containing a forward prediction of the sensory consequences of the motor program (Knille et al. 2019; Ylinen et al. 2015) and/or the sensory goals associated with the motor plan (Niziolek et al. 2013). The mechanism is thought to play an important role in error detection, error correction and speech-motor learning.

Largely ignored in the field of speech production but well-studied in audiology are two major efferent feedback pathways to the auditory periphery: the middle ear muscle reflex (MEMR) reflex and the medial olivocochlear reflex (MOCR; see **Figure 1**). These lower-order CD mechanisms, particularly MEMR, are of interest in the context of SIS. MEMR involves the contraction of the intratympanic muscles, which increases the stiffness of the ossicular chain, thereby altering the acoustic impedance (Metz 1952), particularly below about 1.5kHz, which in turn reduces the input to the cochlea at these frequencies. In quiet environments such as experimental lab conditions, reduced input changes the response in the auditory cortex (Herrmann et al. 2020). Clinical MEMR thresholds to external stimuli are relatively high; about 75dB-SPL for noise and 90dB-SPL for pure-tones in healthy listeners (Liberman & Guinan 1998). Perhaps because of this, a common misconception is that MEMR is irrelevant for normal conversational speech, with voice levels of 60-70dB-SPL. However, electromyography (EMG) data have shown that MEMR can also occur without acoustic stimulation during (and in anticipation of) vocalization at normal vocal effort (Borg & Zakrisson 1975). Furthermore, the effect is stronger during self-vocalization than when presented with external speech (Borg & Zakrisson 1975). SIS is determined by subtracting the magnitude of cortical responses during speaking from the response magnitude during listening to playback of the same speech signal. Since it alters the signal input from the periphery, MEMR thus forms a major confound for SIS measurement. The current study features a novel, non-invasive method to measure MEMR that allows isolation and quantification of the MEMR component of SIS.

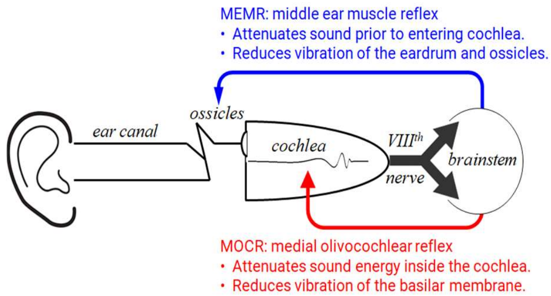

Figure 1: *Two lower-order mechanisms of corollary discharge: the middle ear muscle reflex (MEMR) reflex and the medial olivocochlear reflex (MOCR).*

2. Methods

2.1 Participants

Fifteen young adult speakers of American English (11 females, 4 males; age range = 18–25 years) with normal hearing and speech participated in the study. The first five participated in pre-pilot and development.

10.21437/issp.2024-30

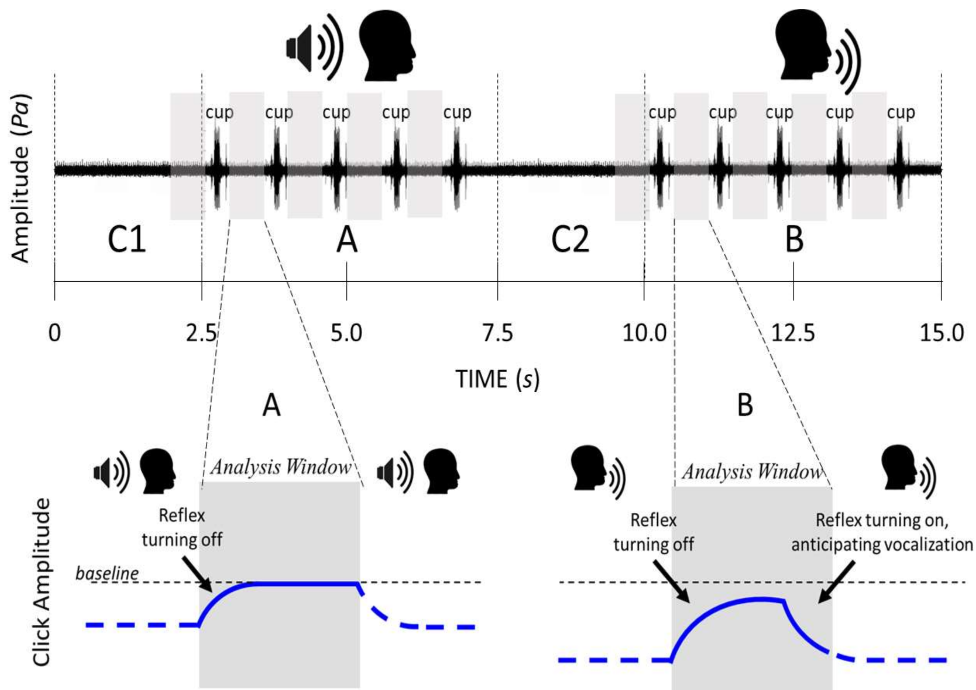

Figure 2: *Experimental paradigm and predicted results. In response to* ***external*** *sound (A), the reflexes will turn on after onset, with a delay of ~100ms. The reflexes will turn off after the sound stops with the same delay. In response to* ***self-produced*** *sound (B), the reflexes will be activated more strongly and prior to onset.*

2.2 Procedures

Participants were seated in a sound-treated booth; stimuli were presented, and ear canal pressure measured binaurally using a 2-channel probe-microphone system (ER10X, Etymotic Research). The experiment consisted of 110 trials, each consisting of four conditions: *Listen*, in which subjects listened to the recording of their own voice playing back the word "cup" five times, a*Speak*, in which subjects were visually cued to produce the word "cup" five times, with a 2.5s inter-stimulus interval, and two *Baseline* conditions. During both *Listen* and *Speak* trials, a train of low-level clicks were played continuously (**Figure 2**). Sound pressure levels of stimuli were equivalent. Baseline conditions, containing clicks only, were interleaved with the *Listen* and *Speak* conditions. The stimulus played in the Listen condition was an ear canal recording of the participant speaking, adjusted so that ear canal sound pressures during Listen and Speak conditions were approximately the same.

2.3 Data processing and analysis

The time-courses and magnitudes of MEMR responses were quantified by measuring the changing amplitudes of the click sounds reflected by the eardrum during the inter-stimulus intervals, compared to the baseline conditions.

3. Results

Figure 3 presents changes in recorded waveform magnitude averaged over five 100 ms time windows during the inter-

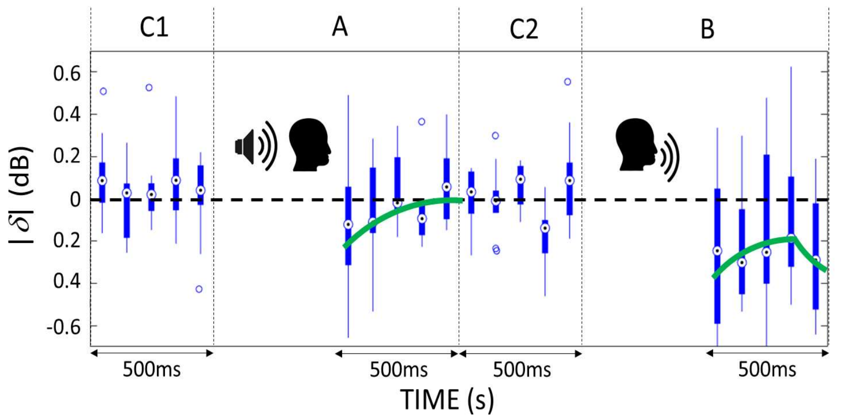

Figure 3*: Changes in recorded waveform magnitude (|δ|) as a function of time within each condition. At each time point, box and whisker plots show the median changes of the group of participants (N=10; see Fig. 5 to observe all individual data points). Open circles with black dots in the center show the group median values. Thicker blue bars show the second and third quartiles. Open blue circles indicated individual participant median values greater or less than 1.5 times the interquartile range. Exponential trend curves of MEMR activation (green lines) show the changing median activation across the 500 ms analysis windows in the conditions Listen (A), Speak (B) compared to baselines (C1 & C2).*

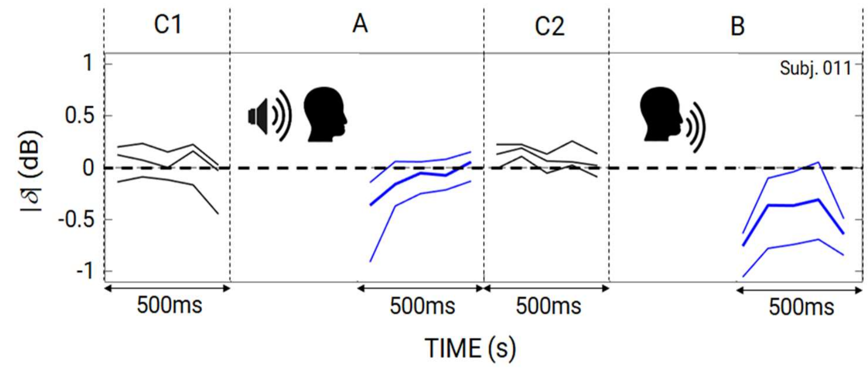

Figure 4: *Example of MEMR activation curves for an individual participant (subj011) as a function of time within each condition. Median changes in recorded waveform magnitude (|δ|) are shown by the thicker center lines, with interquartile range being shown by slightly thinner lines above and below.*

stimulus intervals in *Listen* and *Speak* conditions compared to baselines. An example of MEMR activation curves for an individual subject is presented in **Figure 4**.

Because the data had a less than optimal signal-to-noise ratio, median values were considered when fitting and reporting the data. The filled and open circles in **Figure 5** show the median values of the 10 participants at five time points in each of the four conditions. In each condition, the data for the five-point time series were fit with a straight line (*N*=50 data points). The Matlab Curve Fitting toolbox (R2023b) was used, along with the least absolute residuals (LAR) robust fitting option. The LAR method minimizes the absolute difference of the residuals, rather than the squared differences, so that extreme values have a lesser influence on the fit. For the two control conditions (C1 and C2), the 95% confidence intervals for the slopes and the intercepts contained zero. For the two test conditions (A and B), the 95% confidence intervals for the slopes contained zero, but the intercepts did not. Taken together, the results suggest the test conditions produced decreases in ear canal magnitude compared to baseline. Further, the intercept for condition B (-0.376 dB) was outside of the 95% confidence interval for condition A (-0.307 dB to -0.029 dB), highlighting that the *Speak* condition (condition B) produced a larger effect of the MEMR than the *Listen* condition (condition A).

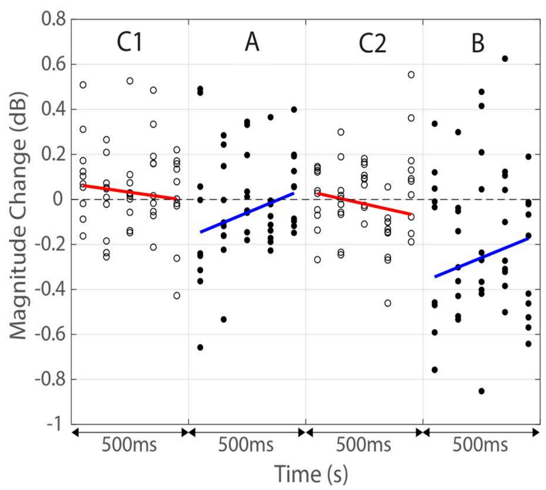

Figure 5: *Changes in recorded waveform magnitude (|δ|) as a function of time within each condition Each circle shows data from a participant. Open circles show control conditions, and filled circles show test conditions. The red and blue lines show best fits to the data.*

4. Discussion and conclusion

Auditory self-monitoring plays an important role in the acquisition and maintenance of intelligible speech. An important mechanism in auditory self-monitoring is SIS, the phenomenon of a reduced response in auditory cortex to self-produced compared to externally-produced speech, putatively triggered by the efference copy from motor regions (Knille et al. 2019; Ylinen et al. 2015). SIS has been found to be deviant in a variety of prevalent disorders such as stuttering (e.g., Beal et al., 2011; Toyomura et al., 2020) and Parkinson's disease (Mollaei et al., 2019; Huang, et al., 2016) and is thought to play an important role in the online adaptation of vocalizations to environmental conditions and speech-motor learning. However, speech studies have focused on auditory cortical activation while alternative peripheral mechanisms of corollary discharge have been largely ignored.

The present study investigated the MEMR, a peripheral mechanism of that might be of particular relevance in this context. The MEMR involves the contraction of the middle ear muscles, which stiffens the ossicular chain and reduces the low-frequency input to the cochlea. Intracellular EMG of action potentials in the stapedius muscle has shown that the MEMR occurs during (and in anticipation of) vocalization and is stronger during self-vocalization than when presented with external speech (Borg & Zakrisson 1975). Since this peripheral mechanism is altered during self-vocalization, it would have bottom-up effects on cortical activity and thus may alter auditory self-monitoring and play a role in speech production, acquisition and relevant disabilities.

The current study explored a novel, non-invasive method that estimates the magnitude of the MEMR by measuring the sound pressure of acoustic clicks that are reflected by the tympanic membrane in-between speech stimuli. The results show that MEMR magnitude can be assessed non-invasively, both during listening and during self-vocalization. Furthermore, the results were consistent with the EMG findings of Borg & Zakrisson (1975) that the MEMR 1) is activated prior to the onset of speech, but not prior to the playback of the recorded utterances, and 2) is stronger for self-generated sound. These findings potentially have specific implications for research into SIS. As MEMR reduces the excitation amplitude in the cochlea and subsequently the response in auditory cortex, measurements of SIS may need to correct for the effect of the MEMR. Experiments involving simultaneous measurements of MEMR and cortical activation for self-vocalizations are being prepared.

5. Acknowledgements

The authors would like to express their gratitude to all participants thank for their time and effort in this study. Additionally, we would like to thank two anonymous reviewers of our conference proposal for their constructive comments.

6. References

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & Luc, F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. *Neuroimage*, *54*(4), 2994-3003.

Borg, E., & Zakrisson, J. E. (1975). The activity of the stapedius muscle in man during vocalization. *Acta oto-laryngologica*, *79*(3-6), 325-333.

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience*, *9*(8), 587-600.

Greenlee, J. D. W., Jackson, A.W., Chen, F., Larson, C.R., Oya, H., Kawasaki, H., et al. (2011) Human Auditory Cortical Activation during Self-Vocalization. *PLoS ONE 6*(3): e14744.

Herrmann, B., Augereau, T., & Johnsrude, I. S. (2020). Neural responses and perceptual sensitivity to sound depend on sound-level statistics. *Scientific Reports*, *10*(1), 9571.

Holst, E. von. & Mittelstaedt, H. (1950). The reafference principle. *Naturwissenschaften 37*, 464–467.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of cognitive neuroscience*, *14*(8), 1125-1138.

Huang, X., Chen, X., Yan, N., Jones, J. A., Wang, E. Q., Chen, L., ... & Liu, H. (2016). The impact of Parkinson's disease on the cortical mechanisms that support auditory–motor integration for voice control. *Human brain mapping*, *37*(12), 4248-4261.

Knolle, F., Schwartze, M., Schröger, E., & Kotz, S. A. (2019). Auditory predictions and prediction errors in response to self-initiated vowels. *Frontiers in Neuroscience*, *13*, 1146.

Liberman, M. C., & Guinan Jr, J. J. (1998). Feedback control of the auditory periphery: anti-masking effects of middle ear muscles vs. olivocochlear efferents. *Journal of communication disorders*, *31*(6), 471-483.

Metz, O. (1952). Threshold of reflex contractions of muscles of middle ear and recruitment of loudness. *AMA Archives of Otolaryngology, 55*(5), 536-543.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, *33*(41), 16110-16116.

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience letters*, *272*(1), 29-32.

Sperry, R. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology, 43*, 482–489.

Toyomura, A., Miyashiro, D., Kuriki, S., & Sowman, P. F. (2020). Speech-induced suppression for delayed auditory feedback in adults who do and do not stutter. *Frontiers in Human Neuroscience*, *14*, 150.

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., ... & Service, E. (2015). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*, *25*(6), 1576-1586.