EFFECTS OF SLOWED RATE OF SPEECH ON EYE GAZE TO THE MOUTH IN YOUNG ADULTS: CLINICAL IMPLICATIONS

Virginia L. Shaw, Claire Bernar, Nik Josafatow, Tyler Bonnell, Fangfang Li

University of Lethbridge, Lethbridge, Alberta, Canada
vl.shaw@uleth.ca

ABSTRACT

The slowed rate of speech used in clinical speech therapy protocols for fluency and motor speech disorders is known to confer improvements in motor-speech coordination and speech intelligibility. We investigated whether a slowed rate of speech might also support improved speech production *indirectly*, by changing the location and duration of eye gaze behaviors of listeners in favor of the mouth, where visual cues to articulation are centered. An eye tracking protocol was designed to measure listeners' total duration of fixations to the mouth in single sentences in both regular and slowed rate conditions, and comparison blocks (regular and loud volume, sentence repetition with no change). Preliminary analysis from 25 typical young adult listeners suggests greater attention to the mouth area occurs in a slowed rate condition. The results add to supportive evidence for the clinical use of a slowed rate of speech (SR) as a therapy strategy.

Keywords: speech rate, perception, attention, gaze

1. INTRODUCTION

Voluntary production of a slowed rate of speech (SR) is well-documented as an effective speech modification strategy for persons with motor speech disorders [1], as well as fluency disorders [2]. Improvements in the smooth, forward flow of speech and/or speech intelligibility with slowed speech rate are primarily attributed to coordination benefits [3, 4] and beneficial sensory feedback [5] during slowed movements.

As an addition to the known motor coordination benefits, the current study aims to explore *whether use of a slowed rate of speech (SR) may provide important secondary benefits for speech perception and motor-speech production, by* ***spontaneously drawing visual attention*** *to the sound placement cues and movements of the oral articulators at the mouth.*

The visual information provided by movement and placement of the articulators during speech production is known to affect *speech perception* [6-9]. Attention to visual speech cues can confirm an ambiguous, incomplete or poorly perceived auditory speech signal, thereby improving message intelligibility and comprehension [6]. Fewer cognitive processing resources may be needed to perceive and understand speech with supportive visual speech cues [10]. Visual cues are known to support the speech perception abilities of hearing impaired speakers [11], and can improve typical listeners' abilities to perceive speech in background noise [6, 7], when learning speech sounds in a second language [8], and when hearing accented [9] or less intelligible speech [12].

The positive effect of adding visual speech cues for accurate perception of the auditory signal is reinforced in recent *eye tracking* studies. Yi, et al. [7] found that listeners achieve higher perceptual accuracy when they spontaneously move their eye gaze closer to the center of the mouth in difficult listening environments (i.e., low signal-noise ratio, multiple speakers), presumably to take additional advantage of mouth movements for intelligibility and speaker identity cues. Banks, et al. [13] noted that longer eye gaze fixations to the mouth improved speech perception accuracy when listening to degraded (i.e., noise vocoded) speech signals.

The visual cues from articulatory movements may also support learning of *speech sound production*. Through Infant Directed Speech (IDS), young children learning speech sounds are exposed to exaggerated articulatory movements from caregivers [14]. In addition, adults have been found to produce "hyperarticulations" of speech sounds in response to less intelligible speech or speech sound production errors [15] by preschool children, providing "both enhanced input to children and an error-corrective signal" (p. 1836). As hyperarticulation creates an audible change in the speech signal, it is likely that a change in the visible movement component of the oral articulators would also occur.

Finally, in clinical speech-language pathology practice, speech therapists routinely use cued visual attention to the mouth to enhance their clinical models and provide effective feedback, particularly in speech sound treatment protocols. Cueing methods may include: verbal prompts to "look at my mouth," use of a mirror for clients to see their own articulatory movements [16], and use of specific hand shapes or touch cues at the mouth area on the clinician's (or client's) face to draw client's

observation and imitation of multisensory cues during speech sound production attempts [17].

Given the benefits for speech perception and speech production learning provided by the visual cues just described, we proposed that the mechanisms for *natural visual attention* may provide a rationale for why a clinically slowed rate of speech would draw more visual attention to the mouth than a regular speech rate. Visual attention can be described as "stimulus-driven" [18], wherein the perceptual characteristics of stimuli (across modalities) are subconsciously evaluated and visual attention is drawn to more salient stimuli [19]. One of the high saliency conditions, that of "novelty" or "surprise", is created when sensory input differs from prior experience or expectations (i.e., stronger, atypical, new, or unpredictable information) [20]. Evidence shows that "surprise is a strong attractor of human attention" (p. 1295), as measured via eye gaze [20].

With respect to speech rate, adult listeners have been found to have expectations for prosodic features based on their experience with both the particular language and speaker [21], providing a baseline or foundation for "typical" prosody characteristics such as speech rate, volume, and pitch contours. Therefore, use of a clinically slowed rate of speech, which is significantly less than the lower boundaries of typical speech rates [22], would be expected to create a "surprise" effect. Moreover, unexpected changes in movement features and timing have been found to draw visual attention [23, 24]. As the mouth is the *source* of the slowed speech signal (both extended movement and auditory components), the "novelty" or surprise effect would likely focus on the mouth. Once visual attention is focused at the mouth, all of the possible benefits of visual speech cues for speech perception and production would be available.

With the foundation of a significantly slowed rate of speech as a novel or surprising stimulus condition, we hypothesized that typical young adult participants would increase their eye gaze to the mouth area significantly more during the Slowed Rate (SR) condition than in the Regular Rate (RR) condition. A Volume condition was added to allow comparison of the effects on visual attention of a less novel change in stimuli. Our study accepted the key assumption that measurement of the location and duration of visual fixation(s) of the eyes will generally relate to the focus of a person's visual attention [25].

2. METHODS

2.1. Study design

A screen-based eye tracker was used to record participants' eye gaze behaviours while passively viewing a video-recorded model speaking single sentences aloud. Three (3) blocks were presented to all participants (within-subject, repeated measures design). Each block contained 9 novel sentences (control condition) and 9 repeated sentences (manipulated condition). Sentence stimuli order was randomized within blocks 2 & 3. The order of blocks 2 & 3 was alternated.

- Block 1 - **Repetition (Rep)**: sentences repeated unchanged (Rep A, Rep B), Rep A always presented before Rep B, 200 syllables per minute (i.e., spm)
- Block 2/3 – **Volume (Vol)**: Regular Volume, Loud Volume (+6 dB) [26], 200 spm
- Block 3/2 - **Rate**: Regular Rate (200 spm), Slowed Rate (120 spm)

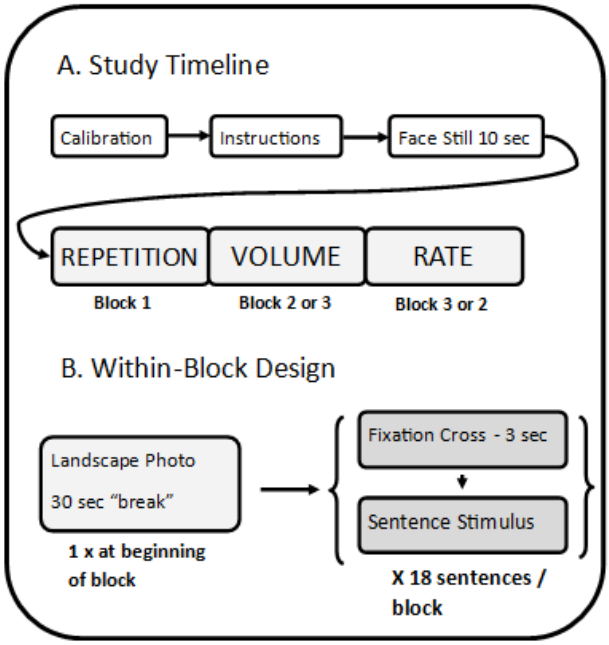

Figure 1: Depiction of study timeline and example block design.

2.2 Participants

A convenience sample of 32 young adults was recruited through a program for psychology student involvement in research at the University of Lethbridge. Twenty-five participants (15 females, 10 males, aged 19-31 years, mean = 22.2 years, sd $\pm$ 3.0) met pre-selected inclusion criteria: English as a first language, normal or corrected-to-normal vision, and no history of speech, language, learning or hearing difficulties (self-reported). Pre-study ethics approval was obtained from the Research Ethics Board of the University of Alberta. Participants

provided voluntary written consent prior to testing and completed a demographic questionnaire.

2.3. Stimuli design & recording

Sentence stimuli were designed to be linguistically consistent (sentence structure, length, gender-neutral subject) and were produced with natural, but neutral intonation.

The slowed speech rate (SR) was achieved through differentiation in the prolongation of vowel sounds [4], to maintain the relative syllable durations consistent with spoken English pronunciation and stress patterns. Continuing consonant sounds were not increased in duration. All speech rates were standardized to include 0.5 seconds of pre-speech inspiration time for each sentence.

Sentence stimuli were video recorded on a Logitech web cam (1080 x 720p x 60fps), with online monitoring of audio recording levels and a consistent mouth-to-microphone distance of 10" (Shure Beta 87A cardioid microphone). The speaker was a 20-year-old female undergraduate student with clear articulation.

Using Tobii Pro Lab software (v.1.181) [27], 3 dynamic Areas of Interest (AOIs) were placed on each video stimulus (Eyes, Mouth, Face).

Figure 2. Designated Areas of Interest (AOIs): Eyes, Mouth, Face. Marked facial areas were not visible to participants.

2.4 Eye tracking recordings and analyses

Eye tracking recordings were performed with a Tobii Pro Fusion screen-based eye tracker and Tobii Pro Lab presentation software (v.1.181) [27]. Participants were seated at a standard distance of 65cm. Participants viewed one of four randomly selected timelines. Preset cut-off values for validation of calibration accuracy and precision (<1.0 deg) and percentage of eye gaze samples (>70%) were met for all participants included in data analysis. Using the Tobii Pro Lab preset "I-VT Fixation" filter, measurements were taken of the Total Duration of Fixations (summed) for the mouth Area of Interest (AOI) over the Time of Interest (TOI) (i.e., "Talk Time" or speech portion) for each stimulus. The amount of time participants spent viewing the mouth was compared across blocks and conditions.

To account for differences in sentence stimuli durations, the Total Duration of Fixation values were normalized to a proportion (Total Duration of Fixations to the Mouth [in ms] / Duration of "Talk Time" per stimuli [in ms] = Proportion of Total Duration of Fixations to Mouth per stimuli).

3. RESULTS

Data analyses were performed in R, v. 4.1.2 [28]. The dependent variable of interest (Normalized Total Duration of Fixation of Eye Gaze to the Mouth) was analyzed using a Bayesian hierarchical linear mixed-effects model [29], with Main effects = Block and Condition, and Random factors = Participant (intercept and slope) and Sentence (slope). A normal distribution was used to model the dependent variable and set weakly informative priors for the parameters (normal), (0,1).

(1) Total_Duration_Fixation_Mouth (Normalized) ~ Block * Condition + (1 + Block + Condition | Participant) + (1 | Sentence)

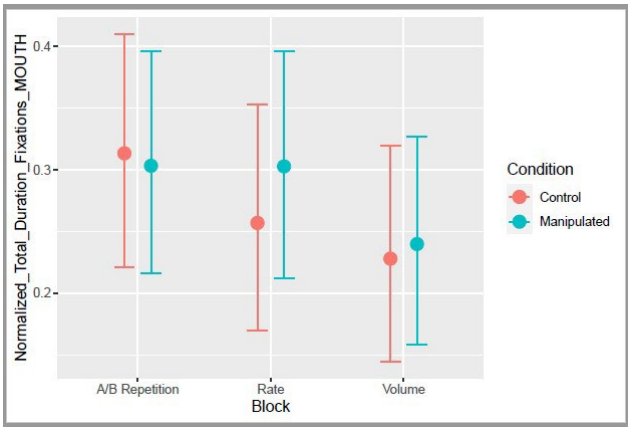

Figure 3: Conditional Effects plot: Amount of predicted change in the (duration normalized) Total Duration of Fixations to the Mouth Area. All Blocks and Conditions (Scale: 0.0 to 1.0, Control – L, Manipulated - R)

The model converged successfully with Rhat values of <1.01 (four chains, 2000 iterations). The significant interaction between Block (Rate) and Condition (Manipulated) of 0.06 (**Table 1**) indicates that when an "average" participant hears a slowed rate of speech, they look 6% more to the mouth area

than when they hear a regular rate of speech (Control condition).

Population Parameter Estimates: Change in Eye Gaze Duration to Mouth				
	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept (Block Rep) (Rep A)	0.32	0.05	0.22	0.41
BlockRate (Regular Rate)	-0.06	0.03	-0.12	0.01
BlockVol (Regular Volume)	-0.09	0.03	-0.14	-0.03
BlockRep: ConditionMan (Rep B)	-0.01	0.02	-0.05	0.03
BlockRate: ConditionMan (Slow Rate)	**0.06**	0.03	**0.00**	**0.11**
BlockVol: ConditionMan (Loud Volume)	0.02	0.03	-0.03	0.08

Table 1: Estimated parameters of the model

The non-significant Block x Condition interaction estimates in the Loud Volume condition (2% increase) and Repetition B condition (1% decrease) indicate that not all selected experimental manipulations of the speech signal result in a meaningful change in eye gaze to the mouth.

The estimated Bayesian Conditional R2 was 0.471, explaining 47% of the variability in the full model. The estimated Marginal R2 value, (excluding the Random Effects of Participant and Sentence) was 0.017. The large difference between the R2 values highlights the importance of the Random Effects in the prediction of the model parameters. With the addition of Participant as a Random Factor, the reference Intercept estimate of 0.32 (Population Level) is estimated to vary substantially (i.e., Intercept estimate for Group Level Effect of Participant: sd Estimate = 0.23, Est Error = 0.04). In contrast, only a minimal effect of including Sentence as a Random Factor was noted (Group Level Effect of Sentence: sd Estimate = 0.02, Est Error = 0.01).

4. DISCUSSION

In confirmation of our hypothesis, preliminary analysis of model results showed an increase in the duration of eye gaze to the mouth in the Slowed Rate (SR) condition compared to the Regular Rate (RR) condition for the average participant. We interpreted this as evidence of increased visual attention to the mouth due to the novel prosody and movement characteristics of a clinically slowed rate of speech (~120spm).

The smaller magnitude of the visual attention response to the mouth in the Loud Volume (LV) condition was thought to have occurred because a significantly slowed speech rate was likely more novel or surprising to listeners than a perceptual "doubling" of volume (+6dB) [26], especially when the volume changes were not paired with sentence content indicating urgency or strong emotion. In addition, the increased volume of the auditory signal was not associated with any significant changes in the movement characteristics of the speaker's mouth.

No increase in focus to the mouth was noted in the second presentation of sentences in the Repetition block. This was expected, as the unchanged sentences would likely have maintained a neutral or decreased salience value for participants.

To interpret the magnitude of the changes in duration of eye gaze fixations to the mouth (6%) with Slowed Rate we considered the context of typical eye gaze behaviours, as the amount of spontaneous gaze to the eyes is generally greater than to the mouth [30]. The preference for eye-directed gaze may be related to factors such as physiological visual biases [31] or the significance of eye contact in social interactions [32]. According to Thompson, et al. [30], there is also a strong, difficult to inhibit, natural tendency to fixate first on the eyes. This "first fixation" bias towards the eyes may be easier to sustain, rather than change to another measurement area or AOI (such as the mouth). As such, even a relatively small increase in eye gaze to the mouth, as shown in the model results, may arguably be interpreted as relevant.

Finally, with respect to potential benefits of a slowed rate of speech for modelling of speech production by clinicians, the study results are complex to interpret. Although the increase in attention to the mouth is positive, the individual variability indicated by the difference between the Conditional and Marginal R2 values suggests it may be difficult to predict how a particular individual might change their eye gaze in response to a slowed rate (SR) of speech. Future analysis of the idiosyncratic eye gaze responses of the participants in our study to SR may give insight into the breadth of personal responses to a slowed rate of speech, with individualization of clinical models to follow.

5. CONCLUSION

Study results indicated a small increase in the duration of eye gaze fixations to the mouth for typical young adult participants, when passively observing a video model of a speaker using a significantly slowed rate of speech. Our preliminary findings provide initial evidence for potential benefits of a slowed rate of speech in enhancing speech perception and/or speech production through spontaneous eye gaze to the mouth.

6. REFERENCES

[1] K. M. Yorkston, M. Hakel, D. R. Beukelman, S. Fager. 2007. Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. *J. Med. Speech-Lang. Pathology* 15, 1-26.

[2] A. K. Bothe, J. H. Davidow, R. E. Bramlett, R. J. Ingham. 2006. Stuttering treatment research 1970-2005: I. Systematic review incorporating trial quality assessment of behavioral, cognitive, and related approaches. *Amer. J. of Speech-Lang. Pathology* 15, 321-341.

[3] A. S. Mefferd, M. S. Dietrich. 2020. Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to Amyotrophic Lateral Sclerosis: Effects of loud, clear, and slow speech. *JSLHR (Online)* 63, 2625-2636.

[4] E. Boberg, D. Kully. 1994. Long-term results of an intensive treatment program for adults and adolescents who stutter. *Journal of Speech and Hearing Research* 37, 1050-1059.

[5] A. S. Mefferd. 2019. Effects of speaking rate, loudness, and clarity modifications on kinematic endpoint variability. *Clinical Linguistics & Phonetics* 33, 570-585.

[6] Y. Yuan, Lleo, Y., Daniel, R., White, A., Oh, Y. 2021. The impact of temporally coherent visual cues on speech perception in complex auditory environments. *Front. in Neuroscience* 15.

[7] A. Yi, W. Wong, M. Eizenman. 2013. Gaze patterns and audiovisual speech enhancement. *JSLHR* 56, 471-480.

[8] A. Chauvin, N. A. Phillips. 2021. Bilinguals show proportionally greater benefit from visual speech cues and sentence context in their second compared to their first language. *Ear and Hearing*.

[9] Y. Zheng, A. G. Samuel. 2019. How much do visual cues help listeners in perceiving accented speech? *Appl. Psycholing.* 40, 93-109.

[10] R. Holt, L. Bruggeman, K. Demuth. 2020. Visual speech cues speed processing and reduce effort for children listening in quiet and noise. *Appl. Psycholing.* 41, 933-961.

[11] J. B. Frtusova, N. A. Phillips. 2016. The auditory-visual speech benefit on working memory in older adults with hearing impairment. *Front. Psychol.* 7, 1-14. Art no. 490.

[12] C. K. Keintz, K. Bunton, J. D. Hoit. 2007. Influence of visual information on the intelligibility of dysarthric speech. *Amer. J. of Speech-Lang. Pathology* 16, 222-234.

[13] B. Banks, E. Gowen, K. J. Munro, P. Adank. 2021. Eye gaze and perceptual adaptation to audiovisual degraded speech. *JSLHR* 64, 3432-3445.

[14] A. Cristià. 2010. Phonetic enhancement of sibilants in infant-directed speech. *J. Acoust. Soc. Amer.* 128, 424-434.

[15] H. M. Julien, B. Munson. 2012. Modifying speech to children based on their perceived phonetic accuracy. *JSLHR* 55, 1836-1849.

[16] W. Secord. 1981. *Eliciting Sounds: Techniques for Clinicians*. The Psychological Corporation.

[17] A. K. Namasivayam, A. Huynh, F. Granata, V. Law, P. van Lieshout. 2021. PROMPT intervention for children with severe speech motor delay: a randomized control trial. *Ped. Res.* 89, 613-621.

[18] O. Räsänen, S. Kakouros, M. Soderstrom. 2018. Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition* 178, 193-206.

[19] C. Ranganath, R. Gregor. 2003. Neural mechanisms for detecting and remembering novel events. *Nature Rev. Neuroscience* 4, 193-202.

[20] L. Itti, P. Baldi. 2009. Bayesian surprise attracts human attention. *Vis. Res.* 49, 1295-1306.

[21] S. Kakouros, N. Salminen, O. Räsänen. 2018. Making predictable unpredictable with style - Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia* 109, 181-199.

[22] T. Kendall. 2013. *Speech Rate, Pause and Sociolinguistic Variation : Studies in Corpus Sociophonetics*. Palgrave Macmillan.

[23] C. J. Howard, A. O. Holcombe. 2010. Unexpected changes in direction of motion attract attention. *Attention, Perception and Psychophysics* 72, 2087-2095.

[24] A. von Mühlenen, M. I. Rempel, J. T. Enns. 2005. Unique temporal change is the key to attentional capture. *Psychol. Sci.* 16, 979-986.

[25] B. T. Carter, S. G. Luke. 2020. Best practices in eye tracking research. *Int. J. Psychophysiol.* 155, 49-62.

[26] E. Sengpiel. Decibel levels and perceived volume change. http://www.sengpielaudio.com/calculator-levelchange.htm (accessed March 28, 2022).

[27] *Tobii Pro Lab*. (2014). Tobii Pro AB, Danderyd, Sweden.

[28] *R: A Language and Environment for Statistical Computing*. (2022). R Foundation for Statistical Computing, Vienna, Austria. [Online]. https://www.R-project.org/

[29] P. Bürkner. 2018. Advanced Bayesian multilevel modeling with the R package brms. 10, 395-411.

[30] S. J. Thompson, T. Foulsham, S. R. Leekam, C. R. G. Jones. 2019. Attention to the face is characterised by a difficult to inhibit first fixation to the eyes. *Acta Psychologica* 193, 229-238.

[31] K. Guo, C. Smith, K. Powell, K. Nicholls. 2012. Consistent left gaze bias in processing different facial cues. *Psychol. Res.* 76, 263-269.

[32] T. Foulsham, J. T. Cheng, J. L. Tracy, J. Henrich, A. Kingstone. 2010. Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition* 117, 319-331.