Segmented Analysis of Eye Gaze Behaviors of Fluent and Stuttered Speech

Analyse segmentée des comportements de fixation du regard sur une élocution fluide et une élocution bégayée

KEY WORDS

STUTTER

STAMMER

FLUENCY

FLUENCY DISORDERS

COMMUNICATION

PERCEPTION

EYE GAZE

EYE TRACKING

Daniel Hudock
Andrew Stuart
Tim Saltuklaroglu
Jianliang Zhang
Nicholas Murray
Joseph Kalinowski
Nicholas Altieri

Daniel Hudock, PhD,
Department of Communication Sciences and Disorders,
Division of Health Sciences,
Idaho State University,
Pocatello, ID
U.S.A.

Andrew Stuart, PhD,
Department of Communication Sciences and Disorders, College of Allied Health Sciences,
East Carolina University,
Greenville, NC
U.S.A.

Tim Saltuklaroglu, PhD,
Department of Audiology and Speech Pathology,
University of Tennessee,
Knoxville, TN
U.S.A.

Jianliang Zhang, PhD,
Department of Allied Professions
School of Education
North Carolina Central University
Durham, NC
U.S.A.

Nicholas Murray, PhD,
Department of Kinesiology,
College of Health and Human Performance, East Carolina University, Greenville, NC
U.S.A.

Joseph Kalinowski, PhD,
Department of Communication Sciences and Disorders,
College of Allied Health Sciences,
East Carolina University,
Greenville, NC
U.S.A.

Nicholas Altieri, PhD,
Department of Communication Sciences and Disorders,
Division of Health Sciences,
Idaho State University,
Pocatello, ID
U.S.A.

Abstract

Purpose: To measure the effect of stuttering on eye-gaze in fluent speakers while viewing video presentations of typical speakers and people who stutter (PWS) speaking because eye-gaze behaviors provide indicators of emotion and communicative integrity.

Method: Sixteen fluent college-age adults, naïve to stuttering, observed six 30-second audiovisual speech samples of three PWS, and three age and gender matched controls who do not stutter (PWNS). A desk-mounted eye-tracker recorded the amount of time participants spent watching four regions of interest (ROIs) in the stimulus videos of PWS and PWNS: eyes, nose, mouth, and “outside” (i.e., any gaze-point not occurring within the eyes, nose, or mouth area). Proportions of gaze-time in each ROI were the dependent variables of interest in the study. Comparisons were made between proportions of time spent in each ROI for the PWS and PWNS speaker groups, and also between fluent versus disfluent speech segments produced by the PWS.

Results: Participants spent significantly more time watching the eyes (e.g., maintaining eye-contact) when viewing PWNS than PWS. They also spent significantly more time observing mouth regions of PWS. When watching the videos of PWS, participants spent significantly more time observing nose and mouth regions when speech was stuttered (PWS-S) than when the speech was fluent (PWS-F).

Conclusions: Overall, the difference in eye gaze patterns across speaker-group is interpreted to indicate negative emotional responses to stuttering. Current findings align with previous research showing that stuttered speech elicits negative reactions from listeners. Specifically, stuttering behaviors avert gaze from the eyes. Gaze aversion is a clear sign of disrupted communication that is visible to PWS and may contribute to their negative reactions to their own stuttering.

Abrégé

Objectif : Mesurer l'effet du bégaiement sur la fixation du regard chez les locuteurs fluides qui visionnent des présentations vidéo de locuteurs typiques et de personnes qui bégaient, parce que les comportements de fixation du regard sont des indicateurs de l'émotion et de l'intégrité dans la communication.

Méthode : Seize adultes d'âge collégial ayant une parole fluide, sans histoire de bégaiement, ont observé six échantillons audiovisuels de parole de trois bègues et de trois personnes ne bégayant pas, jumelées selon l'âge et le sexe. La durée des échantillons était de 30 secondes. Un dispositif de suivi oculaire de type bureau a enregistré le temps passé par les participants à regarder quatre régions d'intérêt dans les vidéos : les yeux, le nez, la bouche et « l'extérieur » (c.-à-d. tout point de fixation du regard situé ailleurs que sur les yeux, le nez et la bouche). Les portions de temps de fixation du regard dans chaque région d'intérêt étaient les variables dépendantes de l'étude. Les comparaisons furent faites entre les portions de temps passées dans chaque région d'intérêt pour les groupes de locuteurs bègues et non bègues et également entre des segments de parole fluide et de bégaiement des bègues.

Résultats : Les participants ont passé un temps significativement plus important à regarder les yeux (par ex., maintenir le contact visuel) en visionnant les non bègues que les bègues. Ils ont également passé significativement plus de temps à observer la région de la bouche des bègues. En visionnant les vidéos des bègues, les participants ont passé significativement plus de temps à examiner les régions du nez et de la bouche quand le locuteur était dans des périodes de bégaiement comparativement à des moments où sa parole était fluide.

Conclusions : Dans l'ensemble, les différences des comportements de fixation du regard à travers les groupes de locuteurs sont interprétées comme indiquant des réponses émotionnelles négatives vis-à-vis le bégaiement. Les constatations présentes vont dans le même sens que des recherches précédentes qui montraient que les périodes de bégaiement suscitent des réactions négatives chez les auditeurs. Spécifiquement, les comportements de bégaiement détournent le regard des yeux. Ce détournement du regard est un signe clair d'un bris de communication, perceptible par les bègues et qui peut contribuer à leurs réactions négatives envers leur propre bégaiement.

Introduction

Stuttering is a disorder characterized by auditory disruptions and visually evident struggle behaviors. Acoustic productions of sound and word repetitions, phoneme prolongations, and postural fixations often co-occur with visually distracting secondary behaviors of excessive lip tension, forceful blinking, involuntary head jerking, and other extremity movements (Bloodstein & Bernstein-Ratner, 2007). These disruptive speech behaviors categorically distinguish stuttering from fluent speech. These disruptions adversely impact the dynamics of a verbal communicative exchange between two individuals, which hypothetically may be signaled by changes in visual attention such as averted eye gaze.

Evidence that people react differently to PWS is supported by self-reported responses and physiological reactions from listeners observing recordings of stuttered and fluent speech. Participants usually self-reported similar negative emotional reactions of increased anxiety, tension, and uneasiness during video observations of PWS speaking, via a bipolar semantic differential state emotional scale (Guntupalli, Kalinowski, Nanjundeswaran, Saltuklaroglu, & Everhart, 2006). Finally, studies have also shown increased average skin conductance and decreased average heart rates in both PWNS (Guntupalli, et al., 2006; Guntupalli, Everhart, Kalinowski, Nanjundeswaran, & Saltuklaroglu, 2007) and PWS (Zhang, Kalinowski, Saltuklaroglu, & Hudock, 2010) when observing video presentations of PWS speaking. This change in skin conductance is likely a manifestation of anxiety/tension since it co-occurs with self-reports of such emotions. Evidence for alterations in attention, changes in physiological responses, and sender receiver dynamics are also indicated by several "stereotyping studies" of stuttering (e.g., Cooper & Cooper, 1996, for a review). In these studies participants are typically requested to complete personality attribute questionnaires regarding hypothetical people who stutter (PWS) and people who do not stutter (PWNS). PWNS typically judge PWS as being more tense, anxious, and uneasy. Similar personality judgments are reported during video observations of PWS as compared to PWNS (Tatchell, van den Berg, & Lerman, 1983). However it should be noted that the previously mentioned physiological reaction and stereotyping studies examine features that are not observerable to the PWS. It is important to look at a reaction that is visible to the PWS as it is these reactions that most likely alter the communicative dynamics.

Eye Gaze and Tracking

Interpersonal communication is a dynamic process of sending and receiving information between individuals. Verbal messages are transmitted synchronously with non-verbal nuances, including hand gestures, posture, facial expressions, proximity, and eye contact, which together signal the emotional tone of the exchange. In fact, it is often easy to understand the emotional tone of a communicative exchange simply by watching two individuals interact. Eye contact in particular provides telling clues regarding integrity and emotional valence of verbal interaction. According to a large body of research, negative avoidance-oriented emotions (e.g., embarrassment, grief, and revulsion) are typically conveyed via gaze aversion (Argyle & Cook, 1976; Blakemore & Frith, 2004; Kleinke, 1986; Rutter, 1984). Humans have the capacity (e.g., Theory of Mind) (Baron-Cohen, 1997) for understanding and reciprocating the subtle cues that encode emotions (Gallese, 2003). This occurs to the extent that when a sender portrays an emotion, receivers understand intentional cues, and display similar emotions back to the sender, therefore dynamically influencing sender-receiver interactions.

It is a common understanding that PWS exhibit decreased eye contact during interpersonal communicative exchanges (Bloodstein & Bernstein-Ratner, 2007). To examine one aspect of this, Atkins (1988) asked 133 college students to judge hypothetical speakers' personality characteristics as a result of their perceived criteria of "good" eye contact (i.e., 90 – 100%) versus minimal or no eye contact. Students judged speakers who used good eye contact much more favorably than speakers with minimal eye contact. The researcher inferred that therapy for PWS should target increasing eye contact. Other researchers more directly examined influences of eye contact and verbal fluency on listeners' personality judgments. For example, Tatchell et al. (1983) examined undergraduate students' personality judgments/ratings from video recordings of an actor who maintained or averted eye contact during fluent and stuttered speech. Participants' perceptions were differentially affected during the four test conditions. Maintained eye contact with fluent speech was judged the most positive. The next highest ranking was maintained eye contact with stuttering. The lowest ranked condition was averted eye gaze and stuttering. Clinicians have argued for the importance of maintaining "good" eye contact for PWS and have integrated it as a crucial component to therapeutic success (Breitenfeldt & Lorenz, 1989; Sheehan, 1970; Tatchell et al., 1983). In fact, amount of eye contact has also been used in clinical transcription during stuttering therapy (Tetnowski & Franklin, 2002). Based on this evidence, one might expect listeners to also avert their eye gaze when witnessing stuttering. However, listeners' eye gaze behaviors when watching a PWS who is maintaining eye contact has yet to be fully explored.

Eye tracking using video presentations are commonly used procedures for exploring interpersonal human social communication. It is pertinent to mention that all of the studies listed in the introduction that examined participants' reactions to stuttering, and many of the experiments that use eye-tracking technology, use prerecorded videos for stimuli. This procedure is typically implemented to insure consistency and standardization of speech characteristics (Mendel & Owen, 2011). Additionally, there are very limited studies comparing participants' reactions to live communication to reactions towards recorded speech. There are certain drawbacks to presenting recordings to participants; first, it limits the interpretations that one can make from the data. Secondly, it is not a true naturalistic situation, so participants' reactions might not be truly representative of what occurs during interpersonal communication exchanges.

The eyes transmit information regarding attention, turn taking, respect, emotion, and intention (Adams & Kleck, 2005; Baron-Cohen & Cross, 1992, Frischen, Bayliss, & Tipper, 2007). From infancy to adulthood, humans and higher order mammals demonstrate a propensity to fixate on eye regions and eye-like objects (Tomalski, Csibra, & Johnson, 2009). This proclivity for directed gaze behavior may be developmentally influenced by anatomical characteristics (i.e., elevated cheek bone, pronounced brows, and contrast of the sclera to the iris) or social-emotional factors involved in nonverbal communication (Kobayashi & Kohshima, 2001; Tomalski et al., 2009). As eye-gaze appears to be largely innate and important for communicative exchanges, gaze aversion leads to a variety of interpretations, including social cueing responses, reducing empathetic connections, increasing cognitive loads, presentation of peripheral stimuli, and decreased interest (see Blakemore & Frith, 2004, for a review). Simply put, although there are many possible interpretations of eye gaze aversion during communicative exchanges, some of the most common relate to turn taking, emotional factors, and altered attention. For example, conspicuous breakdowns in this process are observed in children with autism spectrum disorders, as compared to their fluent peers, who exhibit reduced time spent observing eye regions when viewing social communication situations (Klin, Jones, Schultz & Volkmar, 2003).

Eye-tracking procedures with PWS have examined a variety of factors; initially however, researchers primarily attempted to examine anticipation of stuttering during silent reading (Bakker, Brutten, Janssen, & van der Meulen, 1991; Brutten & Janssen, 1979; Roland, 1972). PWS exhibited more retraces and fixations, although anticipations were not related to stuttering. Participants who stuttered tended to gaze ahead at words that they listed as "difficult to say", and would gaze more often at sections that they had just read. Neither anticipation nor retracing behaviors were significantly related to overt stuttering. More recently, researchers have examined social factors related to the sender or receiver dynamics of speech. Lowe et al. (2012) examined PWS eye gaze towards prerecorded audiences during oral presentations. Participants were initially told that the presentations were being televised to audience members in an adjacent room, however full discloser occurred after the study, as is common practice in deception style studies. Participants who stutter as compared to fluent controls, spent less time viewing audience members with positive reactions as compared to negative and neutral reactions. Furthermore, the extent of gaze aversion from more positive audience members was significantly correlated to self-reported anxiety about the speaking situation.

To examine the receiver aspect of the communication process, researchers presented participants with fluent and stuttered audiovisual segments while they recorded eye gaze (Bowers, Crawcour, Saltuklaroglu & Kalinowski, 2010; Zhang & Kalinowski, 2012). Bowers et al. (2010) employed a single PWS during 20 second (s) audiovisual segments of fluent and disfluent speech. Results revealed that participants decreased observations of eye regions and increased observations of nasal regions during stuttered stimuli. This study provided pilot objective evidence of gaze aversion to stuttering speech. Being the first of its kind, it had a number of constraints, including the use of only one speaker to provide fluent and stuttered speech samples, requiring participants heads to be constrained within a chin mount, and ensuring that the videos of the speaker did not contain head movements. Zhang and Kalinowski (2012) also employed three speakers under both fluent and stuttered conditions, but presented 60 s audiovisual recordings, and examined responses by Caucasian Americans, African American, and Chinese participants. Results revealed both American groups decreased observation of eye regions and increased observation of mouth regions when presented stuttered stimuli. The Chinese group exhibited less eye gaze for both fluent and stuttered stimuli. Different from the American groups, the Chinese also increased observation time of outside regions of interest (ROI) during stuttered speech.

By objectively quantifying eye gazing behaviors during the presentation of audiovisual recordings from both fluent and disfluent speakers, clinicians and researchers can gain insight into factors influencing communication dynamics. As with recent research (Bowers et al., 2010; Zhang & Kalinowski, 2012), the current study sought to explore fluent listeners eye gaze behaviors during observation of stuttered and fluent speech. This study differed systematically from Bowers et al. (2010) in several ways. First, it increased ecological validity by including greater talker variability (i.e., six speakers in two mutually exclusive speaker categories) for the stimulus (e.g., Gilbert, Tamati, & Pisoni, 2013) instead of one and three speakers under two speaker conditions in Bowers et al. (2010) and Zhang and Kalinowski, (2012) respectively. Also, every participant watched each speaker only once, unlike the previous studies that presented the same speaker during fluent and disfluent conditions. The current study is the first to examine duration of direct gaze from the speaker, allowing for more complete quantification of gaze pattern behavior. Most importantly, stuttering is an intermittent pathology. That is, PWS do not stutter in every production. Hence, to better understand the impact of stuttering on eye gaze, it is necessary to examine how listener eye gaze varies as a function of speech fluency (i.e., stuttered vs fluent speech) when watching PWS. This is the first study using eye-tracking that analyzed gaze patterns in response fluent (PWS-F) and disfluent (PWS-S) segments of PWS speech. Previous study designs have not made this distinction. As such, an examination of the extent to which gaze aversion is tied to actual episodes of stuttering versus being a more global phenomenon related to communicative integrity is possible.

We first hypothesized that participants would decrease the proportion of time spent observing eye regions as well as increase proportion of time spent viewing mouth regions when viewing PWS as compared to PWNS. We also hypothesized that participants would decrease observation time of eye regions and gaze more toward the nose/mouth region when observing PWS-S as compared to PWS-F.

Methods

Participants

Twenty-one undergraduate students with no training in speech-language pathology, or self-reported history of cognitive, emotional, hearing, visual, speech, or language abnormalities participated. Participants reported being naïve to stuttering and had not participated in any previous studies of this kind. Two participants were unable to take part in the experiment due to inability in obtaining appropriate calibration (i.e., one had nystagmus, and one had a chip in their bifocals which inhibited accurate calibration). Additionally, three participants data were not used in the analysis because they maintained gaze with only one region throughout all stimulus presentations therefore resulting in sixteen participants (6 male and 10 female; *M* = 21.1 years, *SD* = 3.2). Prior to experimental procedures, informed consent (approved by the University and Medical Center Institutional Review Board at East Carolina University) was obtained from all participants.

Stimuli

Stimuli for the current study were designed similar to those used in Zhang and Kalinowski (2012). However, the current study used six speakers with 30 s durations instead of three speakers under both fluent and disfluent conditions for 60 s durations, as in Zhang and Kalinowski (2012). Video production for the current study required speakers to maintain direct gaze with the camera, were standard definition quality, and framed on the face, therefore omitting shoulder areas (see Figure 1). Audio-visual recordings of three male PWS with three age and gender matched PWNS maintaining directed eye gaze with a teleprompter were used as stimulus videos. Texts consisted of six non-standardized passages from fifth to seventh grade reading levels, as determined by the Flesch-Kincaid reading scale (Kincaid, Fishburne, Rogers, & Chissom, 1975; as reported in Saltuklaroglu, 2004). Reading was chosen to control for content and complexity. Multimedia staff at East Carolina University professionally recorded stimulus videos. Speakers were recorded in a sound-treated studio while wearing a unidirectional collar microphone attached below the viewpoint of the camera. Table 1 presents the behavioral characteristics of speakers. All three PWS samples were considered to be severe via informal assessment from three credentialed speech-language pathologists.

Videos were presented on a 51 cm Dell 2001FP computer monitor. Center point of the monitor was located approximately 60 cm at 0^{0} azimuth and 0^{0} altitude from the participants' line of sight. Stimulus presentation was controlled by a Dell Optiplex GX280 personal computer via GazeTracker software (Version 8.0; Eye Responses Technologies, 2009). Audio sound recordings were presented simultaneously on two speakers (Harman Kardon DP-N 02320V) located adjacent to the monitor. Presentation levels were set at a comfortable listening level (e.g., 65-75 dB SPL). A D6 desk-mounted optical/camera array eye-tracking system

Table 1. Speech and behavioral characteristics of stimulus speakers.

Speaker characteristics

Stimuli Video	Total Syllables Spoken	% Stuttered Syllables	% Total Time Stuttered	Longest Stuttering Episode (s)	Speech Rate (syllables/s)	Duration of eye contact (s)	Number of Eye blinks	Concomitant Stuttering Behaviors
PWS 1	23	26	96	14.7	NA*	29.48	4	Lip protrusions and muscle tension in the forehead
PWS 2	52	19	57	4.1	NA*	27.11	13	Lip protrusions, eyebrow raising, and head movements
PWS 3	69	16	61	9.7	NA*	28.68	20	Lip protrusions and head jerks
Fluent Speaker 1	124	0	0	NA	4.7	28.58	5	NA
Fluent Speaker 2	132	0	0	NA	4.8	28.80	4	NA
Fluent Speaker 3	100	0	0	NA	4.2	29.36	4	NA

Note. Behavioral characteristics of fluent and stuttered speech samples (per 30 s segments) as a function of speaker. Stuttering was operationally defined as part-word or whole word repetitions, phoneme prolongations, and postural fixations. *Indicates that speech rate could not be calculated for the PWS since 50 perceptually fluent contiguous syllables were not produced (Kalinowski, Armson, Roland-Mieszkowski, Stuart, & Gracco, 1993).

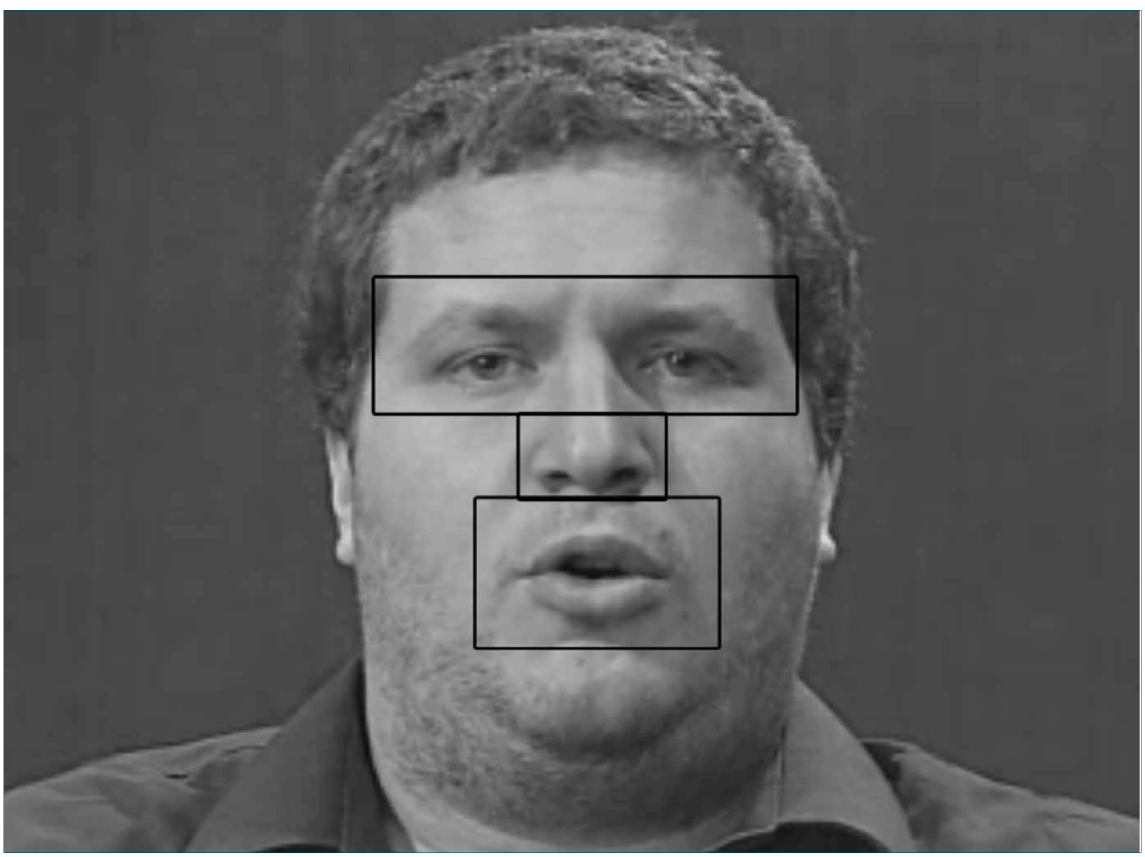

Figure 1. Example of Regions of Interest (ROI) framing used in the current study. Anatomical markers for region creations were retrieved via Bowers, *et al.*, (2010).

(Applied Sciences Laboratories - ASL) was used to collect gaze behaviors. It sat directly below the monitor and housed an infrared camera designed to capture the pupil diameter and corneal reflection of the left eye at 60 Hz. It is common for eye-tracking devices to only record one eye due to synchronous movements of both eyes. The D6 unit also employed a second camera for online video head tracking to adjust for subtle X, Y, and Z spatial head movement. However, participants were instructed to move as minimally as possible during the experiment, especially during the calibration phase. A second Dell Optiplex GX280 personal computer controlled the D6 unit via EYE-TRAC®6 software (ASL) capturing gaze data and allowing for real-time monitoring of gaze behaviors on an 18 cm VGA closed circuit monitor. Prior to data collection, each participant performed a nine-point calibration sequence per manufacturer's specifications. Offline analysis was performed with GazeTracker software that overlaid recorded data onto stimulus videos enabling the creation of dynamic ROI.

Procedure

A fluent research assistant initially briefed participants about general experimental procedures, and requested that they remain as motionless as possible throughout the study, especially during the calibration. Participants then read and signed informed consent documents. They were seated in a stationary chair positioned 61 cm in front of the stimulus screen/eye tracking system. A second researcher was seated behind a partition, out of view of the participants, to calibrate the equipment and start the experiment. Researchers adjusted camera positions through the software controls, acquired adequate pupil contrasts and corneal reflections, and turned on the video-head-tracker for auto calibration. Researchers then calibrated participants according to standardized nine-point ASL calibration procedures. The calibration sequence presented a 1.5 cm grey cross at nine distinct points (three equidistant rows and columns) across the screen. Crosses were displayed only one at a time. Participants fixated on the cross until the software acquired calibration data for each point. Once data for one point was calibrated, the cross would disappear and the next cross would appear sequentially from top left to top right before moving down a row and proceeding from left to right. The final point for calibration was the bottom right. Upon calibration of all nine points, participants fixated on a central cross, if tracking was off more than 1.5 cm from the central cross, the calibration sequence was re-administered. Researchers then started the experiment via GazeTracker that presented random sequences of the six stimulus videos. Each participant observed all six videos in a randomized order. Between each video presentation the central cross was presented to subjectively determine tracking consistency. No participants' tracking data drifted more than 1.5 cm from the outside edges of the central cross, so no recalibrations were performed.

Regions of Interest

Using Gaze-Tracker software, four ROIs, similar to those employed by previous researchers (Bowers et al., 2010; Zhang & Kalinowski, 2012), were generated for offline analysis. They included: both eyes (i.e., approximately superior to the eyebrows, lateral edges of the eye sockets, and zygomatic process protrusion "cheek bone protrusions"); nose (i.e., inferior portion of the nasium, extending in close proximity past the lateral edges of the nostrils, and inferior portion of the nasal bone); mouth (i.e., extending past the inferior portion of the lips, lateral edges of the lips, and inferior portion of the nasium); and "outside" (i.e., any gaze-point not occurring within the eyes, nose, or mouth area) regions (see Figure 1). A visual overlay was used to create unique dynamic ROI using a 0.25 s video playback speed during placement of the ROIs to increase accurate placements without overlap.

Gaze Analysis

Gaze data were exported from GazeTracker for offline analysis. Only the tracked time in the ROI was examined as the dependent variable in the current study. For fixation counts and fixation durations during the observation of stuttered compared to fluent speech, please see Bowers et al. (2010). Proportion of total time tracked data was analyzed to account for any time lost during recordings (i.e., loss of corneal reflection, loss of pupil diameter, and blink artifact). Percent time recorded was typically over 92%.

For further examination, using previously described methods to create ROIs, researchers created ROIs for the PWS videos including sub-segments of fluent (PWS-F) and disfluent (PWS-S) periods. Disfluencies were defined using Stuart, Frazier, Kalinowski, and Vos's (2008) adapted categorization of Conture (2001). Two researchers trained on analyzing stuttering independently categorized fluent and disfluent segments while observing audiovisual presentations on Peak Pro Version 6.0. This enabled researchers to make note of time of onset/offset for disfluencies. Disagreements on stuttering episodes were under 1.8% of total spoken syllables. Disfluencies ranged from approximately 200 ms to 14.3 s. If one syllable in a word was classified as disfluent by both researchers, the entire word was categorized as disfluent for segment analysis. Between word disfluencies that carried into initial syllable disfluencies and phrase repetitions were classified as disfluent segments. For example, if the speaker exhibited a postural fixation that proceeded to a tense production (e.g.,______C-C-C-Carried) the auditory or visual moment when the observers marked the disfluency beginning was included in the PWS-S segment. Only agreed upon stuttering episodes were used for the disfluent segments. When stuttering occurred at the onset of a word, an average onset time as marked by the two researchers was used. Our shortest PWS-S segment was just over 400 ms with the average PWS-S segment being 927 ms (excluding the 14.3 s outlier). However, the outlier was included in the ROI analysis. Additionally the fluent, PWS-F, segments ranged from 320 ms to 6.8 s with an average fluent duration of 4.3 s. As with the previous analysis, GazeTracker exported total time tracked data in region for each ROI for total PWS-F and PWS-S segments, which were then analyzed using SPSS.

Results

In the initial set of analyses, we examined whether eye gaze patterns systematically differed during the viewing of PWS compared to PWNS. Prior to inferential statistical analysis, participants' proportional gaze-times were transformed into arcsine units to reduce endpoint weighting (see Zar, 1996). The means and standard errors for proportion of gaze time averages, as a function of speaker group and ROI, are displayed in Figure 2. First, a two-factor repeated measures analysis of variance (ANOVA) was used to examine proportions of gaze as a function of fluency and ROI. A significant main effect of ROI was observed [$F(1.70, 25.48) = 15.25$, Greenhouse-Geisser $p < 0.0001$, $\eta_p^2 = 0.50$], but not for fluency overall [$F(1, 15) = 2.75$, $p = .12$, $\eta_p^2 = 0.16$]. A significant ROI by fluency (PWS vs. PWNS) interaction, however, was observed [$F(2.59, 38.84) = 5.47$, Greenhouse-Geisser $p = 0.005$, $\eta_p^2 = 0.27$], indicating that effects for fluency may have arisen in certain ROIs.

To examine the source of the interaction, four separate paired samples t-tests were utilized to examine the proportions of gaze time as a function of speaker at each ROI. We used a Bonferroni corrected alpha level of .0125 (.05/4) to correct for multiple comparisons. First, significant differences were not observed in the proportions of gaze time for the nose (t (15) = -0.83, $p = 0.42$, $\eta_p^2 = .044$) and outside regions of interest (t (15) = -0.78, $p = 0.45$, $\eta_p^2 = .039$). There were significant differences in the proportions of gaze time for the eye (t (15) = 4.01, $p = 0.001$, $\eta_p^2 = .52$) and mouth regions (t (15) = -2.89, $p = 0.012$, $\eta_p^2 = .36$) for PWS versus PWNS. These results revealed that participants' gazed more at PWNS eyes, and PWS mouths when examining the entire videos. To obtain a relative index ratio of these differences, an eye gaze to mouth gaze proportion was computed for each participant, then averaged for both of the speaker conditions (i.e., PWNS and PWS). We found that participants viewed eye regions during PWNS stimuli 7.7 times more often than the mouth in comparison to PWS (cf. ratios of 309 vs. 40).

Analysis of stuttered (PWS-S) and fluent (PWS-F) segments from the PWS

A major aim of this study was to investigate the extent to which stuttering episodes versus fluent speaking in PWS influences eye-gaze behavior. This will help determine whether stuttering episodes contribute to shifting of eye gaze to an ROI or alternatively, whether participants treat stuttered and fluent episodes similarly when viewing PWS.

Mean proportion of gaze times in each ROI are displayed for PWS-F and PWS-S episodes in Figure 3, and compared to PWNS. We utilized two-factor repeated measures ANOVAs to examine participants' gaze proportions of PWS video speech as a function of production fluency (i.e., PWS-F or PWS-S) and ROI. (Once again, participants' proportional eye gaze times were transformed into arcsine units prior to inferential analysis). Significant main effects of ROI [$F(1.69, 25.37) = 7.76$, Greenhouse-Geisser $p = 0.004$, $\eta_p{}^2 = 0.34$] and ROI by speaker category were observed [$F(2.18, 32.75) = 3.40$, Greenhouse-Geisser $p = 0.042$, $\eta_p{}^2 = 0.18$]. The main effect of speaker category was not statistically significant [F (1, 15) = 0.25, $p = .63$, $\eta_p{}^2 = 0.016$]. Next, four paired samples t-tests were carried to examine the proportion of gaze-time as a function of fluency (PWS-S vs. PWS-F) at each ROI. A Bonferroni corrected alpha level of .0125 (.05/4) was again implemented to correct for multiple comparisons.

First, we failed to observe significant differences in the proportions of gaze time for eye, although interestingly, there was a trend for greater gaze time when stuttering occurred (t (15) = 2.33, $p = 0.034$, $\eta_p^2 = .015$). Second, the nose region was significant, with greater gaze time directed toward the nose when stuttering occurred (t (15) = 3.62, $p = 0.003$, $\eta_p^2 = .022$). Third, the mouth region was significant; again, with greater gaze time occurring during episodes of stuttering (t (15) = 2.95, $p = 0.01$, $\eta_p^2 = .013$). Finally, there was a non-significant trend for the outside ROI in PWS-S versus PWS-F (t (15) = 2.43, $p = 0.028$,

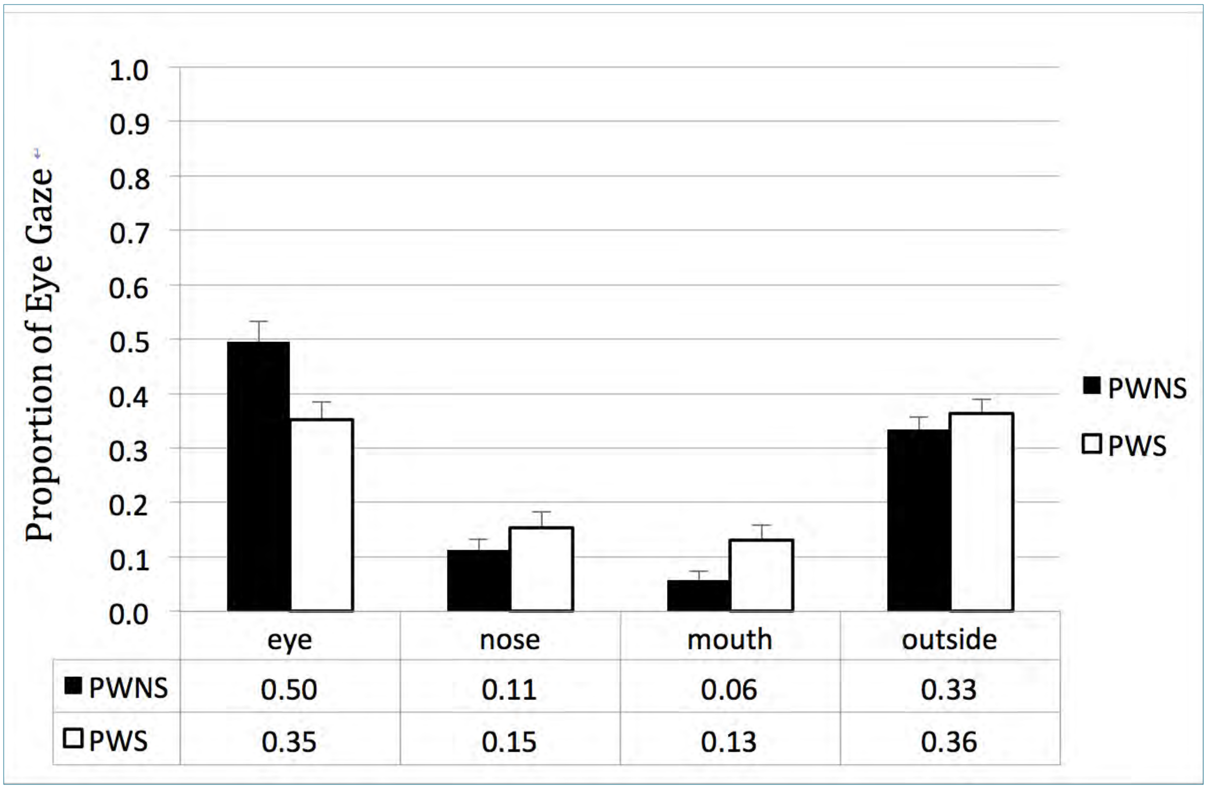

Figure 2. Mean proportion of gaze time for speaker and region (x-axis). The gaze appears to shift from the eyes to mouth in PWS compared to PWNS. Error bars represent plus one standard error of the mean.

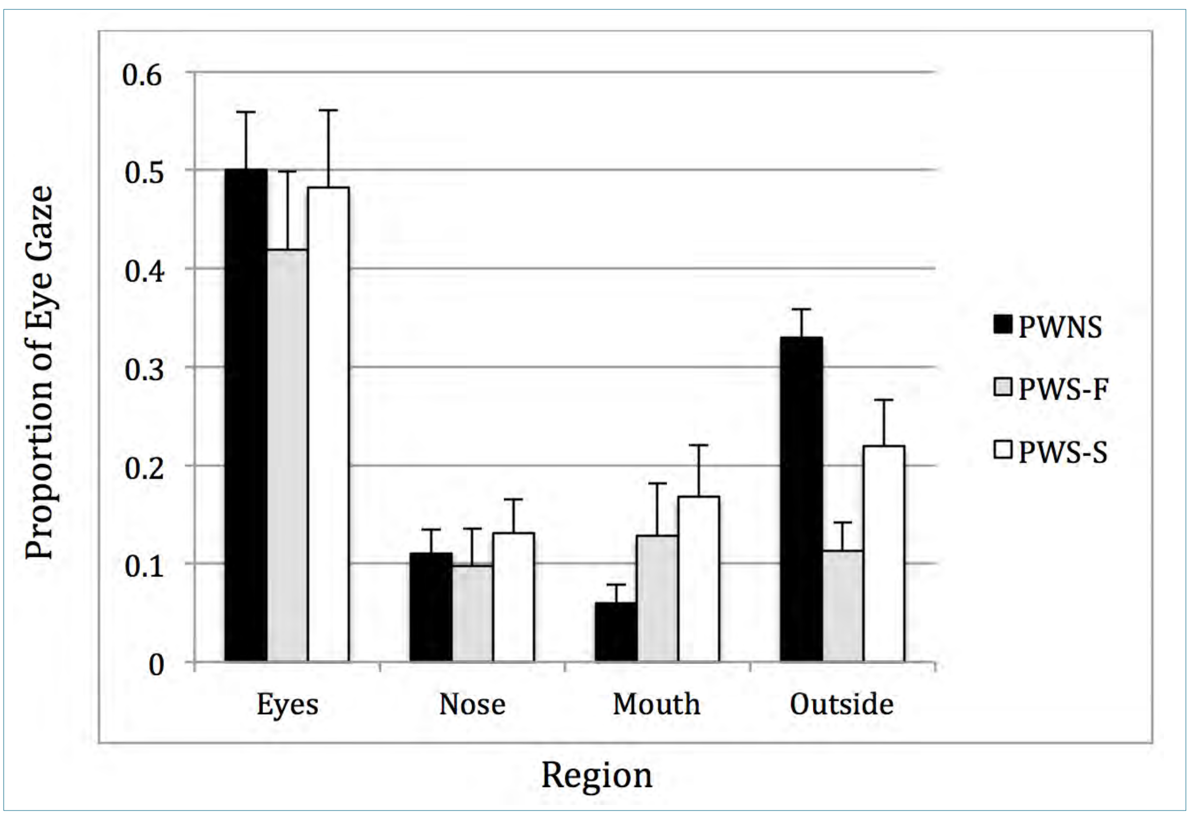

Figure 3. Mean proportion of gaze time in ROI for segment analysis. Error bars represent plus one standard error of the mean.

η_p^2 = .08). Taken together, these results suggest that stuttering contributed to gaze aversion toward the nose-mouth region, as predicted.

Discussion

This study aimed to collect objective data to better understand eye contact and gaze aversion in response to stuttered speech. In doing so the current study produced findings that are consistent with, yet expand upon the findings of Bowers et al., (2010) and Zhang and Kalinowski, (2012). That is, fluent speakers averted their eye gaze when they observed audiovisual recordings of PWS speaking. Most importantly however, this study further decomposed the speech of PWS into fluent versus disfluent segments and demonstrated a more specific pattern of gaze aversion to stuttered events.

First, significant differences were observed between the proportion of time participants spent observing both eye and mouth regions throughout the entire videos. When viewing PWS as compared to PWNS, participants decreased observation of the eyes region by 30% and increased observation of the mouth region by 115%. Our findings supported our initial hypothesis that participants would decrease proportion of time spent observing eye regions and increase proportion of time spent observing mouth regions when viewing PWS compared to PWNS. Findings that fluent speakers spend less time observing the eyes region of PWS during whole video analysis adds converging support to prior results (Bowers et al., 2010; Zhang & Kalinowski, 2012). However, this is the first study to examine fluent and stuttered segments of PWS speech (i.e., PWS-F and PWS-S), in regard to listeners' eye gaze behaviors. This novel examination indicates that eye-gaze behavior does not result from aversion to PWS *per se*, but instead result from aversion to stuttered events themselves.

A commonly accepted theory of gaze aversion is the presence of avoidance-oriented emotions (e.g., increased anxiety, tension, and uneasiness). It could therefore be interpreted that the reduction in proportion of time spent observing the eyes region of PWS supports evidence of negative self-reported emotional states and physiological arousal in listeners when observing PWS speaking (see e.g., Bowers et al., 2010; Guntupalli et al., 2006, 2007; Zhang et al., 2010). These interpretations are also supported by our findings that participants increased observations of nose and mouth regions during PWS-S. The display of anxiety, uneasiness, tension, shame, and guilt on the faces and in the eyes of PWS may lead receivers to avoid observing these manifestations (Bowers et al., 2010; Zhang & Kalinowski, 2012).

Another interpretation for eye gaze shift relates to the "meaningfulness" of eye gaze aversion. Here, participants seem to have been engaging in eye gaze avoidance rather than searching for additional visual cues to understand speech. On the other hand, during presentation of static facial images of prerecorded audiovisual or visual-only speech samples and face-to-face communication, participants typically spend 50-60% of the time observing eye regions (Bowers et al., 2010; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). The current study revealed values within these ranges during perception of PWNS, PWS-S, and PWS-F. However, during perception of PWS videos, these proportions significantly decreased overall compared to PWNS. Proportion of eye gaze decreased during PWS, although this was not the case for PWS-F or PWS-S. This finding demonstrated how variable eye-gaze can be depending extraneous factors including: peripheral distractions, S/N ratio, direction of the senders gaze, duration of presentation, social interpretation, emotional context, situation of presentation (life or video), and the speaker's verbal fluency (Klin et al., 2003).

Differences in observed gaze-time of mouth regions supported Zhang and Kalinowski's (2012) findings, but differed from Bowers et al. (2010) who previously reported differences in observations of eye versus nose regions rather than eyes to mouth. Although there appears to be disparity between these findings, there may be considerable similarity. The one PWS used as the stimulus speaker for Bowers et al. (2010) manifested his visually aberrant stuttering behavior at the level of the nose (i.e., nostril flaring during stuttering). The multiple PWS used for stimuli in both Zhang and Kalinowski (2012) and the current study primarily exhibited their associated visual stuttering behaviors at the level of the mouth (i.e., lip tension, flexion, and fixation). There is another important implication of the participants' tendency to spend more time looking at the mouth of PWS; given that participants can only observe one spot at a time, any increase in time spent in one location implies a move from another location. In other words, our current findings suggest the participants made gaze shifts from eyes to mouth regions when observing PWS. Postural fixations and excessive lip activity may have acted as peripheral stimuli enhancing the avoidance behaviors. Presentation of visual distractions in participants' periphery has been shown to shift directed eye gaze to search for and fixate on the peripheral stimuli (Greene, Mooshagian, Kaplan, Zaidel, & Iacoboni, 2009). In the case of observing PWS, an automatic visual orienting response may be triggered by the anticipation of stuttering, therefore requiring participants to proactively recruit additional gestural information for increased comprehension. This process is similar to that described in speech reading studies and speech threshold studies (Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998). As participants are presented with increased amplitudes of white noise during speech perception, their gaze tends to shift from eyes to central (i.e., nose) or mouth regions. Similarly, when participants are presented with audiovisual recordings of different speakers with auditory masking noise, they fixate on mouth regions more frequently and for longer durations (Buchan, Pare & Munhall, 2008). If presented with the same speaker across various intensities of the same auditory masking noise they tend to have central fixations on nose regions. The findings from Bowers et al. (2010) are relevant here; recall that the study had one speaker for all speaker conditions and that participants shifted eye gaze to the nose region where nostril flaring was occurring. These findings suggested that the shift in gaze may be motivated by an altered state of attention or the presence of peripheral visual distractions instead of searching for additional gestural information.

Conclusion

The current results support related findings that people react differently when viewing PWS speaking as compared to their fluent counter parts. Conclusions can be summarized into three categories. First, participants viewed the eyes region more and the mouth region less when watching the fluent speakers – the inverse, gaze shifting from eyes to mouth region, occurred when watching PWS speaking. Second, participants observed nose and mouth regions more during PWS-S compared to PWS-F. Thirdly, results from the segmented analysis appear to indicate that once a speaker is perceived as disfluent the listeners' gaze is altered during both fluent and disfluent segments. These findings are consistent with avoidance-oriented behaviors and support previous eye tracking and biobehavioral experiments showing negative self-reported and physiological arousal when viewing stuttered stimuli.

It is of interest to later examine what influence the use of a disclosure statement (e.g., "Hi my name is… and I stutter), altered levels of eye contact, or communicative skills training have on sender-receiver dynamics during video presentations and during naturalistic communicative exchanges. Future studies should examine both communication partners in more naturalistic environments to determine the initiator of gaze alteration behaviors with segmented methods of analysis. It is also likely that different results may be obtained during face-to-face interactions as compared to viewing prerecorded videos on a monitor. The use of high definition avatars might later allow for a more precise representation of naturalistic interactions that have limited output responses to defined input parameters (Le, Ma, & Deng, 2012). Additionally, future studies should further investigate dynamic sender-receiver interactions during the perception of typical and various types of disordered speech in order to better understand how to effectively and efficiently increase communication naturalness during interpersonal exchanges when total fluency might not be possible.

References

Adams, R.B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion, 5*, 3-11.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze.* New York: Cambridge University Press.

Atkins, C.P. (1988). Perceptions of speakers with minimal eye contact: Implications for stutterers. *Journal of Fluency Disorders, 13*, 429-436.

Bakker, K., Brutten, G., Janssen, P., & van der Meulen, S. (1991). An eyemarking study of anticipation and disfluency among elementary school stutterers. *Journal of Fluency Disorders, 16*, 25-33.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind.* Cambridge, Massachusetts: MIT press.

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: Evidence for the role of perception in the development of a theory of mind. *Mind and Language, 6*, 173-186.

Blakemore, S.J., & Frith, U. (2004). How does the brain deal with the social world? *Neuroreport, 15*, 119-128.

Bloodstein, O., & Bernstein-Ratner, N. (2007) A handbook on stuttering (6th ed.). Clifton Park, NY: Thomson Delmar Learning.

Bowers, A.L., Crawcour, S.C., Saltuklaroglu, T., & Kalinowski, J. (2010). Gaze aversion to stuttered speech: A pilot study investigating differential visual attention to stuttered and fluent speech. *International Journal of Language and Communication Disorders, 45*, 133-144.

Breitenfeldt, D. H., & Lorenz, D. R. (1989). *Successful stuttering management program (SSMP) for adolescent and adult stutterers.* Cheney: Eastern Washington University School of Health Sciences.

Brutten, G.J., & Janssen, P. (1979). An eye-marking investigation of anticipation and observed stuttering. *Journal of Speech and Hearing Research, 22*, 22-28.

Buchan, J. N., Pare, M., & Munhall, K. G. (2008). The effect of varying talker identiry and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research, 1242*, 162-171.

Conture, E. G. (2001). Stuttering: Its nature, diagnosis and treatment. Boston: Allyn and Bacon.

Cooper, E.B., & Cooper, C.S. (1996). Clinician attitudes toward stuttering: Two decades of change. *Journal of Fluency Disorders, 21*, 119-135.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*, 694-724.

Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology, 36*, 171-180.

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology, 24*, 26-36.

Greene, D.J., Mooshagian, E., Kaplan, J.T., Zaidel, E., & Iacoboni, M. (2009). The neural correlates of social attention: Automatic orienting to social and nonsocial cues. *Psychological Research, 73*, 499-511.

Guntupalli, V. K., Everhart, D. E., Kalinowski, J., Nanjundeswaran, C., & Saltuklaroglu, T. (2007). Emotional and physiological responses of fluent listeners while watching the speech of adults who stutter. *International Journal of Language and Communication Disorders, 42*, 113-129.

Guntupalli, V. K., Kalinowski, J., Nanjundeswaran, C., Saltuklaroglu, T., & Everhart, D. E. (2006). Psychophysiological responses of adults who do not stutter while listening to stuttering. *International Journal of Psychophysiology, 62*, 1-8.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., & Gracco, V. L. (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech, 1*, 1-16.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch report. Memphis Naval Air Station. 8-75.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 78-100.

Klin, A., Jones, W., Schultz, R.T., & Volkmar, F. (2003). Dr. Klin and colleagues reply. *The American Journal of Psychiatry, 160*, 1359.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution, 40*, 419-435.

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception and Psychophysics, 65*, 536-552.

Le, B. H., Ma, X., & Deng, Z. (2012). Live speech driven head-and-eye motion generators. *IEEE Transactions on Visualization and Computer Graphics, 18*, 1902-1914.

Lowe, R., Guastella, A. J., Chen, N. T., Menzies, R. G., Packman, A., O'Brian S., & Onslow, M. (2012). Avoidance of eye gaze by adults who stutter. *Journal of Fluency Disorders, 37*, 263-274.

Mendel, L. L., & Owen, S. R. (2011). A study of recorded versus live voice word recognition. *International Journal of Audiology, 50*, 688-693.

Roland, B. C. (1972). Eye movements of stutterers and nonstutterers during silent, oral and choral reading. *Perceptual and Motor Skills, 35*, 297-298.

Rutter, D. R. (1984). *Looking and seeing: The role of visual communication in social interaction.* New York: Wiley.

Saltuklaroglu, T. (2004). The role of gestural imitation in the inhibition of stuttering. Ph.D. East Carolina University, Greenville, NC.

Sheehan, J. (1970). *Stuttering: Research and therapy.* New York: Harper and Row.

Stuart, A., Frazier, C. L., Kalinowski, J., & Vos. P. W. (2008). The effect of frequency altered feedback on stuttering duration and type. *Journal of Speech Language and Hearing Research, 51*, 889-897.

Tatchell, R.H., van den Berg, S., & Lerman, J.W. (1983). Fluency and eye contact as factors influencing observers' perceptions of stutterers. *Journal of Fluency Disorders, 8*, 221-231.

Tetnowski, J. A., & Franklin, T. C. (2002). The clinical analysis layer of transcription. *Clinical Linguistics and Phonetics, 16*, 361-369.

Tomalski, P., Csibra, G., & Johnson, M.H. (2009). Rapid orienting toward face-like stimuli with gaze-relevant contrast information. *Perception, 38*, 569-578.

Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception and Psychophysics, 60*, 926-940.

Zar, J. H. (1996). *Biostatistical analysis* (3rd ed.). Prentice Hall: Upper Saddle River, NJ.

Zhang, J., & Kalinowski, J. (2012). Culture and listeners' gaze responses to stuttering. *International Journal of Language & Communication Disorders, 47*, 388-397.

Zhang, J., Kalinowski, J., Saltuklaroglu, T., & Hudock, D. (2010). Stuttered and fluent speakers' heart rate and skin conductance in response to fluent and stuttered speech. *International Journal of Language and Communication Disorders, 45*, 169-178.

Authors' Note

Correspondence concerning this article should be addressed to Daniel Hudock, PhD, Department of Communication Sciences and Disorders, Division of Health Sciences, Idaho State University, 921 South 8th Ave, Mail Stop 8116, Pocatello, ID 83209-8116, U.S.A. Email: Hudock@isu.edu.