Spring 2017

Effect of Visual Self-Image on Attention During Speech

David Ballenger
david.ballenger@trincoll.edu

Effect of Visual Self-Image on Attention During Speech

David E. Ballenger

Trinity College

Abstract

When people are made more self-aware either through name priming or the introduction of a mirror, the increase in self-awareness elicits a positive effect in performance of difficult mental tasks. This effect was first documented decades ago, but lacks extensive research. The relationship between self-awareness and seeing oneself has been studied in terms of task completion and self-esteem measures, but never in the realm of speech tasks. The present study focused on the effects of different speech goals on self-attention. Similar to earlier studies, a webcam was used to display the participants' images during tasks, while eye-tracking was used to determine how variation in speech goals affected attention to their visual self-image. It was found that proportionally more time is was spent looking at the eyes during no task, while a possible time-compression limit effect in looking times was observed in speech-relevant looks to participants' mouth/nose area in a recitation of the ABCs casual condition. We also found that attention towards the overall face and speech-relevant areas was more tightly correlated during storytelling tasks. Broadly speaking, visual feedback of one's own face is a unique form of feedback, and does appear to have some effect on attention during speech acts.

Effect of Visual Self-Image on Attention During Speech

Seeing Yourself

Difficult mental tasks elicit a positive effect on effort in task-completion when an individual's self-awareness is manipulated using a mirror (Silvia, 2012; Davis & Brock, 1997; Paulus, Annis & Risner, 1978; Windheim, Veale & Anson, 2011; Devue & Bredart, 2008). A related effect has also been measured using physiological responses to the presence of a mirror, in which individuals heart rates and palm sweating increased simply from the presence of a mirror in the room with them (Paulus, Annis & Risner, 1978). A study by Silvia et al. (2012), testing performance and blood pressure during the d2 attention task (Brickenkamp & Zillmer, 1998), found increased blood pressure in individuals who were self-primed with a mirror. Increased blood pressure here, is an indicator of increased arousal when self-primed. Performance, in this case measured by accuracy rates in the d2 attention test, was not affected by self-priming with a mirror. However, early research has indicated that the presence of a mirror leads individuals to complete tasks more slowly, presumably in order to complete the task with the highest level of accuracy (Silvia, 2012). Although there is no consensus on the exact manner of behavioral change elicited by self-priming, the combination of physiological and effort based measures support the claim that an individual's behavior will be altered by the presence of a mirror. Past research on this effect however, has been focused on taking measures of effort or self-consciousness. There is limited research in regards to how mirrors affect attention specifically, and again when this is studied it has been addressed in terms of negative self-attention's relation to self-consciousness (Paulus, Annis & Risner, 1978; Silvia, 2012).

In addition to these studies, research in the field of self-face perception has found that the

self-face represents a unique attentional object (Devue, Stigchel, Bredart & Theeuwes 2009; Keyes, Brady, Reaily & Foxe 2010; Tacikowski & Nowicka, 2010; Bredart, Delchambre & Laureys, 2000). While self-face research does not generally focus explicitly on the self-awareness effect discussed, it can offer useful insights into how a person interacts cognitively with their own image. For example, faces are highly distracting objects, drawing attention compared to other stimuli (Bredart, Delchambre & Laureys, 2000). Devue and Bredart (2008) presented the self-face as a distraction during a basic counting task, and found it produced a temporary distraction. Self-referential words or objects of attention have also been shown to produce a distracting effect, suggesting that self-referential stimuli, unrelated the human face, are highly distracting (Tacikowski & Nowicka, 2010). Both effects find their synthesis in the self-face, as it is both a human face and a self-referential object. It is important to note however, that the self-face has not been consistently shown to be more distracting than the face of a close associate, indicating an attentional for familiar faces, but not necessary one's own image (Devue, Stigchel, Bredart & Theeuwes, 2009; Gillihan & Farah, 2005). Regardless, there is a substantial body of evidence indicating that the face in general elicits attention in a unique way. As the present research will focus explicitly on how individuals distribute focus across their own image during speech acts, it is essential to understand the face's unique ability to draw attention in most individuals.

Speech

Research on attention during speech has been almost exclusively focused on the attention of conversational partners. In a normal conversational settings, the attention of the listener is mediated both by the goal of the speech produced and any interference making it more difficult

to perceive speech (Buchan, Pare´ & Munhall, 2008; Lansing & McConkie, 2003; Everdell, Marsh, Yurick, Munhall & Pare´, 2007; Lander & Capek, 2012; Vatikiotis-Bateson, Eigsti, Yano & Munhall, 1998). This effect presents itself in two distinct ways, both are demonstrated in a 2008 study by Buchan, Pare´, and Munhall in which gaze patterns were tracked during a normal conversational setting. First, the gaze of the listener is more heavily focused on the eyes when attempting to perceive the emotional content of speech, and more focused on the mouth and nose when attempting to perceive the words. Second, increased difficulty in speech perception, in this case caused by increased background noise, drew the gaze towards the mouth and nose. The first effect suggests that certain areas of the face are associated with conveying certain types of information. In this case the eyes are found to be better at conveying emotion. Similarly, the second effect suggests that certain areas of the face, the mouth and nose, are more important in perceiving informational content. Further studies have support the effects outlined in Buchan et al. (Lansing & McConkie, 2003; Lander & Capek, 2012; Vatikiotis-Bateson, Eigsti, Yano & Munhall, 1998).

Also regarding speech, significant differences have been documented between speech goals and their related speech production strategies (Cooke, King, Garnier & Aubanel, 2014; Theune, Meijs, Heylen & Ordelman, 2006; Smiljanic & Bradlow 2009). This effect has most commonly been noted between two distinct types of speech referred to as clear and casual speech. Casual speech is most often associated with slow speech, including more pauses and breaks, while clear speech is epitomized by faster speech, spoken at a constant rate without pauses and breaks (Smiljanic & Bradlow 2009; Cooke, King, Garnier & Aubanel, 2014). A 2006 study by Theune et al. examined specific qualities of informative vs storytelling speech in an

attempt to replicate storytelling style using a text to voice program. In an examination of speech style in storytelling vs new-casting (informative speech), story-tellers were found to use significantly more pitch shift and pausing, while the news-caster was found to use less pitch shifting and spoke at a consistent rate. There is a substantial body of research confirming the effect of speech goals on speech production (Cooke, King, Garnier & Aubanel, 2014; Theune, Meijs, Heylen & Ordelman, 2006; Smiljanic & Bradlow 2009), however there has been no effort to link this ideas to self-image and attention allocation during speech acts.

Present research

Despite the substantial body of research that exists regarding how visual self-image and how speech relate to attention, there is no research which seeks to address the relationship of speech and attention in the presence of self-image. This presents an issue for the present study, in that none of the foundational research actually addresses speech and self-image simultaneously. However, as video conferences and interviews become more common, it is more important now than ever to begin to understand how speech, and the ability to see oneself, will interact to affect attention. The present study seeks to fill these gaps in understanding. The study involved participants completing a few simple speech tasks, directed towards either casual or clear speech, while viewing their self-image. Participants also completed a non-speech task as well as a no-task condition. Eye-tracking was used as the primary measure of attention during each task. Areas of interest, for the example the eyes or the mouth, were defined prior to trials by a researcher, and focus on these areas was used as an indicator of self attention. Three specific predictions were tested. First, that the no-task condition would result in gaze patterns focused away from the self-face. Second, that speech tasks which require more accuracy on the part of

the participant (clear directed speech), were likely to draw more attention to the mouth, perhaps as a means to integrate visual feedback into speech production. Third, that speech tasks requiring less accurate speech, or speech which is meant to convey emotion rather than information, would draw more attention to eyes. This would be in keeping with the results of past research regarding gaze patterns in conversational settings (Buchan, Pare´ & Munhall, 2008; Lansing & McConkie, 2003; Everdell, Marsh, Yurick, Munhall & Pare´, 2007; Lander & Capek, 2012; Vatikiotis-Bateson, Eigsti, Yano & Munhall, 1998). It is the present studies hope to lay a foundation upon which further research regarding self-attention and speech can be conducted.

Method

Participants

Participants in this study included 20 Trinity college students. Participants included 4 females and 16 males between the age of 19 and 22 with a mean age of 20. All participants were native english speakers and passed a hearing screening with <= 25 dB HL. Students were recruited from Psychology 101 classes, as well as by flyers. All participants were volunteers.

Design

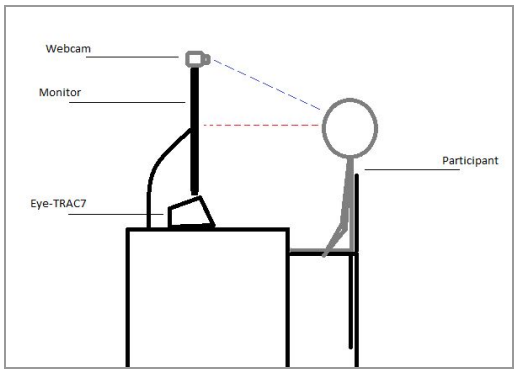
Participants were asked to complete a total of five conditions. The first was a no-task condition in which the participant was silent while viewing their image. The following four tasks consisted of reciting the ABC's and telling the story of Goldilocks, either casually or directed towards clear speech. Video was recorded both by the webcam used to produce the participants self-image, as well as by secondary video camera mounted next to the computer monitor. Total look count and proportional gaze duration were taken as primary measures of attention.

Procedure

The research design of this study is experimental and observational. Participants were asked to perform several speech acts while viewing a live video feed of their face displayed directly in front of them on a video monitor. Each participant completed an informed consent form, as well as undergoing a hearing test using an audiometer. Participants sat in a chair roughly 2' from the monitor showing their image. Calibration of the Eye-Trac 7 was performed prior to beginning each trial. Each participant was first directed to remain quiet and simply view themselves in the mirror, or look elsewhere in the room. While the participants were asked not to move their head excessively, they were not instructed specifically to look at their image in the mirror. Next each participant recited the ABC's in a casual manner, followed by a telling of Goldilocks and the Three Bears. The participant was then asked to perform the same speech tasks, but directed towards an individual who is hard of hearing. Researchers instructed the participant to speak as though they were speaking to an individual who has hearing impairment, necessitating the need for clear and well annunciated speech. Each participant completed each trial once. Measurements were taken based on the number of looks and proportional look duration to a given area of the face. Materials for this study include an Eye-Trac 7 unit, as well as a logitech c270 webcam, a secondary Panasonic HC-WX970 video camera, and 14"x12" computer monitor. The eye-tracking unit was placed directly underneath the monitor, with the webcam mounted on top of the monitor (see figure 1). An audiometer was used to test for any hearing issues in participants prior to the experiment.

Figure 1

Study Design

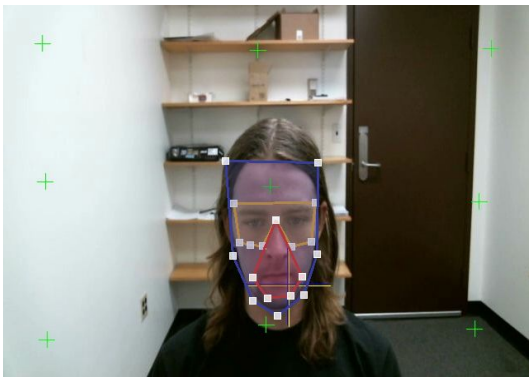


Data Analysis

Data was collected based on participants' gaze patterns including the number of of fixations to a given areas, as well as a the proportional look duration to a given area. Fixations are defined within the ASL results program. A fixation begins once 7 consecutive looks lasting at least .1 seconds each are record in a given area, and ends when 3 consecutive samples deviate from the initial fixation value How frequently a participant looked at a particular area, in conjunction with the duration of each fixation on a given area, was used to measure attention between various speech acts. Areas of interest were assigned to the scene videos by hand using the ASL results plus program supplied by the Eye-Trac 7 team. Three AOI's were drawn in the program for each participants, one on the eyes, one on the mouth and nose, and another around the entire face. (see figure 2)

Figure 2

Drawn AOIs

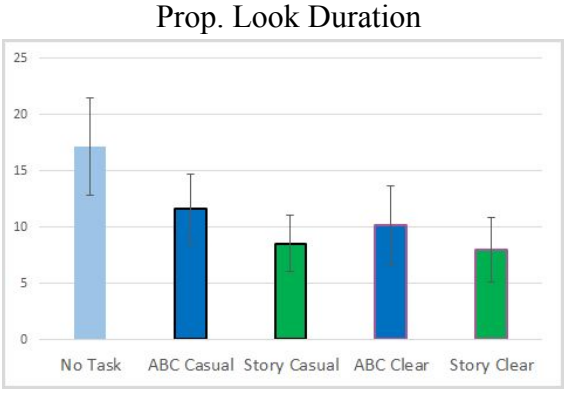


Results

We hypothesized that participants would spend more time looking at speech relevant areas of the face (nose/mouth), when the speech task was clear or informative, and would focus more on the eyes during casual or storytellings tasks.

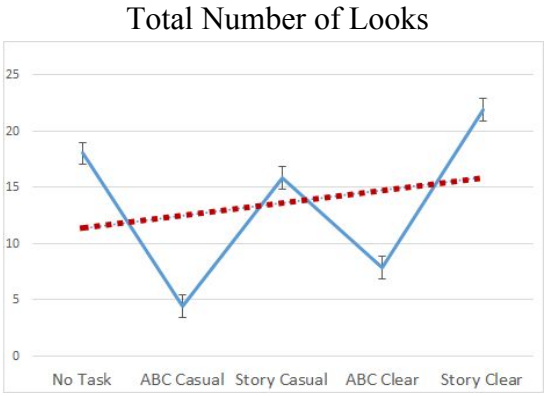
The grand mean for proportional duration looking to the eyes was *G*=11.081 *(SD*=2.831*)*. In the no-task condition, participants had significantly higher proportional looking time directed towards eyes *(M*=17.173*, SD*=4.343*)* than in ABC casual *(M*=11.618*, SD*=3.042*)* *F(*1, 19*)* = 5.55*, p* = .021*),* story casual condition *(M*=8.498*, SD*=2.502*) F(*1,19*)* = 8.675*, p* = .06, and story clear condition *(M*=7.942*, SD*=2.882*) F(*1, 19*)* = 9.231*, p* = .002*. (see figure 3)*

Figure 3



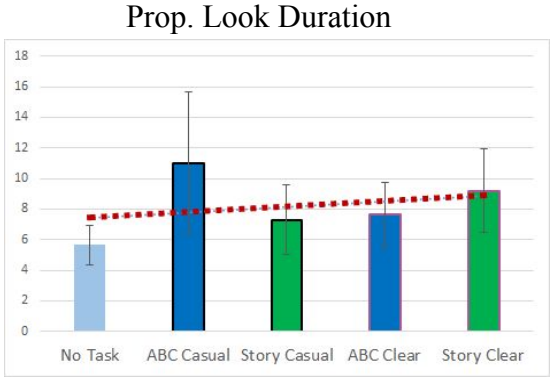
The grand mean for total number of looks to the eyes was *G*=11.59 *(SD*=21.6366). No significant difference was found between the no-task condition and ABC clear condition. Total number of looks to the eyes was consistent across all tasks. (*see figure 4)*

Figure 4



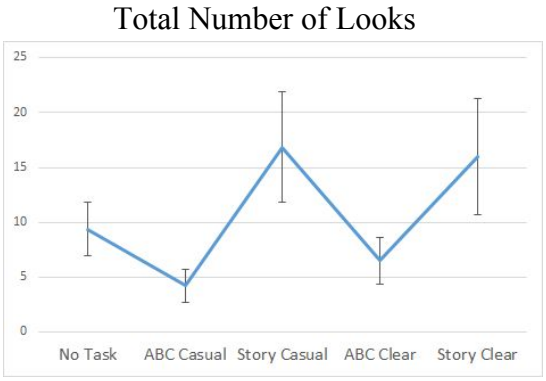
For proportional look duration to the mouth and nose, *G=8.159 (SD = 2.239)* No significant difference was found between ABC casual condition and ABC clear condition in terms of look count to the mouth and nose. Proportional look duration to the mouth and nose was consistent across all tasks. (*see figure 5)*

Figure 5



For looks to the mouth and nose, *G*=16.6 *(SD* = 14.56). In the ABC casual task, there were significantly fewer total number looks to the mouth and nose *(M*=4.4, *SD*=1.118) than in the no-task conditions *(M*=9.4, *SD*=10.753*) F(*1,19*)* = 5.15*, p = .006,* the story casual condition *(M*=16.85, *SD*=22.399*) F(*1,19*)* = 12.6, *p* =.012, and the story clear condition *(M*=16*,* *SD*=23.822) *F(*1,19*)* = 11.75*, p* = .019. *(see figure 6)*

Figure 6



In addition to the above analyses, the proportional look time durations across different tasks and AOIs were examined for correlational relationships. The results are summarized in Table 1. In general, looks to the face were more strongly correlated with looks to the mouth and nose in storytelling conditions. A significant correlation was found between looks the face x looks to the mouth/nose in the story casual task *(r=.553) p<.05,* and in the story clear task *(r=.559) p<.05. (see table 1)*

Table 1

AOI Correlations by Task

Task	*Face/Eyes*	*Face/Nose & Mouth*
No Task	*0.625*	*.261*
ABC Casual	*.33*	*.206*
Story Casual	*0.573*	*0.553*
ABC Clear	*-.155*	*.247*
Story Clear	*.38*	*0.559*

Discussion

Seeing oneself has been demonstrated to increase an individual's self-awareness and effort in task-completion (Brickenkamp & Zillmer, 1998), but this effect has not been documented within the realm of speech. The present research sought to investigate the effects of seeing oneself while speaking. In order to achieve this, participants were asked to complete a variety of speech acts while viewing themselves speaking in real-time on a computer monitor. Given the lack of background research to draw upon, our hypothesis were tentative, and the scope of the study was naturally broad. However, our results do not seem to support our initial hypotheses that that looks to the mouth and nose will increase during informative and clear speech tasks, while looks to the eyes will increase during storytelling and casual speech tasks. Participants tended to give a higher number of looks to their eyes when not speaking, and spent proportionally less time looking at their nose/mouth when saying the ABC's casually. In addition, looks to speech relevant areas of the face (nose and mouth), were more heavily correlated in more complex speech tasks. To our knowledge, this is the first study to address how seeing oneself effects attention during speech.

The lack of an effect found regarding differences in attention across storytelling and informative speech tasks, suggests that this type of manipulation is not affected by the presence one's self-image. The original hypothesis was extrapolated from research conducted explicitly in conversational settings, that suggested looks to the mouth and nose increase during informative tasks, while looks to the eyes increase during emotional speech tasks (Buchan, Pare´, & Munhall, 2008). There was, as stated previously, very little research overall to draw upon in regards to our specific area of study. We were not sure how attention might be affected by viewing oneself

during speech, so we decided to include several major types of speech (ie storytelling, informative, clear, casual). In Buchan's 2008 study it is suggested that looks to the mouth and nose during informative speech are a result of visual feedback integration during speech perception. This effect might present itself in a completely different way when an individual is watching themselves speak rather than watching other people speak. Our original reasoning was that participants would be inclined to integrate visual feedback into speech production in order to speak as clearly as possible in informative speech tasks. However, our results indicated a correlation between looks to the face and looks to the mouth and nose during both clear and casual storytelling tasks. Storytelling is a more complex speech tasks than the ABCs, and as a result may tend to draw participants eyes to the speech relevant areas of the face (mouth/nose).

While our findings regarding gaze fixation during informative speech do not match the results from conversation studies, our finding regarding looks to the eyes in the no-task condition does mimic previous research (Lansing, McConkey, 2003). Specifically, we found that participants spent proportionally more time looking at their eyes during the no-task condition. While Lansing and McConkey's study focused on speech in a conversational setting, they found that participants looked more towards the eyes of their conversation partner in the silence or no-speech conditions. This suggests that the eyes are particularly attention grabbing when an individual is idle, and in a broader sense, that attention is passively drawn to the eyes more than other parts of the face. Lansing and McConkey do not offer a cause for the increase look-time to the eyes during silence, but in their case it may be a result of the eyes conveying non-speech information. It may also be a cultural effect, meaning that individuals look each other in the eyes out of a desire to be polite. This however, does not explain why we found the same effect when

an individual is looking at themselves. They would not be inspired to do so out of a sense of politeness, and they also would likely not be doing so in order to garner emotional information displayed on their faces. Without further information, perhaps it is enough to simply posit that the eyes are particularly interesting for some reason, and that this effect carries over to the present study's unique field of focus.

Another effect we found that warrants further research regards a possible time compression effect observed in the looks to the mouth and nose in ABC casual tasks. We found that while the proportional look duration was the same across all tasks, there were significantly fewer looks to the mouth and nose in the ABC casual task. Given that the number of looks was lower, but the proportional duration was the same across tasks, it stands to reason that the duration of each look must be longer. While task times were not controlled, the ABC casual task was consistently the fastest task in terms of time, with many participants completing the task in just a few seconds. Perhaps then, the long look duration in ABC casual can be explained by a floor on how long a look needs to be in order to garner useful information. It is also interesting that this effect did not carry over to ABC clear task. To explain this, we might look again to the variation in the time is takes to complete each task. Specifically, the ABC clear condition took more time on average than the ABC casual condition, likely due to participants slowing down their speech in order to make it more intelligible.

While our research yielded interesting results regarding attention during speech acts in the presence of one self-image, there were several methodological problems and obstacles we ran into while building the study. Because the study was to be structured as an undergraduate thesis, there was limited time in which we could build the study. As a result, our initial plan to use a

mirror as the feedback device for the participant fell through. The initial decision to use a mirror was however, not made absent mindedly. Based on our research regarding conversational speech and the effect of self-image on task completion, we determined that a mirror would be the most effective means of displaying the participant's image (Lansing & McConkey, 2003; Buchan, Pare´, & Munhall, 2008; Brickenkamp & Zillmer, 1998; Silvia, 2012). Much of the research regarding the effect of being in the presence of one's self-image was done by simply placing mirrors in the room while a participant completed a task (Silvia, 2012). Use of a mirror would also better mimic conversational settings by having the feedback be real time, as opposed to on a short delay. Another important point is that the participants were not looking at themselves in the eyes when they spoke. Rather, they were looking at their face from an angle created by placing the webcam at the top of the computer monitor. As a result, our ability to relate our results back to conversational settings in which both participants would likely be making periodic eye contact, is limited.

Taking all of these issues into account, the main goal of future research should be to design the study such that the a mirror can effectively be used. This might be possible using the eye-tracking apparatus currently available, but further investigation is needed to determine whether or not this is the case. Future researchers should also take into account the lack of an effect when comparing most speech tasks. Specifically, clear vs casual speech tasks did not result in any significant changes in speech. The presence of an effect in both the storytelling conditions, and the no task condition, suggests that these should be the primary areas of focus for future research. While the effect found in no-task condition mimicked earlier research conducted on conversation partners (Lansing & McConkey, 2003), the correlation between face looks and

mouth/nose looks during the storytelling conditions raises questions as to the nature of storytelling as a speech goal. Future research should focus on determining what it is about storytelling tasks as compared to informative tasks, that leads to this correlation. This might be accomplished by introducing a much larger variety of informative and storytelling speech tasks.

While our findings did not indicate significant differences between most speech types and goals, our results suggest that visual feedback is a unique form of feedback does appear to have some effect on attention during speech. Eye in particular, appear to be a particularly attention grabbing area of the face, especially when an individual is not engaged speech activity. Ultimately this study revealed much in terms of how future research in the field of self-image and speech should be conducted. Specifically, we offer tentative suggestions as to which modes of speech are most affected by the variable of self-image, as well as propose several technical obstacles future researchers should take into account.

References

Alsius, A., Wayne, R. V., Paré, M., & Munhall, K. G. (2016). High visual resolution matters in audiovisual speech perception, but only for some. *Attention, Perception, & Psychophysics*, *78*(5), 1472-1487.

Brédart, S., Delchambre, M., & Laureys, S. (2006). Short article one's own face is hard to ignore. *The Quarterly Journal of Experimental Psychology*, *59*(1), 46-52.

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

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of cognitive neuroscience*, *15*(1), 57-70.

Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*(2), 543-571.

Davis, C., & Kim, J. (2004). Audio‐visual interactions with intact clearly audible speech. *The Quarterly Journal of Experimental Psychology Section A*, *57*(6), 1103-1121.

Devue, C., Laloyaux, C., Feyers, D., Theeuwes, J., & Brédart, S. (2009). Do pictures of faces, and which ones, capture attention in the inattentional-blindness paradigm?. *Perception*, *38*(4), 552-568.

Everdell, I. T., Marsh, H., Yurick, M. D., Munhall, K. G., & Paré, M. (2007). Gaze behaviour in audiovisual speech perception: Asymmetrical distribution of face-directed fixations. *Perception*, *36*(10), 1535-1545.

Feld, J. E., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1555-1565.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological bulletin*, *131*(1), 76.

Jones, J. A., & Striemer, D. (2007). Speech disruption during delayed auditory feedback with simultaneous visual feedback. *The Journal of the Acoustical Society of America*, *122*(4), EL135-EL141.

Keyes, H., Brady, N., Reilly, R. B., & Foxe, J. J. (2010). My face or yours? Event-related potential correlates of self-face processing. *Brain and cognition*, *72*(2), 244-254.

Kowal, S., Wiese, R., & O'Connell, D. C. (1983). The use of time in storytelling. *Language and Speech*, *26*(4), 377-392.

Lander, K., & Capek, C. (2013). Investigating the impact of lip visibility and talking style on speechreading performance. *Speech Communication*, *55*(5), 600-605.

Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Attention, Perception, & Psychophysics*, *65*(4), 536-552.

Lusk, L. G., & Mitchel, A. D. (2016). Differential gaze patterns on eyes and mouth during audiovisual speech segmentation. *Frontiers in psychology*, *7*.

Mattheyses, W., & Verhelst, W. (2015). Audiovisual speech synthesis: An overview of the state-of-the-art. *Speech Communication*, *66*, 182-217.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*(4), 527-545.

Silvia, P. J. (2012). Mirrors, masks, and motivation: Implicit and explicit self-focused attention influence effort-related cardiovascular reactivity. *Biological Psychology*, *90*(3), 192-201.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*, *3*(1), 236-264.

Tacikowski, P., & Nowicka, A. (2010). Allocation of attention to self-name and self-face: An ERP study. *Biological psychology*, *84*(2), 318-324.

Theune, M., Meijs, K., Heylen, D., & Ordelman, R. (2006). Generating expressive speech for storytelling applications. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(4), 1137-1144.

Vatikiotis-Bateson, E., Eigsti, I. M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Attention, Perception, & Psychophysics*, *60*(6), 926-940.

Windheim, K., Veale, D., & Anson, M. (2011). Mirror gazing in body dysmorphic disorder and healthy controls: Effects of duration of gazing. *Behaviour research and therapy*, *49*(9), 555-564.