Published in final edited form as:
J Acoust Soc Am. 2007 October ; 122(4): EL135–EL141. doi:10.1121/1.2772402.

Speech disruption during delayed auditory feedback with simultaneous visual feedback

Jeffery A. Jones and
Centre for Cognitive Neuroscience and Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada, jjones@wlu.ca

Danielle Striemer
Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada dstriemer@gmail.com

Abstract

Delayed auditory feedback (DAF) regarding speech can cause dysfluency. The purpose of this study was to explore whether providing visual feedback in addition to DAF would ameliorate speech disruption. Speakers repeated sentences and heard their auditory feedback delayed with and without simultaneous visual feedback. DAF led to increased sentence durations and an increased number of speech disruptions. Although visual feedback did not reduce DAF effects on duration, a promising but nonsignificant trend was observed for fewer speech disruptions when visual feedback was provided. This trend was significant in speakers who were overall less affected by DAF. The results suggest the possibility that speakers strategically use alternative sources of feedback.

1. Introduction

Numerous laboratory studies have shown that altering the auditory feedback speakers hear affects ongoing speech production. For example, exposing speakers to increases in environmental noise causes speakers to increase their speaking volume and the duration of their utterances (Lane and Tranel, 1971; Bauer *et al*., 2006). Selectively filtering frequencies (Garber and Moller, 1979) or modifying the fundamental frequency (Elman, 1981; Burnett *et al*., 1997; Kawahara, 1998; Jones and Munhall, 2000) or the formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006) elicits compensatory productions that mitigate the alterations. However, the most profound disruptions to ongoing vocal productions result when speakers hear their ongoing speech delayed (Lee, 1950). Exposure to delayed auditory feedback (DAF) often results in “stutterlike” disturbances in fluency (Fairbanks, 1955; Fairbanks and Guttman, 1958). These disturbances include speaking-rate decreases, increased speech intensity and pitch, syllable repetitions and omissions, and misarticulation (Black, 1951; Atkinson, 1953; Yates, 1963; Howell and Archer, 1984).

The profound effects caused by DAF (and the effects of certain other altered feedback conditions) led to speculation that speech production is monitored in a closed-loop manner (Lee, 1950; Fairbanks, 1954). According to these servomechanistic accounts, a comparator looks for discrepancies between the intended output of a vocal production and the sensory feedback; the disruptions observed during DAF are a manifestation of the corrective action initiated to overcome the perceived mismatch. However, Borden (1979) contended that speech rate is too quick for auditory feedback to be processed and the corrections implemented before the next segment is produced. Moreover, Howell and Archer (1984) showed that when a 500-Hz square wave matching the amplitude envelope of the speaker’s speech was substituted for the delayed speech signal, speakers suffered a similar reduction in their speech rate, as they did when they heard their true speech signal delayed. Indeed, DAF effects are not limited to

speech behaviors and are found with other motor behaviors such as tapping and music production (Chase *et al*., 1959; Smith *et al*., 1960; Finney and Warren, 2002). Thus, the root cause of DAF effects appears to be due to a more general disruption of the temporal relationship between production and acoustic input and not due to the fact that the feedback system receives incorrect information about the specific articulations (Howell and Sackin, 2002).

There has been limited research conducted on the remedial effects of providing alternative forms of synchronous feedback simultaneously with DAF. Howell and Archer (1984) reported that increasing the volume of DAF led to increased levels of disruption. This finding may suggest that speakers can use their veridical feedback to reduce DAF effects (either transmitted through bone or air). However, given that DAF effects are likely the result of the detection of global asynchronies between production and feedback, it is probably not necessary that an alternative source of feedback provide more than a crude indication of synchrony. Another naturally synchronous signal during speech is the visible movements of the speaker's face. A wealth of research has shown that listeners readily use visual speech cues: whether in noisy environments or under optimal listening conditions, information from a speaker's face significantly enhances auditory intelligibility (Sumby and Pollack, 1954; Davis and Kim, 2004). The current study investigates whether providing speakers with visual feedback regarding the timing of their ongoing speech can reduce DAF effects.

An earlier study conducted by Tye-Murray (1986) found no evidence that visual information could be used to improve speech production. In her study, Tye-Murray asked 11 volunteers to say eight sentences while hearing their voice delayed, with and without the availability of a mirror to monitor their productions. However, Tye-Murray only examined sentence duration and did not look at the number of speech disruptions that occurred. The results showing that the duration of sentence repetition was unaffected by the presence of the visual feedback do not completely exclude the possibility that the number of speech disruptions was reduced. In the current study speakers heard their auditory feedback delayed by 180 ms while they produced sentences with and without visual feedback. Any moderating effect of providing visual feedback was evaluated by measuring both sentence duration and the number of speech disruptions. We additionally looked at whether the provision of visual feedback would differentially affect those speakers who were more affected by DAF compared to those speakers who were less affected. One might predict that if participants experience very few speech disruptions under DAF, it is unlikely that the availability of visual feedback will further reduce the small number of speech disruptions that occur. However, individuals who experience a greater degree of disruption under DAF may benefit more from the availability of visual feedback. Alternatively, it could be the case that individuals who experience fewer speech disruptions under DAF are *better* able to integrate alternative sources of sensory information to aid in speech production. Conversely, individuals who experience greater speech disruption may be *less* able to integrate other sources of sensory information to aid their speech production.

2. Methods

2.1 Participants

Twenty-two right-handed men (mean age 21.6 years) participated in this study. However, data from only 20 participants were analyzed because two of them were statistical outliers (i.e., their data values were over three times the interquartile range, above the third quartile). All 20 remaining participants were university students with no reported prior neurological damage, or speech or language disorders. All participants had normal or corrected-to-normal vision. The procedures were approved by the Wilfrid Laurier University Research Ethics Board and all participants gave informed consent.

2.2 Apparatus and procedure

Participants sat in a double-walled sound booth and wore headphones (Sennheiser HD 280 Pro) and a headset microphone (AKG C420) during the experimental session. Participants' vocal productions were recorded as they repeated the same ten sentences in each experimental condition. Each sentence was made up of between five to eight words (eight to ten syllables). To familiarize participants with the stimuli, they were asked to read the sentences aloud before commencing the experiment. On each trial, participants heard a recording of a sentence and were asked to repeat the sentence at a consistent pace.

Each speaker participated in the six following conditions: (1) The "NAF" (normal auditory feedback) condition was a control condition and served as a baseline for the other five experimental conditions. In this condition, participants repeated the ten sentences while receiving unaltered auditory feedback. They simultaneously stared at a black fixation cross in a 6×6 in. white square over a blue background on a 17-in. monitor. (2) The "DAF" condition required participants to stare at the fixation cross on the screen and repeat the same sentences while their auditory feedback was delayed by 180 ms using a digital signal processor (Tucker-Davis Technologies, RX6 Multifunction Processor). (3) The "NAF mirror" condition required participants to repeat the sentences while hearing their auditory feedback presented without delay, while viewing the movements of their face in a mirror measuring 7 in. in diameter at an approximately constant distance. The mirror and microphone were adjusted so the participants could view their mouth movements as the sentences were recited. (4) The "DAF mirror" condition required participants to repeat the sentences while their auditory feedback was delayed and they simultaneously received visual feedback regarding their mouth movements reflected by the mirror. (5) The "NAF sentence" condition required participants to repeat the sentences while reading the sentences presented orthographically on the 17-in. monitor and hearing their auditory feedback presented without delay. (6) The "DAF sentence" condition required participants to repeat the sentences while their auditory feedback was delayed and they read the sentences on the monitor. Comparing performance during this condition to performance during the other DAF conditions served to ensure that any deficits observed did not result from misremembering the sentences.

Sentence order was randomized across conditions and the order of conditions was counterbalanced across participants. Gaussian noise was presented throughout the experimental session in an effort to mask the participants' real time auditory feedback.

2.3 Data analysis

The durations of utterances were determined manually using Praat (Boersma, 2001). Speech disruptions were identified as a word or syllable repetition, part-word prolongation, inaudible postural fixation, or misarticulation of a word. Both sentence duration and number of speech disruptions were analyzed separately using a 2×3 repeated measures ANOVA with auditory feedback (normal, delayed) and visual cue (fixation, mirror, sentence) as the within subject factors. To determine whether participants who were more affected by DAF responded differently to the presence of visual feedback compared to those participants who were less affected, we performed a median split based on the number of speech disruptions that occurred for each participant during the "DAF" baseline condition. Two groups were formed: a "low"-disruption group (n=10) and a "high"-disruption group (n=10). The number of speech disruptions made by these two groups was analyzed using the same 2×3 repeated measures ANOVA described above.

3. Results

Two participants were identified as statistical outliers and thus were not included in the analysis. Data from the remaining 20 participants were analyzed. Results of the statistical analysis of sentence duration revealed a main effect of auditory feedback [$F(1,19) = 39.54, p < 0.001$]; sentence durations in the DAF conditions were longer than sentence durations in the NAF conditions [see Fig. 1(a)]. No other significant effects were observed for sentence duration. The results for the number of speech disruptions also revealed a main effect of auditory feedback [$F(1,19) = 23.80, p < 0.001$] with a greater number of disruptions made in the DAF conditions than in the NAF conditions [see Fig. 1(b)]. Although there was not a significant interaction between auditory feedback condition and visual cue condition [$F(1,19) = 2.099, p = 0.137$] there was a visible trend for a higher number of speech disruptions to occur during the DAF compared to the DAF with mirror (facial) feedback conditions.

A median split was performed by dividing the participants into a high-disruption group ($n=10$) and a low-disruption group ($n=10$) based on the median number of speech disruptions experienced during the DAF baseline condition. The statistical analysis of the high-disruption group revealed a main effect of auditory stimulus [$F(1,9) = 38.00, p < 0.001$] with DAF conditions eliciting a greater number of speech disruptions than the NAF conditions [see Fig. 2(a)]. No other effects were significant in the high-disruption group. For the low-disruption group, Mauchly's test revealed that there was a violation of sphericity [$W(2) = 0.47, p = 0.048$], so a multivariate analysis was used. The multivariate analysis revealed a significant main effect of auditory stimulus [$F(1,9) = 10.33, p = 0.011$] with DAF conditions leading to more speech disruptions than the NAF conditions [see Fig. 2(b)]. In addition, there was a significant interaction between auditory stimulus and visual stimulus [$F(2,8) = 4.67, p = 0.045$]. Posthoc paired samples *t*tests revealed significantly fewer speech disruptions occurred during the DAF mirror condition compared to the DAF sentence condition [$t(9) = 2.251, p = 0.05$]. No other significant differences were found.

4. Discussion

In the present study, sentence duration and number of speech disruptions were examined in order to determine whether visual cues could reduce the speech disruption typically observed with exposure to DAF. Delaying auditory feedback did consistently lead to increased sentence duration, but providing visual feedback simultaneously with DAF did not reduce this effect. Moreover, all DAF conditions led to a larger number of speech disruptions than when real time auditory feedback was available. In addition, although not reliable, on average a lower disruption rate was observed when speech was delayed and visual feedback was provided compared to when speakers fixated on a fixation point on a computer monitor or read the sentence on the screen. This trend was also apparent when we split our participants into a high-disruption group and a low-disruption group. However, the trend was only statistically significant for the low-disruption group who experienced fewer disruptions when they viewed their face with DAF than when they read the sentences.

Our results both replicate and extend those observed by Tye-Murray (1986). Tye-Murray also found that providing speakers with visual feedback regarding their speech production under conditions with DAF did not reduce durational effects. Based on those finding, she concluded that visual feedback does not diminish the effects of DAF. However, we found promising trends that suggest that visual feedback might instead reduce the number of speech disruptions that speakers experience under DAF.

It is well known that listeners readily use visual speech cues (i.e., facial movements) during speech perception (e.g., Davis and Kim, 2004). In fact, brain imaging work demonstrates that

even in the absence of audible speech, visual speech cues activate the supratemporal auditory cortex in normal-hearing individuals (Calvert *et al*., 1997). However, we believe that it is unlikely that DAF effects arise from a close monitoring of the precision of production. Rather, we believe that DAF effects arise from monitoring a signal that is globally asynchronous (Howell and Sackin, 2002). The visual feedback we provided to speakers was inherently synchronous with their ongoing production and speakers may have been able to monitor this alternative feedback to some degree, thereby reducing the speech disruption caused by exposure to DAF.

Although consistent, this reduced speech disruption was statistically unreliable except for those speakers who performed best under the DAF condition. The reason for this is unclear, but it is known that speakers do adopt strategies to overcome DAF effects (Katz and Lackner, 1977). One possible strategy that speakers may adopt is to ignore the auditory feedback they receive and use alternative feedback pathways. Thus, it is possible that those speakers who experienced fewer speech disruptions were better able to ignore their auditory feedback and focus on alternative feedback sources that were synchronous with their production (e.g., proprioception). The addition of the visual feedback may have enhanced their ability to ignore the DAF and use visual feedback as an alternative and congruent feedback source. It is possible that both high and low performing speakers could be trained to take advantage of visual feedback. Indeed, previous research has demonstrated that DAF effects can be reduced with practice (Goldiamond *et al*., 1962). Perhaps prolonged practice with visual feedback could lead to more pronounced reductions in the disruption rates we observed.

The possibility that visual feedback can ameliorate some of the effects caused by DAF has implications for understanding the remedial effects of similar treatments for certain speech disorders. For example, Lee (1950) was among the first to note the similarity of DAF effects to stuttering and since then a number of researchers have hypothesized the irregular use of auditory feedback as a cause of stuttering (Fairbanks, 1954; 1955; Max *et al*., 2004). This view was paradoxically bolstered by reports that dysfluency could be reduced by providing people who stutter with DAF (Kalinowski *et al*., 1993; 1996; Van Borsel *et al*., 2003; Alm, 2006). How altered auditory feedback reduces stuttering is the subject of intense debate. However, our results are especially interesting in light of Kalinowski and colleague's recent report that fluency is enhanced when people who stutter view the silent articulations of another speaker producing the same utterance as themselves (Kalinowski *et al*., 2000).

In closing, our data suggest speakers may be able to use visual feedback to reduce the speech disruption that results from exposure to DAF. The literature on DAF has in large part focused on the strategies that speakers use to regain fluency under DAF conditions. An accepted assumption of some researchers is that speakers learn to ignore auditory feedback and may avail themselves of alternative feedback sources such as proprioception. These strategies may in fact underlie the beneficial effects observed in clinical populations where DAF is used as a treatment for stuttering. However, it is difficult to test these assumptions because most obvious forms of alternative feedback are difficult or impossible to manipulate experimentally. Visual feedback is easily manipulated and offers a window into speakers' strategic use of alternative sources of feedback. It will be interesting in future work to manipulate factors such as the salience of the visual feedback relative to the auditory feedback, as well as the amount of practice and training speakers receive.

Acknowledgements

A NSERC Discovery Grant and a Research Fellowship from Wilfrid Laurier University supported this work. We thank Dwayne Keough and Michelle Jarick for comments regarding an earlier draft.

References and links

Alm PA. Stuttering and sensory gating: A study of acoustic startle prepulse inhibition. Brain Lang 2006;97:317–321. [PubMed: 16388843]

Atkinson CJ. Adaptation to delayed sidetone. J Speech Hear Disord 1953;18:386–391. [PubMed: 13109865]

Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. J Acoust Soc Am 2006;119:2363–2371. [PubMed: 16642849]

Black JW. The effect of delayed sidetone upon vocal rate and intensity. J Speech Hear Disord 1951;16:56–60.

Boersma P. Praat, a system for doing phonetics by computer. Glot Int 2001;5:341–345.

Borden GJ. An interpretation of research on feedback interruption in speech. Brain Lang 1979;7:307–319. [PubMed: 455050]

Burnett TA, Senner JE, Larson CR. Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. J Voice 1997;11:202–211. [PubMed: 9181544]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. Science 1997;276:593–596. [PubMed: 9110978]

Chase RA, Harvey S, Standfast S, Rapin I, Sutton S. Comparison of the effects of delayed auditory feedback on speech and key tapping. Science 1959;129:903–904. [PubMed: 13635035]

Davis C, Kim J. Audio-visual interactions with intact clearly audible speech. Q J Exp Psychol 2004;A 57:1103–1121.

Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. J Acoust Soc Am 1981;70:45–50. [PubMed: 7264071]

Fairbanks G. Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem. J Speech Hear Disord 1954;19:133–139. [PubMed: 13212816]

Fairbanks G. Selective vocal effects of delayed auditory feedback. J Speech Hear Disord 1955;20:333–345. [PubMed: 13272227]

Fairbanks G, Guttman N. Effects of delayed auditory feedback upon articulation. J Speech Hear Res 1958;1:12–22. [PubMed: 13576494]

Finney SA, Warren WH. Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. Percept Psychophys 2002;64:896–908. [PubMed: 12269297]

Garber SR, Moller KT. The effects of feedback filtering on nasalization in normal and hypernasal speakers. J Speech Hear Res 1979;22:321–333. [PubMed: 491559]

Goldiamond I, Atkinson CJ, Bilger RC. Stabilization of behavior and prolonged exposure to delayed auditory feedback. Science 1962;135:437–438. [PubMed: 13899770]

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science 1998;279:1213–1216. [PubMed: 9469813]

Howell P, Archer A. Susceptibility to the effects of delayed auditory feedback. Percept Psychophys 1984;36:296–302. [PubMed: 6522222]

Howell P, Sackin S. Timing interference to speech in altered listening conditions. J Acoust Soc Am 2002;111:2842–2852. [PubMed: 12083218]

Jones JA, Munhall KG. Perceptual calibration of FO production: Evidence from feedback perturbation. J Acoust Soc Am 2000;108:1246–1251. [PubMed: 11008824]

Kalinowski J, Armson J, Roland-Mieszkowski M, Stuart A, Gracco VL. Effects of alterations in auditory feedback and speech rate on stuttering frequency. Lang Speech 1993;36:1–16. [PubMed: 8345771]

Kalinowski J, Stuart A, Rastatter MP, Snyder G, Dayalu V. Inducement of fluent speech in persons who stutter via visual choral speech. Neurosci Lett 2000;281:198–200. [PubMed: 10704777]

Kalinowski J, Stuart A, Sark S, Armson J. Stuttering amelioration at various auditory feedback delays and speech rates. Eur J Disord Commun 1996;31:259–269. [PubMed: 8944848]

Katz DI, Lackner JR. Adaptation to delayed auditory feedback. Percept Psychophys 1977;22:476–486.

Kawahara, H. Hearing voice: Transformed auditory feedback effects on voice pitch control. In: Rosenthal, DF.; Okuno, HG., editors. Computational Auditory Scene Analysis. Erlbaum; Mahwah, NJ: 1998. p. 335-349.

Lane H, Tranel B. The Lombard sign and the role of hearing in speech. J Speech Hear Res 1971;14:677–709.

Lee BS. Some effects of side-tone delay. J Acoust Soc Am 1950;22:639–640.

Max L, Guenther FH, Gracco VL, Ghosh SS, Wallace ME. Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering. Contemporary Issues in Communication Science and Disorders 2004;31:105–122.

Purcell D, Munhall K. Compensation following real-time manipulation of formants in isolated vowels. J Acoust Soc Am 2006;119:2288–2297. [PubMed: 16642842]

Smith WM, McCrary JW, Smith KU. Delayed visual feedback and behavior. Science 1960;132:1013–1014. [PubMed: 17820673]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am 1954;26:212–215.

Tye-Murray N. Visual feedback during speech production. J Acoust Soc Am 1986;79:1169–1171. [PubMed: 3700869]

Van Borsel J, Reunes G, Van den Bergh N. Delayed auditory feedback in the treatment of stuttering: Clients as consumers. Int J Lang Commun Disord 2003;38:119–129. [PubMed: 12745932]

Yates AJ. Delayed auditory feedback. Psychol Bull 1963;60:213–232. [PubMed: 14002534]

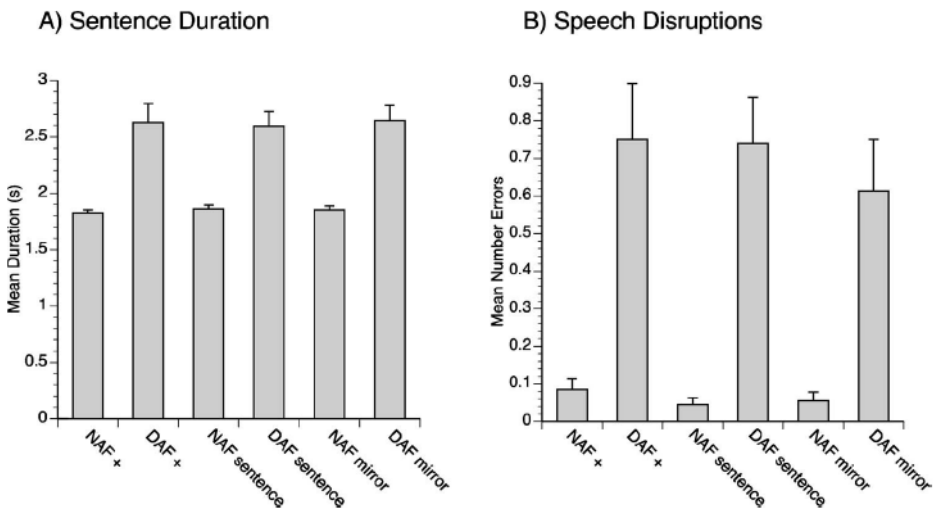

Fig. 1.

(a) Mean sentence duration for each of the six experimental conditions. (b) Mean number of speech disruptions for each of the six experimental conditions.

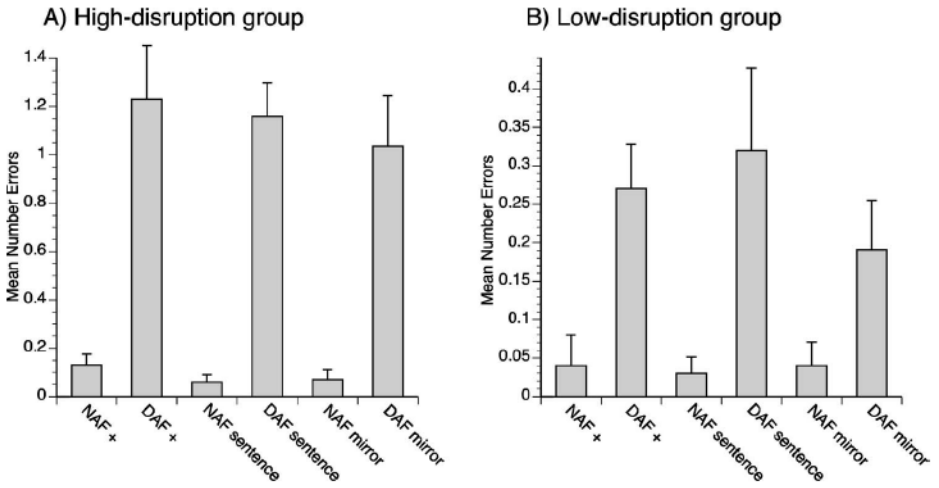

Fig. 2.

Mean number of speech disruptions for each of the six experimental conditions in the (a) "high-disruption" and (b) "low-disruption" groups.