Check for updates

The Relationship Between Attentional Capture by Speech and Nonfluent Speech Under Delayed Auditory Feedback: A Pilot Examination of a Dual-Task Using Auditory or Tactile Stimulation

Osamu Ishida[1,2]*, Daichi Iimura[1,3] and Shoko Miyamoto[4]

[1]*Graduate School of Comprehensive Human Sciences, University of Tsukuba, Ibaraki, Japan, [2]Saitama Municipal Nakamoto Elementary School, Saitama, Japan, [3]Japan Society for the Promotion of Science, Tokyo, Japan, [4]Faculty of Human Sciences, University of Tsukuba, Ibaraki, Japan*

Delayed auditory feedback (DAF) leads to nonfluent speech where the voice of a speaker is heard after a delay. Previous studies suggested the involvement of attention to auditory feedback in speech disfluency. To date, there are no studies that have revealed the relationship between attention and nonfluent speech by controlling the attention allocated to the delayed own voice. This study examined these issues under three conditions: a single task where the subject was asked to read aloud under DAF (single DAF task), a dual task where the subject was asked to read aloud while reacting to a pure tone (auditory DAF task), and a dual task where the subject was asked to read aloud while reacting to the vibration of their finger (tactile DAF task). The subjects also performed the single and dual tasks (auditory/tactile) under nonaltered auditory feedback where no delayed voices were involved. Results showed that the nonfluency rate under the auditory DAF task was significantly greater than that under the single DAF task. In contrast, the nonfluency rate under the tactile DAF task was significantly lower compared with that of the single DAF task. Speech became nonfluent when attention was captured by the same modality stimulus, i.e., auditory tone. In contrast, speech became fluent when attention was allocated to the stimulus that is irreverent to auditory modality, i.e., tactile vibration. This indicates that nonfluent speech under DAF is involved in attention capture owing to the delayed own voice.

Keywords: sensory modality, speech motor control, attentional capture, involuntary attention shift, communication disorders

INTRODUCTION

In day-to-day conversations, we adjust the sound pressure, pitch, and speech speed, among other parameters, based on our auditory feedback. This depends on the environment where speakers are present. Auditory feedback includes air and bone conduction. Air conduction feedback means that the voice is transmitted to a speaker's own eardrums as air vibrations, and the speaker perceives it. Bone conduction feedback means that the speaker perceives the vibration of their

cranial bones caused by articulation. These feedbacks are transmitted *via* the auditory nerve. It is difficult to adjust speech in environments where such auditory feedback is disrupted. In a loud environment, the sound pressure and pitch of the speech increase to compensate for the disruption of auditory feedback (Lombard effect, Siegel and Kennard, 1984).

In delayed auditory feedback (DAF), where the voice of speaker is heard with a delay on the order of a few tens of milliseconds to a few hundred milliseconds, nonfluent speech could occur in the form of word revisions (e.g., yesterday, Mom...Dad), syllable repetitions (e.g., m, m, Mom), or prolongations (e.g., Mo-m). This disfluency effect has been used in previous studies to investigate the relationship between speech and auditory feedback (e.g., Black, 1951; Yates, 1963; MacKay, 1968; Borden, 1979; Siegel et al., 1980; Fukawa et al., 1988; Stuart et al., 2002). In contrast, nonfluent speech occurs less frequently under normal auditory feedback (NAF), where the speaker is heard in real time with no delay. This difference in the occurrence rate of nonfluent speech between DAF and NAF is called the DAF effect (Lee, 1950; Black, 1951). Notably, nonfluent speech significantly increased when feedback was delayed for 200 ms compared with other time delay conditions (100 ms, 400 ms, and 800 ms; Fairbanks, 1955).

Auditory attention to delayed own voice has been suggested to be involved in such nonfluent speech in previous behavioral or neuroimaging studies (MacKay, 1968; Hashimoto and Sakai, 2003; Takaso et al., 2010; Ishida et al., 2019). Nonfluent speech rarely occurs even under DAF when it is difficult to read aloud, such as reading at a fast pace or reading color names that are different from the color of the letters (Stroop task; Fillenbaum, 1963; Zanini et al., 1999). We assume that one cannot pay attention to our speech in these conditions. Therefore, it is considered that speech becomes fluent when attention is focused on a sentence stimulus or speech movement, as this decreases the allocation of attention to DAF (the details have been provided in the "Discussion" section). Furthermore, neuroimaging studies have reported that cerebral blood flow in the superior temporal gyrus, which is involved in auditory attention, significantly increased under DAF compared to NAF (Hirano et al., 1997; Hashimoto and Sakai, 2003; Sakai et al., 2009; Nota et al., 2011); the positive correlation was found between the activity and the frequency of nonfluent speech (Hashimoto and Sakai, 2003; Takaso et al., 2010). These findings could imply the involvement of auditory attention toward nonfluent speech. To date, no studies have been conducted to examine the relationship between attention and nonfluent speech by controlling attention allocation to the delayed own voice under DAF. We assume that forcing the direction of attention toward some stimulus, i.e., DAF, could influence the fluency, and we could realize this situation by performing a dual task.

The phenomenon in which attention is directed to stimuli unrelated to the task, regardless of the subject's intention, is called attention capture (Theeuwes, 2010). In general, humans can actively allocate attention to and differentiate between stimuli (target stimuli) related to a task. In contrast, when the nontarget stimuli, which are highly salient or novel, or when unexpected stimuli (deviant stimuli) are present, such as in oddball tasks, attention is directed to that stimuli even if it is unrelated to the task (Yantis and Jonides, 1984; Theeuwes, 1991, 1994, 2010; Miller and Buschman, 2013). Attention allocation and capture are also related to task performance. The rate of correct answer improves when attention is allocated to task-related stimuli (Reeves and Sperling, 1986; Cheal and Lyon, 1989). In contrast, distracting effects arise when attention is captured by stimuli unrelated to the task, as the processing involved in the task is interrupted (Schröger and Wolff, 1998). Thus, the effect of attention capture differs depending on whether the disruption stimulus is a task-relevant stimulus.

In a dual task, a subject is asked to perform multiple tasks comprising a primary task and secondary task. According to the capacity theory of attention (Kahneman, 1973), and due to limited attentional resources, the attention resources allocated to each task are considered to be fewer than those allocated to a single task. While the performance decreased in some dual task paradigms compared to single tasks (Appelbaum et al., 2012; Finke et al., 2015; Marsh et al., 2015), performance increased in other dual tasks (Melzer et al., 2001; Swan et al., 2004). It is also reported that, by using the dual task of semantic or phonological similarity (auditory), the performance in a dual task varied between different sensory modalities and a single sensory modality (Salamé and Baddeley, 1982; Hirst and Kalmar, 1987).

We suppose that, under a dual task, nonfluent speech also occurs depending on the deviance of the disruption stimulus (i.e., unimodal/task-relevant or bimodal/task-irrelevant). We set a dual task paradigm in which subjects were asked to read aloud under DAF (this is a single task for control) or read aloud under DAF while reacting to auditory stimulation (unimodal, the modality of the concurrent task was relevant to the primary task) or tactile stimulation (cross-modal, the modality of the concurrent task was irrelevant to the primary task). We hypothesized that speech would become nonfluent when attention was allocated to the own voice under DAF (unimodal task), while nonfluent speech would become attenuated when attention was allocated to sensory feedback rather than to the own voice (cross-modal task).

There are also individual differences regarding the DAF effect (Umeo and Ichinose, 2008; Nota et al., 2011; Chon et al., 2012). As a significant positive correlation between the activity of the superior temporal gyrus and the frequency of nonfluent speech was reported (Hashimoto and Sakai, 2003; Sakai et al., 2009), auditory attention to the delayed own voice could also be explained in the context of individual differences in the DAF effect. Therefore, we hypothesized that speech fluency would increase for speakers who were capable of allocating attention to sensory feedback. We also investigated the relationship between the DAF effect and individual differences of the DAF effect.

SUBJECTS

We initially conducted an experiment for 22 healthy adults (14 men, six women; average age: 25.20 ± 7.42 years old, and schooling history: 15.95 ± 1.15 years). We calculated the sample size required for the present experiment (power = 0.8,

FIGURE 1 | Conditions of voice feedback.
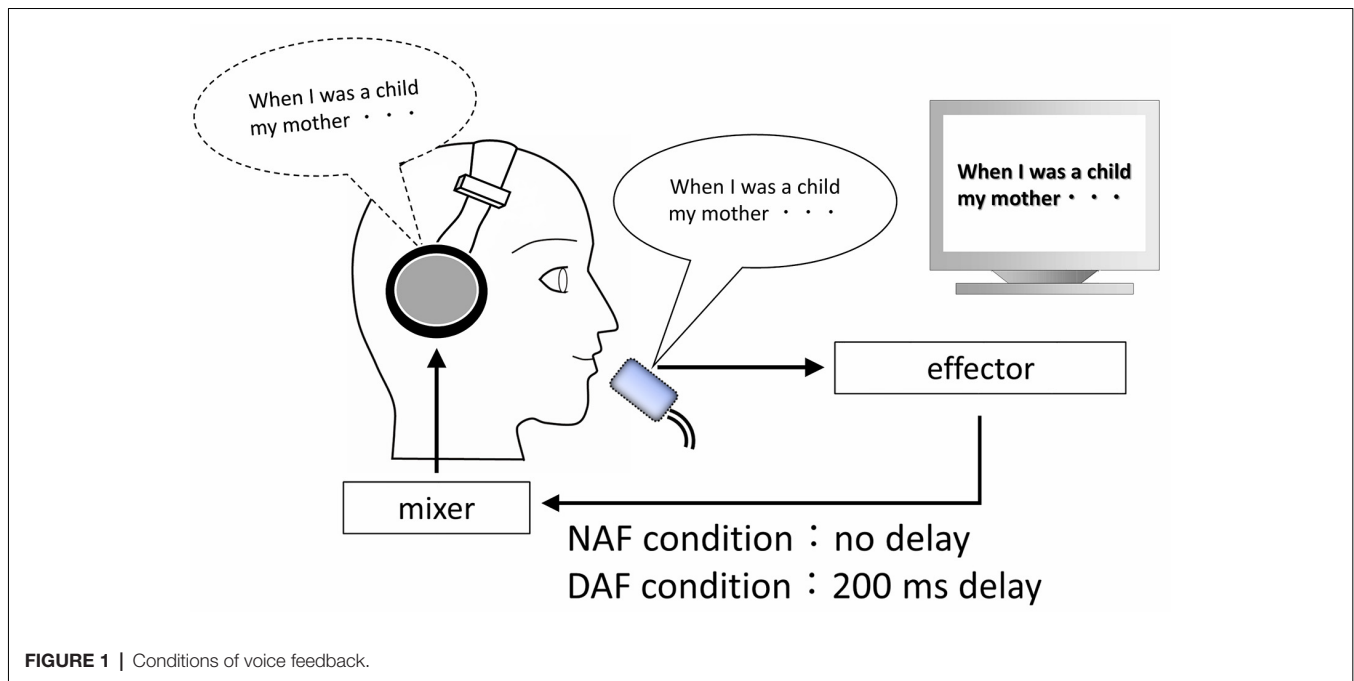
α = 0.05, *d* = 0.5; medium effect, i.e., Cohen, 1988) using G∗power3.1. (Faul et al., 2007). Finally, the subjects comprised 36 healthy adults (17 men, 19 women, average age: 23.56 ± 4.61 years old, schooling history: 16.39 ± 1.10 years). The average score of the analyzed subjects on the handedness test was 0.90 ± 0.25.

EXPERIMENTAL TASKS

Stimulation and Condition of a Single Task
This study was performed in accordance with the approval of Tsukuba faculty of human sciences (Permit number: 29–134). All subjects obtained written informed consent in accordance with the Declaration of Helsinki. We excerpted 18 sentences from a workbook for the Japanese Language Proficiency Test [NPO Research Institute for Japanese Language Education (1995)]. The average stimuli of the sentences were 154.83 ± 4.24 morae, and they were displayed horizontally in the center of a computer monitor with a character size of approximately 1 cm × 1 cm.

In the experiment, the subjects were asked to read aloud the sentences that appeared on the monitor placed 60 cm in front of their eyes. The subject's voice was captured during the task using a microphone (SHURE SM58) placed within 15 cm of his/her mouth, and this was provided as feedback through closed headphones (SONY MDR-7506) worn by the speaker over both ears through a mixer (MACIE 802VLZ4) and effector (LEXICON MX300). The headphones were used to prevent the subject's voice from coming outside the headphones and overlap with the subject's voice recorded by the microphone and played on the headphones. In addition, the subject's voice was recorded on an integrated circuit (IC) recorder (OLYMPUS Voice Trek V-863) with a microphone also placed within 15 cm of his/her mouth.

During the single task (reading-aloud task), the auditory feedback of the subject's voice was given under NAF or DAF (**Figure 1**). Under NAF, the subject's voice was feedback to the subject immediately with almost no delay (in real time), while he/she was reading the sentences aloud. In contrast to NAF, the subject's voice was provided as feedback to the subject with a delay of 200 ms by using an effector under single DAF. One reading-out-loud task lasted for 15 s per trial, and the subjects were asked to perform three trials under each condition with different reading stimuli.

The delay time was set to 200 ms in the present study because the frequency of nonfluent speech was high in this delay time compared to other delay times (Fairbanks, 1955; MacKay, 1968; Siegel et al., 1980; Stuart et al., 2002).

Stimulation and Condition of Dual Tasks
During the dual tasks, the subjects were asked to react to tactile or auditory stimuli while reading aloud sentences (**Figure 2**). In the tactile condition, which requires cross-modal attention allocation, the rod-shaped solenoid vibratory stimulation element of a 1-CH tactile stimulation device (UCHIDA FB-1200LP) was attached to the subject's left index finger by using surgical tape, and vibration stimuli were presented through the element. The stimuli were presented pseudorandomly 10 times in one trial with a Stimulus Onset Asynchrony (SOA) of 600 ms to 2,400 ms. We asked the subject to prioritize reading the sentences when it was displayed on the monitor and to press a button on the keyboard by using right index finger as quickly as possible every time if vibration stimulation was provided to the left index finger while reading. In the auditory condition, which requires unimodal attention allocation, a pure tone (1,000 Hz) was presented for 100 ms through a mixer and effector to the subject from the headphones.
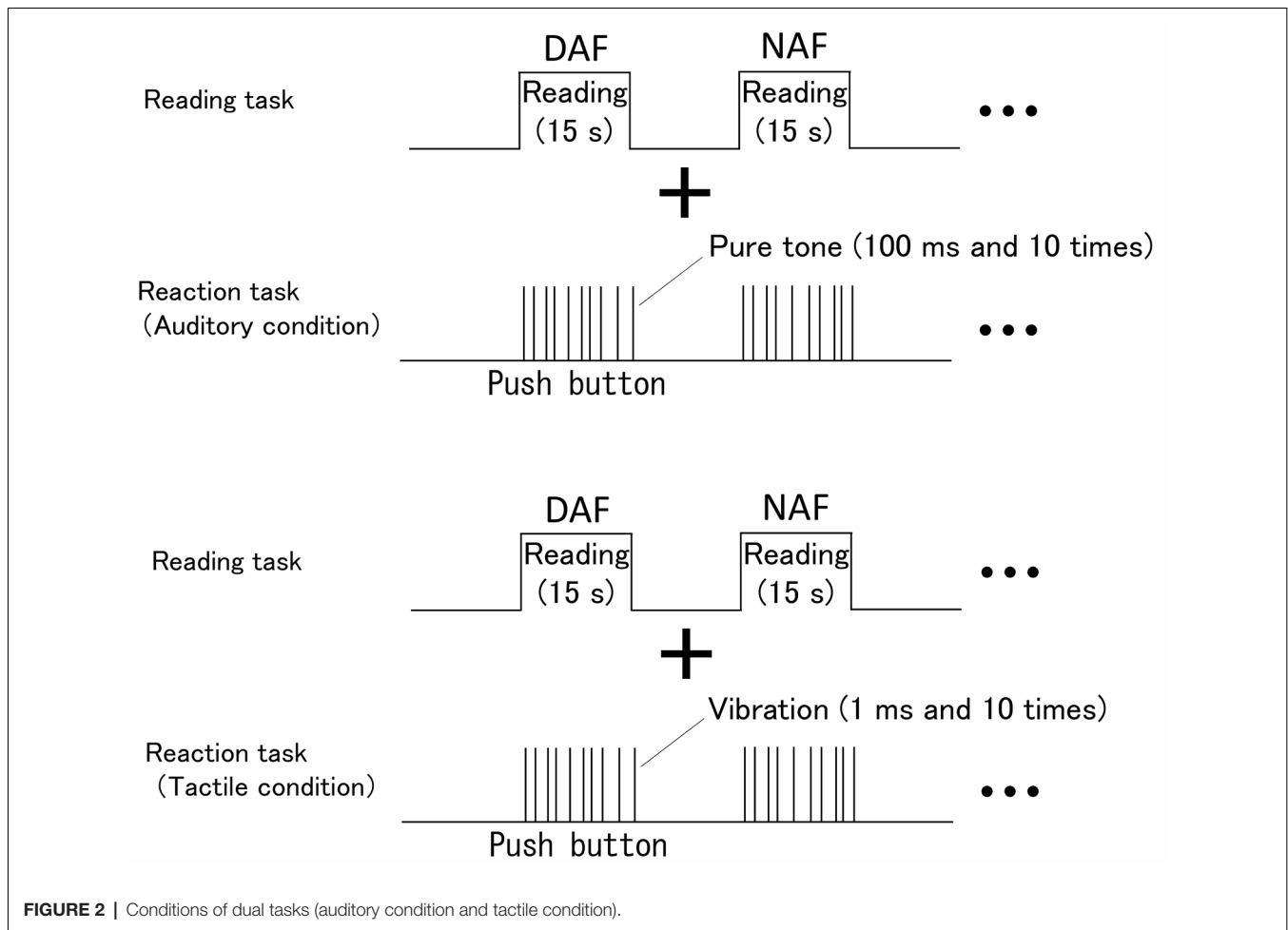
FIGURE 2 | Conditions of dual tasks (auditory condition and tactile condition).

The pure tone was presented 10 times pseudorandomly in one trial with an SOA of 600–2,400 ms. We asked the subject to prioritize reading the sentence when it was displayed on the monitor, and to press a button as quickly as possible every time if the pure tone was presented while continuing reading aloud.

The dual tasks involving tactile or auditory stimulation were both performed under the conditions of NAF and DAF. The same sentences of the single task were used. Each dual task was performed thrice with one trial lasting 15 s. E-Prime 2.0, a psychological experiment software, was used for presenting and controlling tactile and auditory stimulation.

PROCEDURE

We investigated the relationship between feedback and attention under six conditions: single NAF condition, single DAF condition, tactile NAF condition, tactile DAF condition, auditory NAF condition, and auditory DAF condition. Each of these six tasks were performed thrice, while alternating between NAF and DAF [e.g., tactile NAF condition (15 s) ⇒ tactile DAF condition ⇒ tactile NAF condition ⇒ tactile DAF condition ⇒ tactile NAF condition ⇒ tactile DAF condition ⇒ single NAF condition ⇒ single DAF condition . . .]. The sentences, the order of stimulus

modality (single/tactile/auditory), and the order of feedback condition (NAF → DAF or DAF → NAF) were counterbalanced among the subjects. A fixation point "+" was presented every 5 s before and after each trial, and the subjects were asked to gaze at the center of the point when presented.

ANALYSIS

We calculated the reaction time that elapsed between the presentation of stimulation and button press (within 600 ms) and the omission error rate for each subject. For the reaction time and omission error rate, we performed a 2 (NAF and DAF)-by-2 (tactile and auditory) intrasubject two-way analysis of variance (ANOVA) with Bonferonni correction. As we have analyzed the data twice (when the sample size was 22 and 36), an alpha level of 0.025 was set for the analysis in order to maintain an experiment-wise error rate of 5%.

In addition, a professional teacher calculated the nonfluency rate of each subject under each condition by calculating the proportion of nonfluent speech (nonfluent phrase/all phrases×100). Phrase is a unit that contains one independent word. Based on the work by Sakai et al. (2009), we divided nonfluency into nine classifications: (1) sound repetition (w,

	Reaction time (ms)		Omission error rate (%)	
	Tactile	**Auditory**	**Tactile**	**Auditory**
NAF	260.9 ± 58.7	312.0 ± 69.5	6.3 ± 5.7	8.7 ± 7.4
DAF	311.6 ± 85.3	365.6 ± 92.0	15.1 ± 12.2	16.7 ± 12.1

NAF, nonaltered auditory feedback; DAF, delayed auditory feedback.

w, when I ...); (2) part-word repetition (whe, when, I...); (3) word and phrase repetition (when I, when I was...); (4) prolongation (wh-en I); (5) break (when...I was...); (6) distortion; (7) syllable repetition (when, when I was ...); (8) revision (where, when I was...); and (9) error (where I was...). To ensure the measurement reliability of the nonfluency ratings, the nonfluency rate was recalculated by the same professional teacher 1 month after first assessment, and 10% segments were selected at random (intra-rater reliability). Inter-rater reliability was determined by a second professional teacher assessing nonfluent part for 10% of the sample selected at random. Then, an intraclass correlation coefficient analysis was conducted. We performed a 2 (NAF and DAF)-by-3 (single, tactile, and auditory) intrasubject two-way ANOVA to compare the nonfluency rate among different conditions. We also calculated the DAF effect (the difference in the nonfluency rate between the single NAF condition and single DAF condition) to investigate the susceptibility to DAF and the relationship between increased/decreased fluency in each modality. Furthermore, we calculated the Pearson correlation coefficient between the increased/decreased fluency rate (tactile DAF condition minus single DAF condition) under the tactile condition and DAF effect, and the coefficient between the increased/decreased fluency rate (auditory DAF condition minus single DAF condition) under the auditory condition and DAF effect. SPSS version 25.0 was used for statistical analysis.

RESULTS

Reaction Time and Omission Error Rate

The reaction time and omission error rate under the tactile and auditory conditions are listed in **Table 1**.

The mean of the reaction time in the tactile condition was 260.9 ± 58.7 ms under the NAF condition and 311.6 ± 85.3 ms under the DAF condition. The mean reaction time in the auditory condition was 312.0 ± 69.5 ms under the NAF condition and 365.6 ± 92.0 ms under the DAF condition. There was a significant main effect of the modality (tactile/auditory; $F_{(1,35)} = 42.2$, $p < 0.005$, $\eta_G^2 = 0.10$), and the reaction time was significantly shorter under the tactile condition than under the auditory condition. The feedback (NAF/DAF) also showed significant main effects ($F_{(1,35)} = 43.12$, $p < 0.005$, $\eta_G^2 = 0.11$), and the reaction time was significantly shorter under NAF than under DAF. No significant interaction between the modalities and feedback was observed ($F_{(1,35)} = 0.55$, $p = 0.82$, $\eta_G^2 = 0.00$).

The mean of the omission error rate in the tactile condition was 6.3 ± 5.7% under the NAF condition and 15.1 ± 12.2% under

the DAF condition. The mean of the omission error rate in the auditory condition was 8.7 ± 7.5% under the NAF condition and 16.7 ± 12.1% under the DAF condition. There were significant main effects of the feedback (NAF/DAF; $F_{(1,35)} = 30.3$, $p < 0.005$, $\eta_G^2 = 0.16$), and the omission error rate was significantly higher under DAF than under NAF. However, the omission error rate of the modalities (tactile/auditory) showed no main effects ($F_{(1,35)} = 2.5$, $p = 0.13$, $\eta_G^2 = 0.01$) and interaction between the modalities and feedback ($F_{(1,35)} = 0.3$, $p = 0.61$, $\eta_G^2 = 0.00$).

Nonfluency Rate

The nonfluency rates in the single task, dual tactile task, and dual auditory task are shown in **Figure 3**. With respect to nonfluent judgment, we obtained sufficient reliability of assessment (intra-rater reliability: 0.98 and inter-rater reliability: 0.72).

The mean nonfluency rate in the single task was 2.1 ± 2.9% under the NAF condition and 9.6 ± 6.7% under the DAF condition. The mean nonfluency rate in the tactile condition was 2.3 ± 2.2% under the NAF condition and 6.5 ± 4.7% under the DAF condition. The mean nonfluency rate in the auditory condition was 2.6 ± 3.4% under the NAF condition and 15.4 ± 8.4% under the DAF condition. There were significant main effects of the feedback (NAF/DAF: $F_{(1,35)} = 83.56$, $p < 0.005$, $\eta_G^2 = 0.38$) and modalities (single/tactile/auditory: $F_{(2,70)} = 38.84$, $p < 0.005$, $\eta_G^2 = 0.12$). The nonfluency rate was significantly higher under the DAF condition than under the NAF condition. Moreover, there was a significant interaction
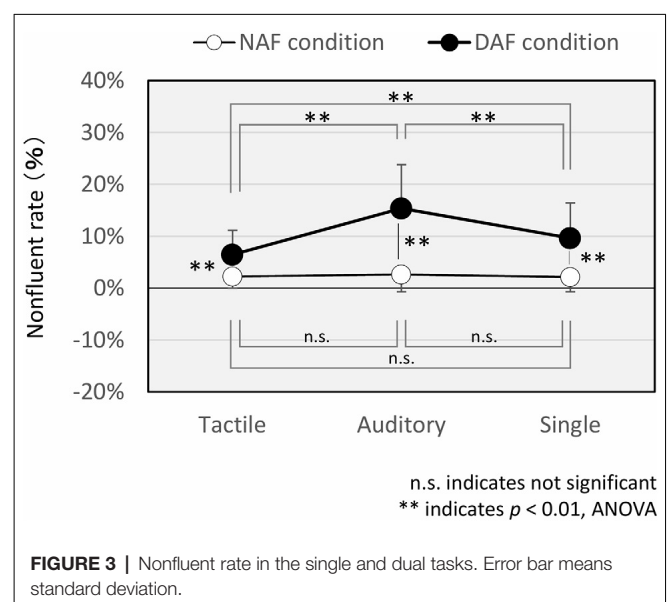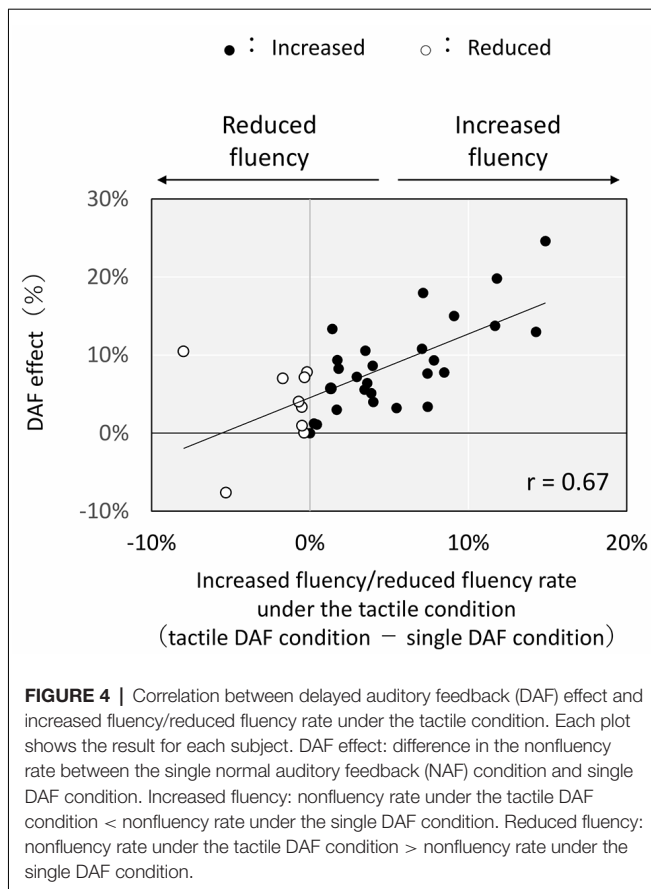

FIGURE 3 | Nonfluent rate in the single and dual tasks. Error bar means standard deviation.

FIGURE 4 | Correlation between delayed auditory feedback (DAF) effect and increased fluency/reduced fluency rate under the tactile condition. Each plot shows the result for each subject. DAF effect: difference in the nonfluency rate between the single normal auditory feedback (NAF) condition and single DAF condition. Increased fluency: nonfluency rate under the tactile DAF condition < nonfluency rate under the single DAF condition. Reduced fluency: nonfluency rate under the tactile DAF condition > nonfluency rate under the single DAF condition.

FIGURE 5 | Correlation between DAF effect and increased fluency/reduced fluency rate under the auditory condition. Each plot shows the results for each subject. DAF effect: difference in the nonfluency rate between the single NAF condition and single DAF condition. Increased fluency: nonfluency rate under the auditory DAF condition < nonfluency rate under the single DAF condition. Reduced fluency: nonfluency rate under the auditory DAF condition > nonfluency rate under the single DAF condition.

between the feedback and modalities ($F_{(2,70)} = 44.56$, $p < 0.005$, $\eta_G^2 = 0.10$). We performed a main-effect test, and there was no significant difference in the nonfluency rate among the different NAF conditions (single/tactile/auditory; $F_{(2,70)} = 0.63$, $p = 0.53$, $\eta_G^2 = 0.01$). In contrast, the nonfluency rate of the DAF condition was significantly lower under the tactile condition than under the single condition ($t_{(35)} = 3.56$, $p < 0.005$, $d = 0.54$). However, it was significantly higher under the auditory condition than under the single condition ($t_{(35)} = 7.46$, $p < 0.005$, $d = 0.75$). The nonfluency rate was significantly higher under the auditory condition than under the tactile condition ($t_{(35)} = 9.24$, $p < 0.005$, $d = 1.30$). In addition, there was a significant difference between NAF and DAF conditions for the single task and dual tasks with tactile and auditory stimulation (tactile: $F_{(1,35)} = 35.48$, $p < 0.005$, $\eta_G^2 = 0.25$; auditory: $F_{(1,35)} = 98.02$, $p < 0.005$, $\eta_G^2 = 0.50$; single: $F_{(1,35)} = 53.07$, $p < 0.005$, $\eta_G^2 = 0.34$). The nonfluency rate under the DAF condition was higher than under NAF conditions in all tasks (single/tactile/auditory).

Correlation With DAF Effect

There was a significant positive correlation between the DAF effect and the increased/decreased fluency rate (tactile DAF condition–single DAF condition) under the tactile condition (**Figure 4**; $r = 0.67$, $p < 0.005$). In contrast, there was no correlation between the DAF effect and the increased/decreased
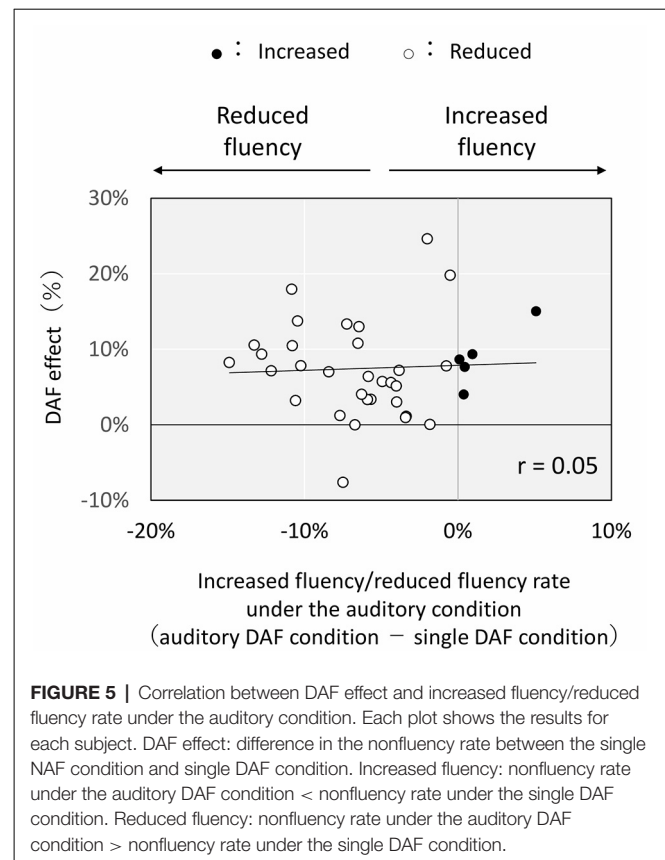
fluency rate (auditory DAF condition–single DAF condition) under the auditory condition (**Figure 5**; $r = 0.05$).

DISCUSSION

Relationship Between Delayed Auditory Feedback and Attention

This study investigated the relationship between nonfluent speech and auditory attention to vocal feedback by using reading tasks. In the single task, the subjects had to read aloud sentences by actively allocating attention to the sentences (task-relevant stimulus). DAF (task-irrelevant stimulus) deviates the subjects from their internal prediction signal generated in parallel with the movements of the subject's speech, and it is considered to be a highly salient stimulus. In particular, auditory stimuli are reported to be more likely to capture attention than visual stimuli (Berti and Schröger, 2001; Boll and Berti, 2009; Leiva et al., 2015). We presumed that speech became nonfluent because attention, which was originally allocated to the reading task, was being captured by the delayed voice (deviant, task-irrelevent stimulus).

One possible explanation for the occurrence of nonfluency under DAF is that the time lag between auditory feedback and internal signal of that efferent motor control that predicts the auditory feedback of the motor command is influenced under DAF (Howell, 2004; Max et al., 2004). The auditory feedback-

based online correction to minimize the internal signal and feedback information may contribute to speech nonfluency (Max et al., 2004). Attention capture could be one of the explanations of the DAF effect since there could be a room for attention system in the feedback control, and we should further elaborate the hypothesis. Though other feedback parameters can change (e.g., frequency-altered auditory feedback), we believe that this attention capture effect that could cause disfluent speech is specific to the delay of own voice. Because the phenomenon of disfluent speech was only found under DAF, and not under any other conditions; e.g., frequently altered auditory feedback (Natke et al., 2001).

Dual Tasks and Attention Capture by Delayed Voice

We also investigated the relationship between auditory attention captured by a delayed voice and nonfluent speech using dual tasks. In a dual task, the subjects are asked to perform a secondary task in addition to the primary task. This causes limited attentional resources to be allocated to each task (Kahneman, 1973). The tactile DAF condition is a cross-modal condition wherein the secondary task (reacting to tactile vibration) is irrelevant to the primary task (reading aloud under DAF). In contrast, the auditory DAF condition is a unimodal condition wherein the secondary task (reacting to auditory tone) is relevant to the primary task (reading aloud under DAF). The correct answer rate for the reaction task was over 80% under all conditions. Hence, in the dual tasks, the subjects continuously allocated attention to not only the sentences but also the target of the secondary task under each modality. This seems apparently inconsistent because, if the DAF increases the auditory attention of the subject and the attention of the auditory feedback involved an auditory task performance, performance parameters such as reaction time or the omission error rate could improve. We assumed that attention allocation under the DAF condition could not be generalized in the context of attention system. Under DAF, their fluency indeed decreased. It was also found that the performance under the DAF condition was poorer than under the NAF condition, and a significant interaction did not occur between the modality (tactile/auditory) and feedback (NAF/DAF) for both the reaction time and omission error rate. Thus, the decrease in performance under DAF is not limited in the auditory task.

Regarding the nonfluency rate in the primary task, there was no significant difference between the conditions under NAF. However, there were significant differences between the conditions under DAF. The nonfluency rate in the dual task with auditory stimulation was significantly higher than that in the single task, while the nonfluency rate in the dual task with tactile stimulation was significantly lower than that in the single task. It is assumed that, in the auditory task under DAF condition, the subjects allocated attention to not only the sentences but also the auditory modality in order to promptly react when the secondary target arrived. The subjects seem to constantly prepare for the detection of and a reaction to the next target when the target was successively presented, and the delayed voice was superimposed within such task sets. A predetermined task set could cause

attention to be captured by the subsequent distracting stimuli (Eimer and Kiss, 2008). Thus, we suggested that the nonfluency rate could increase in the auditor task because of the enhanced attention capture to a delayed voice that was designed for the detection of the auditory target.

This leads to the question of what attentional processes took place under the tactile condition, where nonfluent speech decreased compared to that in the single task. Compared with the auditory condition, the subjects actively allocate attention to the visual-modality sentences (primary task) and the tactile-modality target (secondary and cross-modal task). Attention allocation to auditory stimuli decreased when attention was directed to sequences of tactile stimulus (Marsja et al., 2018). Neuroimaging studies also revealed that the activity of the auditory cortex is attenuated when visual stimuli are presented, while the activity of the visual cortex is attenuated when tactile or auditory stimuli are presented (Laurienti et al., 2002; Merabet et al., 2007). We suggested that, under the tactile condition, the auditory perception of the delayed voice could be inhibited, and attention was captured by the delayed voice to a lesser degree than under the other conditions of DAF. This is because attention was allocated to visual and tactile sensory input. The distracting effect of the delayed voice was attenuated, and nonfluent speech was reduced as a result. Through our findings, we could interpret that nonfluent speech occurs depending on the deviance of the disruption stimulus. That is, task-irrelevant and unimodal stimulus divert attention from the speech, thus resulting in less nonfluency and vice versa.

There are individual differences in the reduction of nonfluent speech. As shown in **Figure 4**, subjects who had a larger DAF effect (i.e., speech disfluency frequently occurs under single DAF condition rather than under single NAF condition) have a tendency to decrease their disfluency under the tactile DAF condition. We assume that disfluent speech resulted in the present study because of inappropriate allocation of attention to delayed voice. The tactile task made the subjects who exhibited a larger DAF effect distract their attention into the tactile (cross-modal) stimulus. Our results suggest that this could increase speech fluency.

In contrast, no significant difference was observed in the nonfluency rate among the different conditions under NAF. The relationship between attention allocation and nonfluent speech was only found under DAF. It is supposed that the subject could allocate sufficient attention to secondary (tactile/auditory) tasks.

The present study showed that attention capture by delayed own voice could result in nonfluent speech. However, it remains unclear if attention capture by delayed own voice is involved in matching or the subsequent movement correction process. A study that used event-related potential (ERP) showed that this matching and movement correction are different processes (Rietdijk et al., 2014). Further studies investigating the impact of attention capture on nonfluent speech using neuroimaging approaches are required. Please note that this study used only DAF at a delay of 200 ms. As our findings regarding nonfluency occurrence by auditory attention capture is at a preliminary level rather than an origin causality, and further studies including conditions such as changing the delay time

or frequency of the auditory feedback are needed to elaborate our findings.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed for this study are not publicly available due to specifications on data availability within the ethics approval. Data are, however, available from the corresponding author upon reasonable request and with the permission of the ethics committee.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tsukuba faculty of human sciences, University of Tsukuba, Japan. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work. OI and DI contributed to the writing of the manuscript. SM contributed to manuscript revision and read and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Appelbaum, L. G., Boehler, C. N., Won, R., Davis, L., and Woldorff, M. G. (2012). Strategic allocation of attention reduces temporally predictable stimulus conflict. *J. Cogn. Neurosci.* 24, 1834–1848. doi: 10.1162/jocn_a_00209

Berti, S., and Schröger, E. (2001). A comparison of auditory and visual distraction effects: behavioral and event-related indices. *Cogn. Brain Res.* 10, 265–273. doi: 10.1016/s0926-6410(00)00044-6

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *J. Speech Disord.* 16, 56–60. doi: 10.1044/jshd.1601.56

Boll, S., and Berti, S. (2009). Distraction of task-relevant information processing by irrelevant changes in auditory, visual, and bimodal stimulus features: a behavioral and event-related potential study. *Psychophysiology* 46, 645–654. doi: 10.1111/j.1469-8986.2009.00803.x

Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain Lang.* 7, 307–319. doi: 10.1016/0093-934x(79)90025-7

Cheal, M. L., and Lyon, D. (1989). Attention effects on form discrimination at different eccentricities. *Q. J. Exp. Psychol. A* 41, 719–746. doi: 10.1080/14640748908402391

Chon, H., Kraft, S. J., Zhang, J., Loucks, T., and Ambrose, N. G. (2012). Individual variability in delayed auditory feedback effects on speech fluency and rate in normally fluent adults. *J. Speech Lang. Hear. Res.* 56, 489–504. doi: 10.1044/1092-4388(2012/11-0303)

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edn. Hillsdale, NJ: Erlbaum.

Eimer, M., and Kiss, M. (2008). Involuntary attentional capture is determined by task set: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 20, 1423–1433. doi: 10.1162/jocn.2008.20099

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *J. Speech Hear. Disord.* 20, 333–346. doi: 10.1044/jshd.2004.333

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf031 93146

Fillenbaum, S. (1963). Impairment in performance with delayed auditory feedback as related to task characteristics. *J. Verb. Learn. Verb. Behav.* 2, 136–141. doi: 10.1016/s0022-5371(63)80077-8

Finke, M., Sandmann, P., Kopp, B., Lenarz, T., and Büchner, A. (2015). Auditory distraction transmitted by a cochlear implant alters allocation of attentional resources. *Front. Neurosci.* 9:68. doi: 10.3389/fnins.2015.00068

Fukawa, T., Yoshioka, H., Ozawa, E., and Yoshida, S. (1988). Difference of susceptibility to delayed auditory feedback between stutterers and nonstutterers. *J. Speech Hear. Res.* 31, 475–479. doi: 10.1044/jshr.3103.475

Hashimoto, Y., and Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an FMRI study. *Hum. Brain Mapp.* 20, 22–28. doi: 10.1002/hbm. 10119

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., et al. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *Neuroreport* 8, 2379–2382. doi: 10.1097/00001756-199707070-00055

Hirst, W., and Kalmar, D. (1987). Characterizing attentional resources. *J. Exp. Psychol. Gen.* 116, 68–81. doi: 10.1037/0096-3445.116.1.68

Howell, P. (2004). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Res.* 1, 31–46.

Ishida, O., Kono, T., and Hisao, K. (2019). Brain activity during disuent speech with delayed auditory feedback: a functional near-infrared spectroscopy study. *J. Spec. Educ. Res.* 7, 69–77. doi: 10.6033/specialeducation.7.69

Kahneman, D. (1973). *Attention and Effort*. New Jersey: Prentice-Hall.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., and Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cogn. Neurosci.* 14, 420–429. doi: 10.1162/089892902317361930

Lee, B. S. (1950). Effects of delayed speech feedback. *J. Acoust. Soc. Am.* 22, 824–826.

Leiva, A., Parmentier, F. B., and Andrés, P. (2015). Distraction by Deviance: comparing the effects of auditory and visual deviant stimuli on auditory and visual target processing. *Exp. Psychol.* 62, 54–65. doi: 10.1027/1618-3169/a000273

MacKay, D. G. (1968). Metamorphosis of a critical interval: age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *J. Acoust. Soc. Am.* 43, 811–821. doi: 10.1121/1.1910900

Marsh, J. E., Sörqvist, P., and Hughes, R. W. (2015). Dynamic cognitive control of irrelevant sound: increased task engagement attenuates semantic auditory distraction. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1462–1474. doi: 10.1037/xhp0000060

Marsja, E., Neely, G., and Ljungberg, J. K. (2018). Investigating deviance distraction and the impact of the modality of the to-be-ignored stimuli. *Exp. Psychol.* 65, 61–70. doi: 10.1027/1618-3169/a000390

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: a theoretical model of stuttering. *Spring* 31, 105–122. doi: 10.1044/cicsd_31_s_105

Melzer, I., Benjuya, N., and Kaplanski, J. (2001). Age-related changes of postural control: effect of cognitive tasks. *Gerontology* 47, 189–194. doi: 10.1159/000052797

Merabet, L. B., Swisher, J. D., McMains, S. A., Halko, M. A., Amedi, A., Pascual-Leone, A., et al. (2007). Combined activation and deactivation of visual cortex during tactile sensory processing. *J. Neurophysiol.* 97, 1633–1641. doi: 10.1152/jn.00806.2006

Miller, E. K., and Buschman, T. J. (2013). Cortical circuits for the control of attention. *Curr. Opin. Neurobiol.* 23, 216–222. doi: 10.1016/j.conb.2012.11.011

Natke, U., Grosser, J., and Kalveram, K. T. (2001). Fluency, fundamental frequency, and speech rate under frequency-shifted auditory feedback

in stuttering and nonstuttering persons. *J. Fluency Disord.* 26, 227–241. doi: 10.1016/s0094-730x(01)00099-7

Nota, Y., Masaki, S., and Ichinose, Y. (2011). "Brain activity during speech under delayed auditory feedback," in *Proceedings of 9th International Seminar on Speech Production*, Montreal, Canada, 329–335.

NPO Research Institute for Japanese Language Education. (1995). *Nihongo Nouryoku Siken Taisaku Koumoku Seiri 1kyu Mondaishu (in Japanese).* Tokyo, Japan: Bonjinsha.

Reeves, A., and Sperling, G. (1986). Attention gating in short-term visual memory. *Psychol. Rev.* 93, 180–206. doi: 10.1037/0033-295x.93.2.180

Rietdijk, W. J., Franken, I. H., and Thurik, A. R. (2014). Internal consistency of event-related potentials associated with cognitive control: N2/P3 and ERN/Pe. *PLoS One* 9:e102672. doi: 10.1371/journal.pone.0102672

Sakai, N., Masuda, S., Shimotomai, T., and Mori, K. (2009). Brain activation in adults who stutter under delayed auditory feedback: an FMRI study. *Int. J. Speech Lang. Pathol.* 11, 2–11. doi: 10.1080/17549500802588161

Salamé, P., and Baddeley, A. (1982). Disruption of short-term memory by unattended speech: implications for the structure of working memory. *J. Verb. Learn. Verb. Behav.* 21, 150–164. doi: 10.1016/s0022-5371(82)90521-7

Schröger, E., and Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport* 9, 3355–3358. doi: 10.1097/00001756-199810260-00003

Siegel, G. M., Fehst, C. A., Garber, S. R., and Pick, H. L. Jr. (1980). Delayed auditory feedback with children. *J. Speech Hear. Res.* 23, 802–813. doi: 10.1044/jshr.2304.802

Siegel, G. M., and Kennard, K. L. (1984). Lombard and sidetone amplification effects in normal and misarticulating children. *J. Speech Hear. Res.* 27, 56–62. doi: 10.1044/jshr.2701.56

Stuart, A., Kalinowski, J., Rastatter, M. P., and Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *J. Acoust. Soc. Am.* 111, 2237–2241. doi: 10.1121/1.1466868

Swan, L., Otani, H., Loubert, P. V., Sheffert, S. M., and Dunbar, G. L. (2004). Improving balance by performing a secondary cognitive task. *Br. J. Psychol.* 95, 31–40. doi: 10.1348/000712604322779442

Takaso, H., Eisner, F., Wise, R. J., and Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. *J. Speech Lang. Hear. Res.* 53, 226–236. doi: 10.1044/1092-4388(2009/09-0009)

Theeuwes, J. (1991). Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Percept. Psychophys.* 49, 83–90. doi: 10.3758/bf03211619

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 799–806. doi: 10.1037/0096-1523.20.4.799

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychol.* 135, 77–99. doi: 10.1016/j.actpsy.2010.02.006

Umeo, K., and Ichinose, Y. (2008). "Effect of auditory attention on utterance in delayed auditory feedback experiments," in *Proceedings of the Spring Meeting of Acoustical Society Japan* Chiba, Japan, 605–606.

Yantis, S., and Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 601–621. doi: 10.1037/0096-1523.10.5.601

Yates, A. J. (1963). Delayed auditory feedback. *Psychol. Bull.* 60, 213–232. doi: 10.1037/h0044155

Zanini, S., Clarici, A., Fabbro, F., and Bava, A. (1999). Speaking speed effects on delayed auditory feedback disruption of speech fluency. *Percept. Mot. Skills* 89, 1095–1109. doi: 10.2466/pms.1999.89.3f.1095