

Published in final edited form as:

J Speech Lang Hear Res. 2010 April ; 53(2): 226–236. doi:10.1044/1092-4388(2009/09-0009).

The effect of delayed auditory feedback on activity in the temporal lobe while speaking: A PET study

Hideki Takaso^{#1}, Frank Eisner^{#2}, Richard JS Wise³, and Sophie K Scott²

¹Dept of Psychology, UCL. Gower Street, London WC1E 6BT

²Institute of Cognitive Neuroscience, UCL, 17 Queen Square London WC1N 3AR

³MRC Clinical Sciences Centre, Hammersmith Hospital, Ducane Road, London W12 0NN

These authors contributed equally to this work.

Abstract

Purpose—Delayed auditory feedback is a technique that can improve fluency in stutterers, while disrupting fluency in many non-stuttering individuals. The aim of this study was to determine the neural basis for the detection of and compensation for such a delay, and the effects of increases in the delay duration.

Method—Positron emission tomography (PET) was used to image regional cerebral blood flow changes, an index of neural activity, and assessed the influence of increasing amounts of delay.

Results—Delayed auditory feedback led to increased activation in the bilateral superior temporal lobes, extending into posterior-medial auditory areas. Similar peaks in the temporal lobe were sensitive to increases in the amount of delay. A single peak in the temporal parietal junction responded to the amount of delay but not to the presence of a delay (relative to no delay).

Conclusions—This study permitted distinctions to be made between the neural response to hearing one's voice at a delay, and the neural activity that correlates with this delay. Notably all the peaks showed some influence of the amount of delay. This result confirms a role for the posterior, sensori-motor 'how' system in the production of speech under conditions of delayed auditory feedback.

Introduction

Speech production is continuously regulated with respect to the current acoustic environment. In noisy situations, talkers unconsciously raise the level of their voices (Lombard, 1911), and if the apparent volume of their voices is raised, talkers unconsciously lower the level of their voices (and vice versa) (Fletcher, Raff & Parmley, 1918). This suggests that, in addition to voluntary modulations, talkers continuously relate the sound of their voices to their acoustic context in a flexible, obligatory, manner. There is debate in the literature as to the extent that this requires explicit monitoring during speech production (Howell, 2004).

At a neural level, there is now considerable evidence that the neural response to a talker's own voice during speech production differs from the neural response to the voices of others. Instead, regions in the secondary auditory cortex, which normally show a robust response to speech sounds, are suppressed in response to self-produced speech (Houde et al, 2002; Wise et al, 1999). Work with non-human primates indicates that these inhibitory effects arise before vocalization physically begins (Eliades and Wang, 2003; 2005). This suppression may be part of a mechanism to ensure that sensory properties of self-initiated actions are processed differently from actions initiated by other sources (e.g. Blakemore Wolpert and Frith, 1998). Alternatively, feed forward projections of actions are proposed as mechanisms for the monitoring of self-initiated actions, via the comparison of these output representations with the sensory input arising from the actions (feedback). Work with non-human primates indicates that many of the neurons which show suppression during vocalization increase in activation when the sound of the voice is distorted (Eliades and Wang, 2008), which may indicate that these neurons are involved in feed forward, monitoring mechanisms. The interaction of feedback and feed forward connections in speech production is one of the features of the DIVA model of speech production (Guenther, Ghosh and Tourville, 2006), which identifies feedforward projections from premotor cortex to somatosensory and auditory cortex, and auditory input about articulations being processed in auditory cortex. In contrast, work in the nonhuman primate literature has identified both somatosensory and auditory responses in the caudo-medial belt area of auditory cortex (Fu et al, 2003, Smiley et al, 2007), and this has been identified in humans as a candidate region for integrating somatosensory and auditory information during speech production (Dhanjal et al, 2008).

In humans, distortions of speech in pitch, spectrum or time affect a talker's speech. Alterations of timing feedback are associated with increasing levels of disfluency in many talkers (Black, 1951; Fukawa et al., 1988; Howell, 1990; Langova et al., 1970; Lee, 1950, 1951; Mackay, 1968; Siegel et al., 1982; Stager et al., 1997; Stager and Ludlow, 1993.). In delayed auditory feedback (DAF) the disfluency is related to the amount of the delay: as the delay increases the speech rate can slow, there can be more vowel prolongations and blocks, (complete cessation of speech output), stuttering can occur, there can be changes in speech prosody, errors can be made on the accuracy of articulation, and phonemes can be produced in the wrong positions (e.g. Fairbanks, 1955; Black, 1951; Ham et al., 1984; Lee, 1950; Siegel et al., 1982; Stager and Ludlow, 1993). These factors increase from short delays up to a maximal disfluency at around 200 ms (Black, 1951, Stuart et al, 2002). The degree to which speakers are affected by DAF varies somewhat across different talkers: many speakers can become extremely disfluent and have difficulty talking at all under DAF, while others can continue to speak under DAF (though their speech production may still be changed - they may speak more slowly, flatten their pitch, and make errors). For speakers with preexisting, developmental fluency problems, this pattern is reversed: individuals with developmental stuttering often find their speech production becomes more fluent under conditions of altered speech feedback, including DAF, and pitch-shifted auditory feedback (or even speaking in noisy environments) (Bloodstein and Ratner, 2008).

The influence of changes to the timing, pitch or timbre of self-produced speech on fluency is likely due to several factors. One element, as evidenced by the role of the length of delay on

fluency, is the timing of speech production: speech is affected by delayed feedback starts at around the level that the delay can be detected (around 40 ms) (Natke, 1999, Jänke, 1991), and increase as the delay increases, to a maximal effect around 200 ms (Black, 1951). A 200 ms delay is around the duration of a syllable and this length of delay may thus be disruptive, as talkers are starting to utter one syllable when hearing the onset of the last (Black, 1951). This may explain the lack of preemptive errors (e.g. a biece of bread) under delay conditions, and the increase in preservative errors (e.g. hypodermic nerdle) (Hashimoto and Sakai, 2003). Furthermore, the effects of delayed auditory feedback are not specific to speech, and DAF also affects the production of rhythmic tapping, and of musical performance (e.g. Pfordesher and Benitez, 2007). This suggests that the effects of DAF are not due to lexical features of what is being produced interacting with what is being heard.

Functional imaging studies of altered auditory feedback during speech production have shown neural activity in the posterior superior temporal gyrus, inferior parietal lobes and frontal lobes for DAF conditions (Hashimoto and Sakai, 2003) and altered pitch feedback (Toyomura et al, 2007), relative to normal speech conditions. Furthermore, these same posterior fields are activated, along with other brain areas, when talkers speak while spectral properties of their speech are altered in real time (Tourville, Reilley and Guenther, 2008), or when attempts are made to mask self-produced speech during overt production (Christoffels et al, 2007). Some of these papers linked the temporal lobe activations expressly to issues of auditory feedback mechanisms in speech production, in addition to compensatory mechanisms in the right frontal lobe (Tourville et al, 2008).

A role for posterior auditory fields in speech production, when speaking under delayed or distorted auditory feedback, has implications for the functional neuroanatomy of speech perception and production. Neural perspectives on the perceptual processing of speech posit that there are at least two different routes along which heard speech is processed, an anterior 'what' processing stream running forwards down the temporal lobe, along the lateral STG and STS, and a posterior 'how' processing stream, running posterior and medial to primary auditory cortex (Scott and Johnsrude, 2003). While the comprehension of spoken language has been linked to the anterior stream of processing (Scott et al, 2000; 2006), the posterior 'how' pathway is activated by both speech perception and production (Okada and Hickok, 2006) and by articulation, even if the articulation is silent (Wise et al, 2001, Hickok et al, 2000). This 'how' pathway has been identified as central to the perceptual control of speech production (Dhanjal et al, 2008, Rauschecker and Scott, 2009). Posterior auditory fields have been specifically linked to the representations of 'do-able' articulations (Warren Wise and Warren, 2005), which map sensori-motor transformations involved in both speech perception and production processes. The posterior-medial auditory fields have also been argued to be not specific to speech in their responses, as they are involved in the rehearsal of both speech and non-speech acoustic information (Hickok et al, 2003). These same posterior-medial auditory fields have been reported to be activated in several functional imaging studies of altered auditory feedback (Hashimoto and Sakai, 2003; Tourville, Reilley and Guenther, 2008). Potentially, posterior auditory areas are recruited to compute and overcome timing difficulties or correct errors in speech production, whether these errors and timing differences are consciously detected or not (Tourville et al, 2008). The aim of the

current study is to vary the *amount* of delay in DAF, and to identify whether the neural systems, which respond to the delayed speech can be dissociated from those that are sensitive to the amount of delay (and which might be showing a purely acoustic response to the extra vocal information in the delay conditions).

Methods

PET neuroimaging was used to image changes in regional cerebral blood flow (rCBF) across conditions, and four different delay conditions were employed: 0 ms (no DAF), 50 ms, 125 ms and 200 ms. PET scanning was used because it is relatively insensitive to the movements associated with overt speech production (compared to fMRI), and the subjects can be scanned while they speak continuously (Blank et al, 2002).

Subjects who could continue to speak under conditions of DAF were tested. These subjects did not 'block', such that there was no complete cessations of speech production. This allowed the delineation of neural systems sensitive to the delay and the amount of delay, without any confounds of the neural responses to a total cessation of speech production. Even though the subjects selected could continue to talk under DAF, their speech production will not have remained entirely normal, and it was likely that there were some changes in errors, rate and pitch profile. Comparing the neural basis of DAF in talkers who can continue to speak under DAF will help us to interpret the activations seen in future studies on speakers who find DAF much more disruptive to their speech production. The use of speakers who became markedly disfluent under DAF would be problematic for a PET study. PET studies are limited in the number of scans that can be run for any one subject, and it would not be possible to repeat a scan lost because a subjects stopped speaking, or to collect enough scans in each condition such that scans which contained speech blocks could be excluded from the analysis.

PET scanning

Eight right-handed native English-speaking volunteers (seven men, one woman), none of whom reported any hearing problems, were recruited and scanned. None of these subjects had any developmental fluency problem. These subjects were pre-tested to ensure that they could continue to speak under DAF - that their speech production was not interrupted by blocking or long prolongations. The mean age was 42, with a range of 35-53. Each participant gave informed consent prior to participation in the study, which was approved by the Research Ethics Committee of Imperial College School of Medicine/Hammersmith, Queen Charlotte's & Chelsea & Acton Hospitals. Permission to administer radioisotopes was given by the Department of Health (UK).

PET scanning was performed with a Siemens HR++ (966) PET scanner operated in high-sensitivity 3D mode. Sixteen scans were performed on each subject, using the oxygen-15-labelled water bolus technique. All subjects were scanned whilst lying supine.

There were four different scanning conditions, which were presented in a pseudo-random order, to ensure that all the presentations of each condition did not occur consecutively. To do this, random sequences were generated, and sequences in which the conditions were not

consecutive were selected. A different order was used for each subject. The subjects were always asked to read aloud into a microphone. The microphone was connected to a delay line, which introduced a delay into the signal, which was then presented to the subjects over headphones. The level of this feedback was set at a comfortable level for each subject when they spoke into the microphone with no delay added. The amounts of delay were: 0 ms (no delay at all), 50 ms, 125 ms and 200 ms. These levels were selected since 50 ms is a barely noticeable level of delay, 125 ms is an intermediate level of delay and 200 ms is a maximally disruptive level of delay for speakers whose speech production is affected by DAF (Black, 1951).

The stimuli were pages of text from a children's book (*The Borrowers*). A text aimed at children was chosen so that the items to be read would be relatively simple in syntax and semantic content. These were photocopied onto A3 sheets so that the font size was large enough to be readable, and attached to a boom in front of the PET scanner. This was lowered so that the subjects could read the text. The pages were counterbalanced in order across the scans. The subjects were instructed to read the passage aloud, and if they made an error, to continue with as brief delay as possible. The subjects were instructed to read aloud 15 seconds before the scanning commenced, and they were not told what auditory feedback condition they were in. The subjects were told to keep reading until they were asked to stop at the end of each scan (approximately 1.5 minutes after starting to read aloud). After every scan, the subjects were asked to rate (out of 10) how accurate they thought they had been in their articulations (where 10 would be very accurate), how difficult they had found it to read aloud (where 10 was very difficult), and how well they thought they had timed their reading – did they feel they read at the right speed? (where 10 would be very well timed). These measures were used as a subjective index of the perceived effects of the DAF on the task of speaking aloud.

Analysis

The images were analysed using statistical parametric mapping (SPM99, Wellcome Department of Cognitive Neurology, <http://www.fil.ion.ucl.ac.uk/spm>), which allowed manipulation and statistical analysis of the grouped data. All scans from each subject were realigned to eliminate head movements between scans and normalized into a standard stereotactic space (the Montreal Neurological Institute [MNI] template was used, which is constructed from anatomical MRI scans obtained on 305 normal subjects). Images were then smoothed using an isotropic 12 mm, full width half-maximum, Gaussian kernel, to allow for variation in gyral anatomy and to improve the signal-to-noise ratio. Specific effects were investigated, voxel-by-voxel, using appropriate contrasts to create statistical parametric maps of the t statistic, which were subsequently transformed into Z scores. The analysis included a blocked AnCova with global counts as confound to remove the effect of global changes in perfusion across scans. The functional imaging data were then analysed by generating contrasts to test the specific hypotheses at a whole brain level of analysis (Friston et al, 2007). Results which were significant at a level of $t < 3.85$, i.e. $p < 0.0001$ (uncorrected for multiple comparisons), and with cluster size of (at least) 40 contiguous voxels are reported.

The first contrast, delayed auditory feedback>no delay ([50, 125 and 200 ms]>0 ms) identified brain areas more activated by the three DAF conditions than by the no-DAF condition. This contrast was entered as a planned comparison, in which negative signs are attached to the condition(s) on one side of the comparison, and positive signs are attached to the conditions(s) on the other side of the comparison. The positive and negative conditions are weighted such that they sum to zero. To identify brain areas more active for the three delayed conditions (50, 125 and 200 ms) over no-delay condition (0 ms), the 0 ms condition was weighted as -3 and the three delay conditions each weighted as 1.

The second planned comparison identified brain areas that increased in activation as the amount of delay increased, i.e. brain areas activated more by 200 ms than 125 ms, and more by 125 ms than 50 ms, and more by 50 ms than 0 ms (0 ms <50 ms<125 ms<200 ms). This planned comparison was modeled with 0 ms weighted to -2, 50 ms weighted to -1, 125 ms weighted to 1 and 200 ms weighted to 2.

The third planned comparison modeled for brain areas sensitive to the increase in delay, but which were not activated in the DAF>no DAF contrast. This was modeled for by running the second contrast, identifying brain areas activated by increases in the delay across conditions, and masking exclusively for brain regions activated by the first, any delay over no delay, contrast. This was an attempt to identify brain regions *not* activated by the delay over no delay contrast (which could be showing a purely acoustic response to the 'extra' delayed speech signal), but which also showed a sensitivity to increases in delay.

Negative versions of each contrast were run, with inverted weightings for the conditions. No significant activation was found for either of these contrasts. Finally, all three sets of subjective data were used as correlations in the functional imaging analysis, to identify brain areas which correlated with difficulty, or accuracy of timing or articulation. None of these correlations were significant.

Results

Behavioural

The subjects all reported finding that as the amount of delay increased, that the perceived difficulty increased, and that the perceived accuracy of articulation and of timing decreased with delay (Figure 1). The behavioural data for two subjects were not available so these values are based on six subjects. These effects were tested with three separate one-way repeated measures ANOVAs. There was a significant effect of amount of delay on each measure: articulation ($F(3,15) = 8.33, p=.002$, partial $\eta^2 = 0.626$), difficulty ($F(3, 15) = 7.9$, $p=.002$, partial $\eta^2 = 0.613$); timing ($F(3, 15) = 12.94, p<.0001$), partial $\eta^2 = 0.721$).

Functional imaging

Contrast 1 – Delayed auditory feedback>no delay: Table 1 shows the peak activations associated with the contrast (50, 125 and 200 ms)>0 ms. This contrast gives the peak activations seen when there is any delayed auditory feedback, irrespective of the amount of delay. This contrast revealed bilateral activation in the left and right superior temporal gyri.

The main peaks in the left superior temporal gyrus is shown in figure 2, with plots of the neural response (as percentage change in regional cerebral blood flow, rCBF) in the main peak. Some activation is spreading across the Sylvain fissure into ventral motor and premotor cortex, but this does not become a separate peak.

Contrast 2 – increasing the length of delay: The second contrast modeled for brain areas that increased in activation with the amount of delay, i.e. $0 \text{ ms} < 50 \text{ ms} < 125 \text{ ms} < 200 \text{ ms}$. This contrast revealed a range of peaks in the left and right STG, described in Table 2. Peak activations in the left STG are shown in Figure 3, with plots of the relationship between the amount of delay and the neural response (shown as percentage change in rCBF). Again, activation is spreading into ventral motor and premotor cortex, but this does not become a separate peak.

Contrast 3 – effects of increasing delay, masking for existence of delay: The final planned comparison looked for brain regions which increased in activity with increasing amount of delay ($0 \text{ ms} < 50 \text{ ms} < 125 \text{ ms} < 200 \text{ ms}$) while masking exclusively for regions which show any response to the DAF>none contrast ($[50, 125 \text{ and } 200 \text{ ms}] > 0 \text{ ms}$). This is a conservative measure, designed to see if the sensitivity to the delay has a neural correlate, which fully dissociates from the acoustic response. This contrast revealed one small peak of activation in the left posterior temporal parietal junction, shown in Figure 4 and Table 3, with the neural response shown as percentage change in rCBF.

Discussion

Bilateral superior temporal responses, extending into sensory-motor cortex, were observed to delayed auditory feedback, relative to the activation seen to normal speech production, at a whole brain level of analysis. This confirms a role for posterior superior temporal lobe areas in the neural response to delayed or distorted auditory feedback (Hashimoto and Sakai 2003; Tourville, Reilly and Guenther 2008). When neural responses to the amount of delay were modeled, this activation was also mainly restricted to the bilateral posterior superior temporal lobes (extending into sensory-motor cortex) and the temporal-parietal junction. The peaks of activation were very close to those for the first contrast, and also included one more posterior peak. Indeed, there were no peak activations in any contrast which were sensitive to the existence of delayed auditory feedback, but were insensitive to the amount of delay: where there were neural responses to delayed auditory feedback, there was always some sensitivity to the amount of delay. The opposite pattern was however seen: in a left temporal-parietal junction, a small area was found which responded to the amount of delay but not to the basic DAF>no DAF contrast.

These results demonstrate that these posterior auditory responses to hearing one's voice at a delay while speaking (Hashimoto and Sakai, 2003) are involved in processing the amount of the delay and consequent compensation (Tourville et al, 2008). This result is consistent with the argument that there are comparisons between feed-forward and feedback signals in superior temporal lobe areas, with the possibility that these comparisons are important in error detection (Guenther et al, 2006, Tourville et al, 2008). The cortical fields which had a maximal response to delayed auditory feedback were not found in the more anterior

temporal lobe areas associated with intelligibility in speech perception (Scott et al, 2000; 2006). This may be consistent with the finding that the effects of DAF are not restricted to speech production, but also influence the playing of music (Pfordresher and Benitez, 2007).

In the current design, it is not possible to distinguish between neural responses important in perceiving and computing the response to (or compensating for) the delay. However, the pattern of activations suggest that the posterior temporal lobes are associated with both detection of and compensation for delays, and that this may extend into ventral motor regions.

A role for posterior temporal lobes in compensatory mechanisms when speaking under DAF is consistent with an important role for this area in speech production. Anatomically, there is evidence that dorsal and medial auditory areas of the superior temporal plane are part of a 'how' stream of processing, a route for linking sensori-motor activation in speech (Scott and Johnsrude, 2003; Warren Wise and Warren, 2005). Activation is seen in posterior medial auditory cortex during articulation, even if that articulation is silent (Wise et al, 2001). In the current study, a single peak in the left temporal-parietal junction which showed a response to the amount of delay but not to the main effect of delay (over no delay), suggesting that this region is not processing the acoustic information in the voice, but is directly involved in controlling for speech production used these conditions of delayed feedback. This peak lay within 9mm of the peak reported for articulation in Wise et al (2001) in the x plane, 2mm in the y plane and 4mm in the z plane, once transformed into the same Talairach co-ordinates. The more extensive activations that are sensitive to both the presence of delayed auditory feedback and the amount of delay also encompass this region. These posterior-medial auditory fields have recently been shown to be activated both by speech production and silent, simple movement of the articulators, e.g. opening and closing the jaw (Dhanjal et al, 2008), and it has been described as an auditory area which is central to the perceptual control of speech production.

As noted previously, this posterior auditory activation has been linked to error monitoring processes in speech production. In Tourville and colleagues' paper on spectral distortion of the first formant (F1) (Tourville et al, 2008), the amount of change of F1 was below the level of awareness for the subjects, although participants changed their articulations to compensate for the distortion over the course of the production of a single long word. Since these posterior auditory areas are also activated by silent articulation (Wise et al, 2001), it is possible that error monitoring is not the sole function of this region, or that what is monitored includes motoric and somatosensory information, as well as acoustic information (Dhanjal et al, 2008). In the current study the subjects read aloud for relatively long periods of time, any disruption was present (or not) from the outset, so the activations seen here cannot be linked precisely to error monitoring – instead they may reflect processes associated with the monitoring of delays and/or implementing compensatory strategies to overcome rhythmic interference. These processes need not be speech specific (e.g. Hickok et al, 2003, Howell 2004, Pfordresher and Benitez, 2007).

There are limitations to the current study. There were no significant neural changes associated with the behavioural measures: while this may reflect the relatively small

variation in the subjective scores across conditions, further studies with more scans per condition (e.g. using fMRI) should be able to determine some neural correlates of these scores. This kind of design would also be able to specifically code for changes in speech production (e.g. errors, changes in rate or speech melody) that varied with the amount of delay. Running more scans per delay condition would also allow the investigation of time dependent changes in neural activity, to identify any adaptation to the delayed auditory feedback. Recordings of the speech of most of the subjects were not available due to technical problems, which means that we cannot specifically link the neural changes to changes in the rate or the pitch of the speech produced in the different conditions. Future studies using sparse fMRI and DAF (e.g. Dhanjal et al, 2008) and recordings of the speech produced will be well placed to investigate these relationships since many more trials can be run in each condition.

Only activity in the posterior temporal lobes, extending into ventral sensori-motor cortex, and the inferior parietal lobes was seen in the current study. The previous study of delayed auditory feedback, using fMRI (Hashimoto and Sakai 2003) reported activation in frontal cortical fields, both ventral prefrontal bilaterally and more dorsal prefrontal areas on the right, in addition to the bilateral superior temporal/inferior parietal activation. Furthermore, the study of distorted spectral feedback in speech production (Tourville et al, 2008) also found right prefrontal cortical areas, which were significant when corrected for false discovery rates. These prefrontal areas did not survive correction for multiple comparisons in Hashimoto and Sakai (2003), and prefrontal areas were not found in the current study, even when a low threshold (not corrected for multiple comparisons) was used. A candidate reason for this discrepancy is statistical power: Hashimoto and Sakai (2003) had 32 trials for their delayed speech condition over the whole scanning session, and 64 trials for the normal speech condition. Likewise, the subjects in Tourville et al (2008) performed 3 or 4 runs of the experimental sequence, in which 16 of 64 spoken words were spectrally distorted, leading to a total of between 48 or 64 distorted feedback trials, and between 144 and 192 normal production trials. In the current study, each subject was run in 12 DAF trials and 4 normal speech production trials. Since PET scans are limited in the total number of scans by the dose administered, this problem cannot be easily addressed. However, the strong similarity of the neural responses in posterior-medial temporal and inferior parietal areas by DAF in the two modalities (PET and fMRI) is strong evidence for the involvement of these regions in the detection of and compensation for delay.

A final issue is that as the DAF increased, it is likely that the subjects' speech production changed, for example in terms of rate and pitch profile, even if they were able to continue speaking. It is thus possible that the neural activity seen was influenced by the subjects hearing this altered speech. This possibility could be directly tested by conducting a speech perception study in which listeners were presented with speech from talkers recorded speaking with and without DAF. Involvement of the same brain areas seen in this production study in the perception of speech produced under DAF would indicate that this perceptual change was driving the activation seen in posterior auditory fields.

In conclusion, therefore, this study has provided further evidence that posterior auditory areas are sensitive to the presence of distortion of feedback during speech during speech

production. This basic finding is elaborated to show that these same posterior auditory fields are also sensitive to the *amount* of delay in delayed auditory feedback, which strongly implicates that these regions are involved in both computing of the delay, and in compensation for the consequences of this delay. In terms of anatomy, this is further evidence that posterior auditory areas are involved in a ‘how’ pathway, important in the sensori-motor co-ordination of speech. Further studies will investigate how this cortical profile is modified in speakers who experience greater disfluency under DAF, and whether this approach can be helpfully applied to studies of disfluency, for example developmental stuttering.

Acknowledgments

FE and SKS are supported by a SRF award to SKS from the Wellcome Trust (WT074414MA)

References

- Black JW. The effect of delayed sidetone upon vocal rate and intensity. *The Journal of Speech Disorders*. 1951; 16:56–60. [PubMed: 14804702]
- Blakemore SJ, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. *Nature Neuroscience*. 1998; 1:635–40.
- Bloodstein, O.; Ratner, N. O. A handbook on stuttering. 6th edition. Delmar Learning; Canada: 2008.
- Dhanjal NS, Handunnetthi L, Patel MC, Wise RJ. Perceptual systems controlling speech production. *Journal of Neuroscience*. 2008; 28:9969–75. [PubMed: 18829954]
- Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*. 2003; 89:2194–207. [PubMed: 12612021]
- Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex*. 2005; 15:1510–23. [PubMed: 15689521]
- Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 2008; 453:1102–6. [PubMed: 18454135]
- Fletcher, H.; Raff, GM.; Parmley, F. Study of the effects of different sidetones in the telephone set. Western Electrical Company; 1918. Report no. 19412, Case no. 120622
- Friston, KJ.; Ashburner, JT.; Kiebel, SJ.; Nichols, TE.; Penny, WD. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier; New York: 2007.
- Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE. Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*. 2003; 23:7510–5. [PubMed: 12930789]
- Fukawa T, Yoshioka H, Ozawa E, Yoshida S. Difference of susceptibility to delayed auditory feedback between stutterers and nonstutterers. *Journal of Speech and Hearing Research*. 1988; 31:475–479. [PubMed: 3172765]
- Ham R, Fucci D, Cantrell J, Harris D. Residual effect of delayed auditory feedback on normal speaking rate and fluency. *Perceptual and Motor Skills*. 1984; 59:61–62. [PubMed: 6493955]
- Howell P. Changes in voice level caused by several forms of altered auditory feedback in fluent speakers and stutterers. *Language and Speech*. 1990; 33:325–338. [PubMed: 2133911]
- Guenther FH, Ghoshm SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*. 2006; 96:280–301. [PubMed: 16040108]
- Hashimoto Y, Sakai KL. Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Human Brain Mapping*. 2003; 20:22–8. [PubMed: 12953303]
- Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*. 2003; 15:673–82. [PubMed: 12965041]

- Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil KA. Functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neuroscience Letters*. 2000; 287:156–60. [PubMed: 10854735]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*. 2002; 14:1125–38. [PubMed: 12495520]
- Howell P. Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stuttering. *Stammering Research*. 2004; 1(1):31–46. [PubMed: 18259594]
- Jäncke, L.; Peters, HFM.; Hulstijn, A.; Starkweather, CW. *Speech Motor Control and Stuttering*. Elsevier Scientific Publishers; Amsterdam: 1991. The 'audio-phonatoric coupling' in stuttering and nonstuttering adults: Experimental contributions; p. 171-180.
- Langova J, Moravsek M, Novak A, Petrik M. Experimental interference with auditory feedback. *Folia Phoniatrica*. 1970; 22:191–196. [PubMed: 5495788]
- Lee BS. Effects of delayed speech feedback. *Journal of the Acoustical Society of America*. 1950; 22:824–826.
- Lee BS. Artificial stutterer. *The Journal of Speech Disorders*. 1951; 16:63–65.
- Lombard E. Le signe de l'elevation de la voix. *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx*. 1911; 37:101–119.
- Mackay D. Metamorphosis of critical interval: Age linked changes in the delay in auditory feedback that produces maximal disruptions of speech. *Journal of the Acoustical Society of America*. 1968; 43:811–821. [PubMed: 5645830]
- Natke U. Die Kontrolle der Phonationsdauer bei stotternden und nichtstotternden Personen: Einfluss der Rückmeldelaustärke und Adaptation. *Sprache - Stimme - Gehör*. 1999; 23:198–205.
- Okada K, Hickok G. Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*. 2006; 98:112–7. [PubMed: 16716388]
- Pfordresher PQ, Benitez B. Temporal coordination between actions and sound during sequence production. *Human Movement Science*. 2007; 26:742–56. [PubMed: 17764772]
- Rauschecker JP, Scott SK. Maps and Streams in the Auditory Cortex: How work in non-human primates has contributed to our understanding of human speech processing. *Nature Neuroscience*. 2009; 12(6):718–724.
- Scott SK, Blank SC, Rosen S, Wise RJS. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*. 2000; 123:2400–2406. [PubMed: 11099443]
- Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*. 2003; 26:100–7. [PubMed: 12536133]
- Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *Journal of the Acoustical Society of America*. 2006; 120:1075–83. [PubMed: 16938993]
- Siegel GM, Schork EJ Jr, Pick HL Jr, Garber SR. Parameters of auditory feedback. *Journal of Speech and Hearing Research*. 1982; 25:473–475. [PubMed: 7176623]
- Stager SV, Ludlow CL. Speech production changes under fluency-evoking conditions in nonstuttering speakers. *Journal of Speech and Hearing Research*. 1993; 36:245–253. [PubMed: 8487517]
- Stager SV, Denman DW, Ludlow CL. Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions. *Journal of Speech, Language, and Hearing Research*. 1997; 40:832–847.
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Comparative Neurology*. 2007; 502:894–923. [PubMed: 17447261]
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. 2008; 39:1429–43. [PubMed: 18035557]
- Toyomura A, Koyama S, Miyamaoto T, Terao A, Omori T, Murohashi H, Kuriki S. Neural correlates of auditory feedback control in human. *Neuroscience*. 2007; 146:499–503. [PubMed: 17395381]

- Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences*. 2005; 28:636–43. [PubMed: 16216346]
- Wise RJS, Scott SK, Blank SC, Mummery CJ, Warburton E. Identifying separate neural sub-systems within Wernickes area. *Brain*. 2001; 124:83–95. [PubMed: 11133789]
- Wise RJS, Greene J, Büchel C, Scott SK. Brain systems for word perception and articulation. *The Lancet*. 1999; 353:1057–1061.

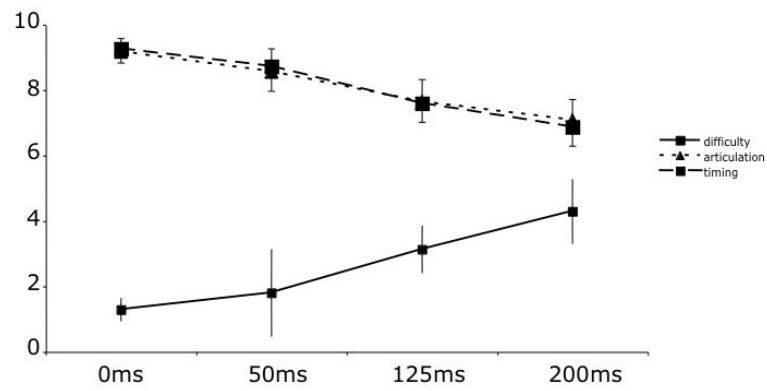


Figure 1. The mean ratings for difficulty of speaking (out of 10, where 10 is very difficult), timing of speech (where 10 is accurate timing) and accuracy of articulation (where 10 is accurate articulation).

The participants rated each score after every scan. Error bars show standard error of the mean.

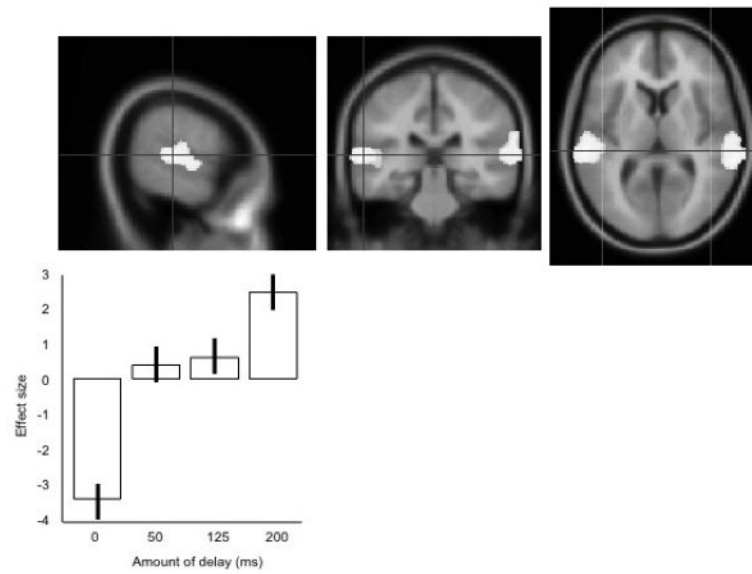


Figure 2. The main peak activation for the presence of any delay of vocal feedback, over the no delay condition ([200 ms, 125 ms 50 ms]>0 ms).

The upper panels show the data projected onto an averaged structural MRI image, showing (from left to right) sagittal, coronal and transverse sections. The lower panel shows the mean effect sizes at this peak location, across the four conditions, centred around zero. Error bars show the standard errors of the parameter estimates.

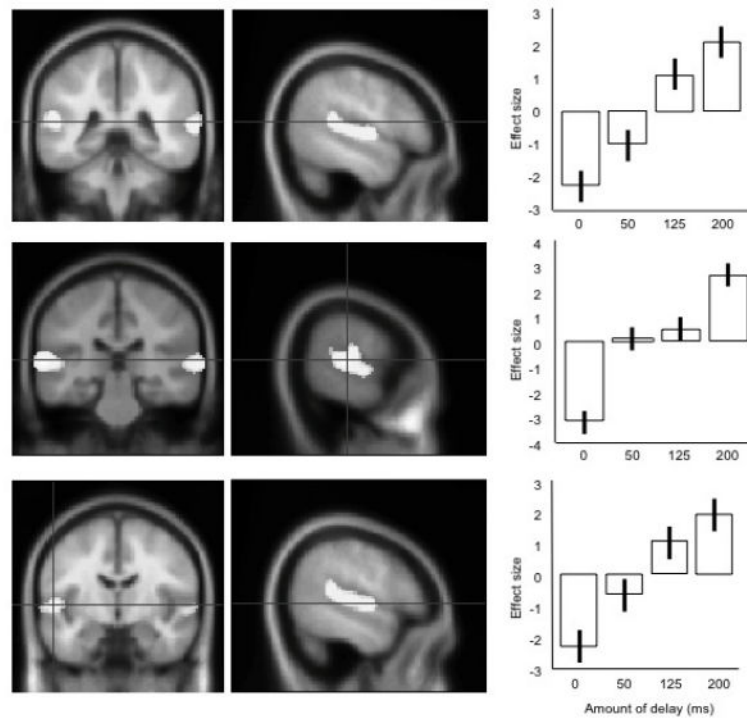


Figure 3. The three peaks of activation for the contrast identifying rCBF increases with amount of delay (200 ms>125 ms>50 ms>0 ms).

The left and middle panels show the peaks projected on mean structural MRI images on coronal and sagittal sections: the right-most panels show the average effect sizes for each condition at these peaks, centred around zero. Moving from top to bottom, the peaks are shown in a posterior to anterior direction. Error bars show the standard errors of the parameter estimates.

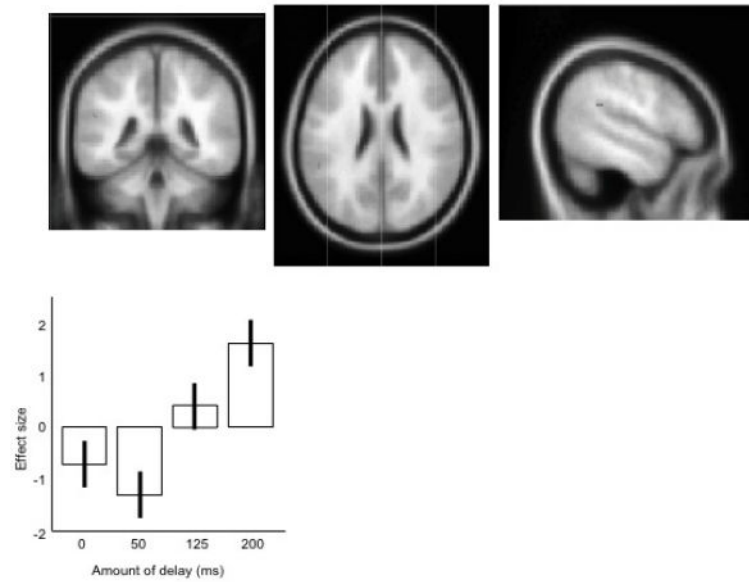


Figure 4. The peak activation for the contrast (200 ms > 125 ms > 50 ms > 0 ms), masking exclusively for the contrast ([200 ms, 125 ms 50 ms] > 0 ms). The upper panels show the data projected onto an averaged structural MRI image, showing (from left to right) coronal, transverse, and sagittal sections. The lower panel shows the mean effect sizes at this peak location, across the four conditions, centred around zero. Error bars show the standard errors of the parameter estimates.

Table 1
The peaks of activation for the contrast of any delay in the speech feedback over the no delay condition.

The main peaks are shown in bold. The x, y, z co-ordinates give the spatial location of each peak in mm in MNI stereotactic space.

Region	P(corrected)	t	x	y	z
Left superior temporal gyrus	0.000	6.88	-60	-30	8
	0.005	5.53	-52	-12	0
Right superior temporal gyrus	0.000	6.77	64	-38	10
	0.012	5.33	54	-20	6

Table 2
Peaks of activation for the contrast identifying increases of activation with increased amount of delay.

The main peaks are shown in bold. The x, y, z co-ordinates give the spatial location of each peak in mm in MNI stereotactic space.

Region	P(corrected)	t	x	y	z
Left superior temporal gyrus	0.000	7.18	-62	-28	8
	0.002	5.82	-54	-8	2
	0.002	5.74	-52	-40	12
Right superior temporal gyrus	0.000	7.15	68	-36	10
	0.000	6.12	54	-20	4

Table 3

Peak of activation for the contrast identifying increases of activation for increasing amounts of delay, masking exclusively for any areas which are activated by the any delay>no delay contrast. The x, y, z co-ordinates give the spatial location of each peak in mm in MNI stereotactic space.

Region	P(uncorrected)	t	x	y	z
Left posterior temporal parietal junction	0.000	3.59	-52	-40	24