
The Effect of Delayed Auditory Feedback on Activity in the Temporal Lobe While Speaking: A Positron Emission Tomography Study

Hideki Takaso

Frank Eisner

University College London (UCL),
United Kingdom

Richard J. S. Wise

Hammersmith Hospital, London

Sophie K. Scott

UCL

Purpose: Delayed auditory feedback is a technique that can improve fluency in stutterers, while disrupting fluency in many nonstuttering 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 was used to image regional cerebral blood flow changes, an index of neural activity, and to assess 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, sensorimotor "how" system in the production of speech under conditions of delayed auditory feedback.

KEY WORDS: delayed auditory information, speech production, positron emission tomography

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 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, Nagarajan, Sekihara, & Merzenich, 2002; Wise, Greene, Büchel, & Scott, 1999). Work

with nonhuman primates indicates that these inhibitory effects arise before vocalization physically begins (Eliades & 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, & 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 nonhuman primates indicates that many of the neurons that show suppression during vocalization increase in activation when the sound of the voice is distorted (Eliades & 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 (Directions Into Velocities of Articulators) model of speech production (Guenther, Ghosh, & Tourville, 2006), which identifies feed-forward 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, Handunnetthi, Patel, & Wise, 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, Yoshioka, Ozawa, & Yoshida, 1988; Howell, 1990; Langova, Moravak, Novak, & Petrik, 1970; Lee, 1950, 1951; Mackay, 1968; Siegel, Schork, Pick, & Garber, 1982; Stager, Denman, & Ludlow, 1997; Stager & 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., Black, 1951; Fairbanks, 1955; Ham, Fucci, Cantrell, & Harris, 1984; Lee, 1950; Siegel et al., 1982; Stager & Ludlow, 1993). These factors increase from short delays up to a maximal disfluency at around 200 ms (Black, 1951; Stuart, Kalinowski, Rastatter, & Lynch, 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 & 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 begins being affected by delayed feedback at around the level that the delay can be detected (around 40 ms; Jänke, 1991; Natke, 1999), and this increases as the delay increases, to a maximal effect around 200 ms (Black, 1951). A 200-ms delay is roughly 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 & Sakai, 2003). Furthermore, the effects of DAF are not specific to speech, and DAF also affects the production of rhythmic tapping and of musical performance (e.g., Pfordresher & 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 & 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, Reilly, & Guenther, 2008), or when attempts are made to mask self-produced speech during overt production (Christoffels, Formisano, & Schiller, 2007). Some of these articles 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 forward down the temporal lobe, along the lateral superior temporal gyrus and superior temporal sulcus, and a

posterior “how” processing stream, running posterior and medial to the primary auditory cortex (Scott & Johnsrude, 2003). While the comprehension of spoken language has been linked to the anterior stream of processing (Scott, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006), the posterior “how” pathway is activated by both speech perception and production (Okada & Hickok, 2006) as well as by articulation, even if the articulation is silent (Hickok et al., 2000; Wise, Scott, Blank, Mummery, & Warburton, 2001). This “how” pathway has been identified as central to the perceptual control of speech production (Dhanjal et al., 2008; Rauschecker & Scott, 2009). Posterior auditory fields have been specifically linked to the representations of “do-able” articulations (Warren, Wise, & Warren, 2005), which map sensorimotor 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 nonspeech acoustic information (Hickok, Buchsbaum, Humphries, & Muftuler, 2003). These same posterior-medial auditory fields have been reported to be activated in several functional imaging studies of altered auditory feedback (Hashimoto & Sakai, 2003; Tourville et al., 2008). Potentially, posterior auditory areas are recruited to compute and overcome timing difficulties or correct errors in speech production, whether or not these errors and timing differences are consciously detected (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 that respond to the delayed speech can be dissociated from those that are sensitive to the amount of delay (and that might be showing a purely acoustic response to the extra vocal information in the delay conditions).

Method

Positron emission tomography (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 functional magnetic resonance imaging [fMRI]), and the subjects can be scanned while they speak continuously (Blank, Scott, Murphy, Warburton, & Wise, 2002).

Subjects who could continue to speak under conditions of DAF were tested. These subjects did not “block” (i.e., there was no complete cessation 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 would 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 subject 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 and 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 pretested to ensure that they could continue to speak under DAF (i.e., that their speech production was not interrupted by blocking or long prolongations). The mean age was 42 years, 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 U.K. Department of Health.

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-labeled water bolus technique. All subjects were scanned while lying supine.

There were four different scanning conditions, which were presented in a pseudorandom 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 he or she 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 because 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*; Norton, 1952). 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 (297 mm × 420 mm) so that the font size was large enough to be readable, and these were 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 s 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 min 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 (10 = *very accurate*), how difficult they had found it to read aloud (10 = *very difficult*), and how well they thought they had timed their reading (i.e., did they feel they read at the right speed?; 10 = *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 analyzed using statistical parametric mapping (SPM99, Wellcome Department of Cognitive Neurology, 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 (using the Montreal Neurological Institute [MNI] template, 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 analysis of covariance with global counts as confound to remove the effect of global changes in perfusion across scans. The functional imaging data were then analyzed by generating contrasts to test the specific hypotheses at a whole brain level of analysis (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007). Results that were significant at a level of $t < 3.85$ (i.e., $p < .0001$; uncorrected for multiple comparisons) and

with a cluster size of (at least) 40 contiguous voxels are reported.

The first contrast, DAF > 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 condition(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—that is, 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 that were sensitive to the increase in delay but 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 that were *not* activated by the delay over no delay contrast (which could be showing a purely acoustic response to the “extra” delayed speech signal) but that also showed a sensitivity to increases in delay.

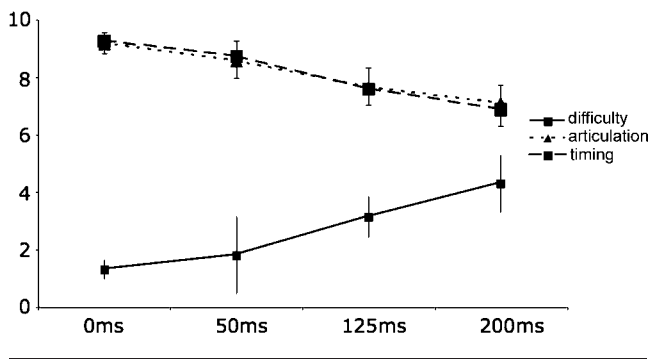
Negative versions of each contrast were run, with inverted weightings for the conditions. No significant activation was found for either of these contrasts. Finally, all three sets of subjective data were used as correlations in the functional imaging analysis, to identify brain areas that correlated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

Results

Behavioral

The subjects all reported finding that as the amount of delay increased, the perceived difficulty increased, and the perceived accuracy of articulation and of timing decreased with delay (see Figure 1). The behavioral 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 analyses of

Figure 1. The mean ratings for difficulty of speaking (out of 10, where 10 = *very difficult*), timing of speech (where 10 = *accurate timing*), and accuracy of articulation (where 10 = *accurate articulation*). The participants rated each score after every scan. Error bars show standard error of the mean.



variance. There was a significant effect of amount of delay on each measure: articulation, $F(3, 15) = 8.33, p = .002$, partial $\eta^2 = .626$; difficulty, $F(3, 15) = 7.9, p = .002$, partial $\eta^2 = .613$; timing, $F(3, 15) = 12.94, p < .0001$, partial $\eta^2 = .721$.

Functional Imaging

Contrast 1: DAF > 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 was any DAF, 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 are shown in Figure 2, with plots of the neural response (as percentage change in 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—that is, 0 ms < 50 ms < 125 ms < 200 ms. This contrast revealed a range of peaks

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

Region	p (corrected)	t	x	y	z
Left superior temporal gyrus	.000	6.88	-60	-30	8
	.005	5.53	-52	-12	0
Right superior temporal gyrus	.000	6.77	64	-38	10
	.012	5.33	54	-20	6

Note. The main peaks are shown in bold. The x , y , and z coordinates give the spatial location of each peak in millimeters in Montreal Neurological Institute (MNI) stereotactic space.

in the left and right superior temporal gyrus, described in Table 2. Peak activations in the left superior temporal gyrus 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.

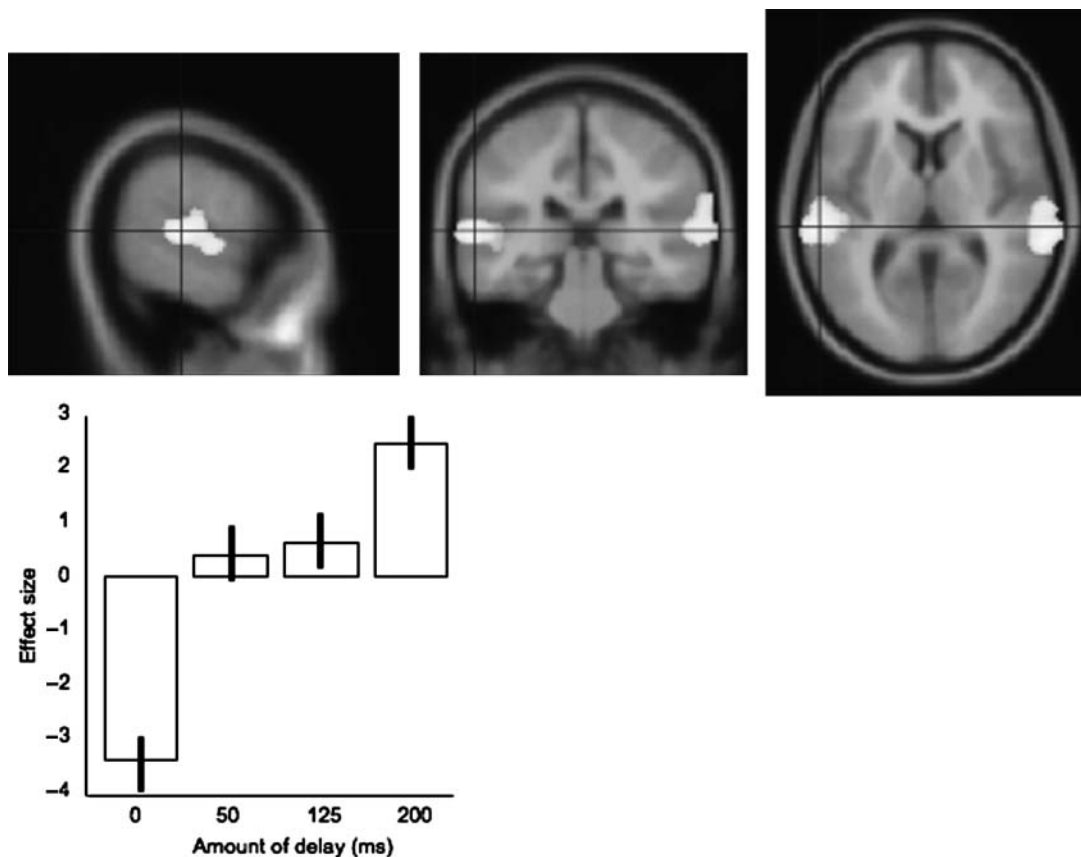
Contrast 3: effects of increasing delay, masking for existence of delay. The final planned comparison looked for brain regions that increased in activity with increasing amount of delay (0 ms < 50 ms < 125 ms < 200 ms) while masking exclusively for regions that showed any response to the DAF > none contrast ([50, 125, and 200 ms] > 0 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 sensorimotor cortex, were observed to DAF, 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 & Sakai, 2003; Tourville et al., 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 sensorimotor 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 that were sensitive to the existence of DAF but were insensitive to the amount of delay: Where there were neural responses to DAF, there was always some sensitivity to the amount of delay. The opposite pattern, however, was seen: In a left temporal-parietal junction, a small area was found that 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 & 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 that had a maximal response to DAF were not found in the more anterior

Figure 2. The main peak activation for the presence of any delay of vocal feedback, over the no delay condition ([50, 125, and 200 ms] > 0 ms). The upper panels show the data projected onto an averaged structural magnetic resonance imaging (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, centered around zero. Error bars show the standard errors of the parameter estimates.



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 & Benitez, 2007).

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

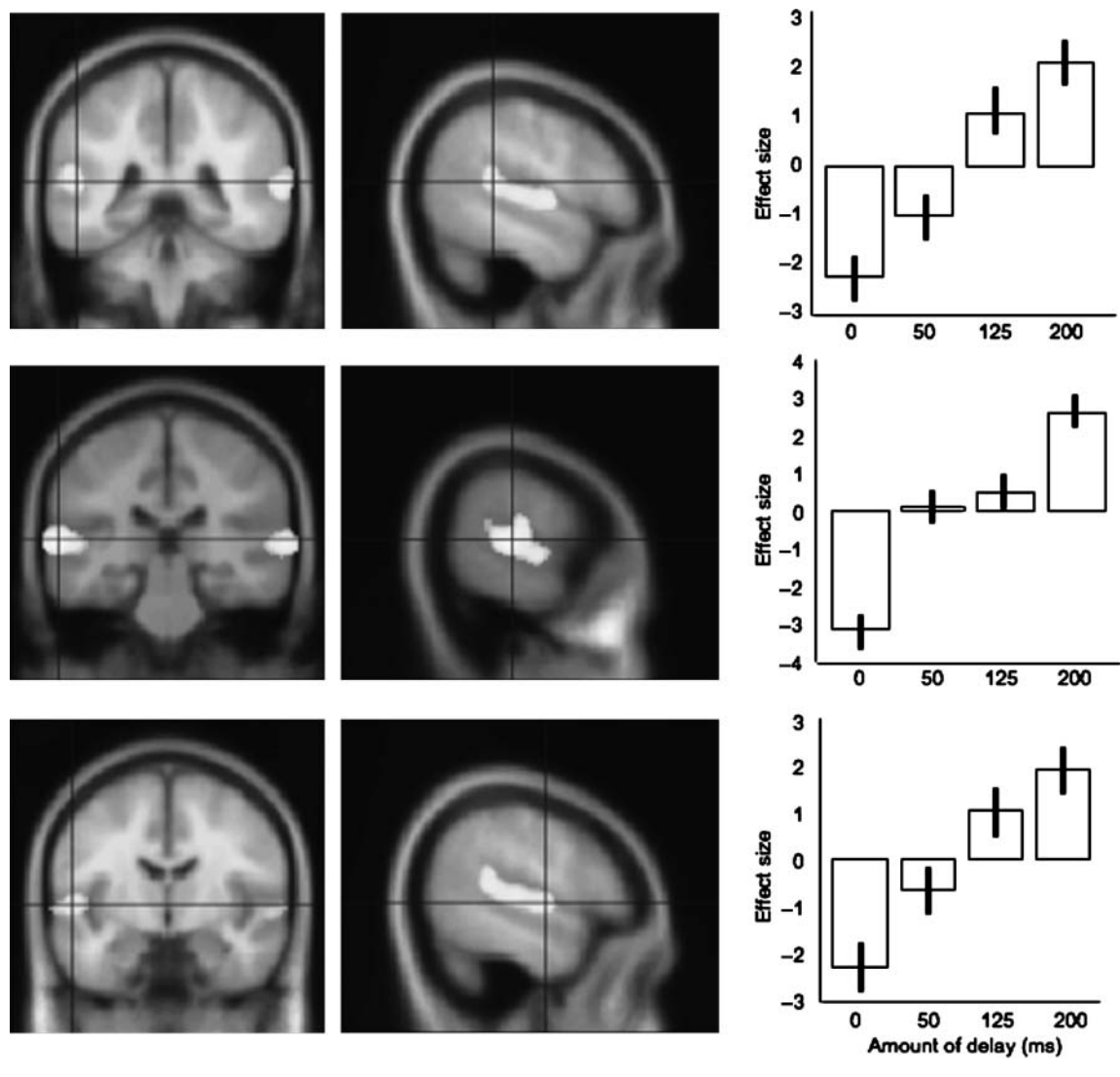
Region	<i>p</i> (corrected)	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>
Left superior temporal gyrus	.000	7.18	-62	-28	8
	.002	5.82	-54	-8	2
	.002	5.74	-52	-40	12
Right superior temporal gyrus	.000	7.15	68	-36	10
	.000	6.12	54	-20	4

Note. The main peaks are shown in bold. The *x*, *y*, and *z* coordinates give the spatial location of each peak in millimeters in MNI stereotactic space.

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 suggests 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 sensorimotor activation in speech (Scott & Johnsrude, 2003; Warren et al., 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 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

Figure 3. The three peaks of activation for the contrast identifying regional cerebral blood flow 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 rightmost panels show the average effect sizes for each condition at these peaks, centered 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.

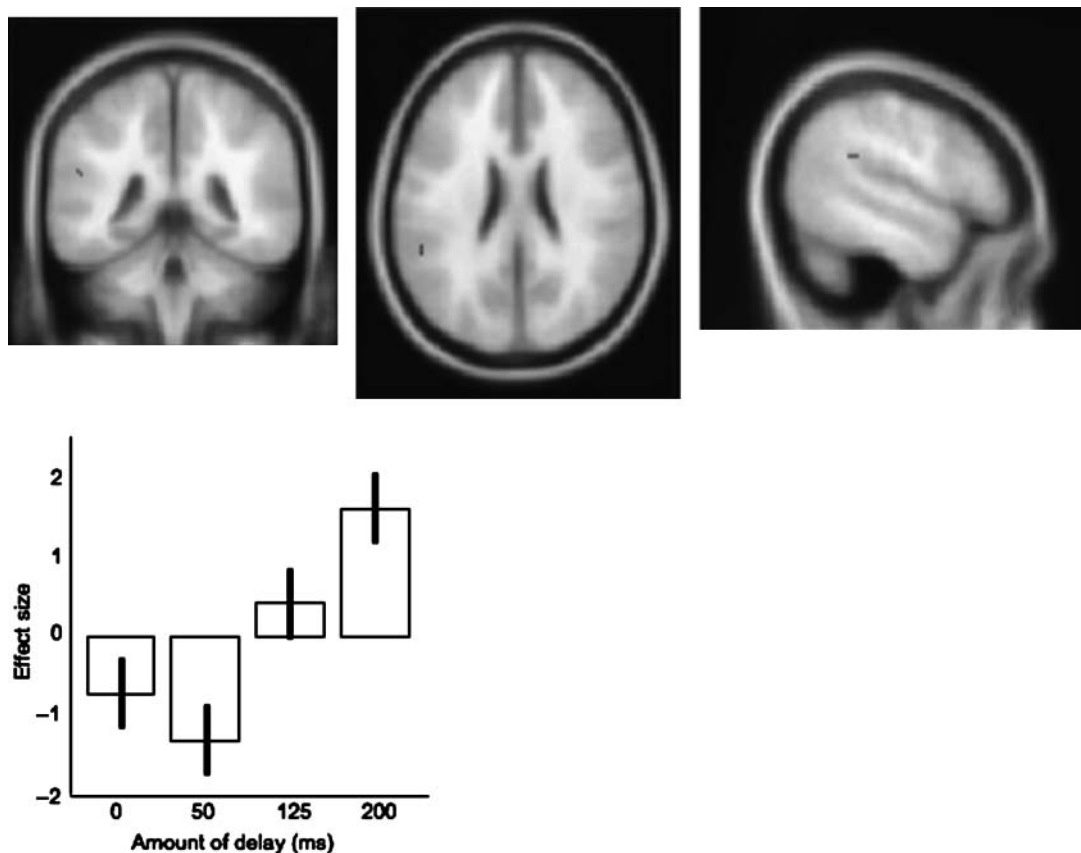


the voice but is directly involved in controlling for speech production used in these conditions of delayed feedback. This peak lay within 9 mm of the peak reported for articulation in Wise et al. (2001) in the x plane, 2 mm in the y plane, and 4 mm in the z plane, once transformed into the same Talairach coordinates. The more extensive activations that are sensitive to both the presence of DAF 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, such as opening and closing the jaw (Dhanjal et al., 2008), and this

region has been described as an auditory area that 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' article 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

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, and 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, centered around zero. Error bars show the standard errors of the parameter estimates.



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, and 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

Table 3. Peak of activation for the contrast identifying increases of activation for increasing amounts of delay, masking exclusively for any areas activated by the any delay > no delay contrast.

Region	<i>p</i> (uncorrected)	<i>t</i>	<i>x</i>	<i>y</i>	<i>z</i>
Left posterior temporal parietal junction	.000	3.59	-52	-40	24

Note. The *x*, *y*, and *z* coordinates give the spatial location of each peak in millimeters in MNI stereotactic space.

compensatory strategies to overcome rhythmic interference. These processes need not be speech-specific (e.g., Hickok et al., 2003; Howell, 2004; Pfordresher & Benitez, 2007).

There are limitations to the current study. There were no significant neural changes associated with the behavioral 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 or 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 DAF. 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 because many more trials can be run in each condition.

Only activity in the posterior temporal lobes, extending into ventral sensorimotor cortex, and the inferior parietal lobes was seen in the current study. The previous study of DAF, using fMRI (Hashimoto & 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 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 three or four runs of the experimental sequence, in which 16 of 64 spoken words were spectrally distorted, leading to a total of between 48 and 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 four 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

production. This basic finding is elaborated to show that these same posterior auditory fields are also sensitive to the *amount* of delay in DAF, which strongly implies that these regions are involved in both computing of the delay and 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 sensorimotor coordination 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

Frank Eisner and Sophie K. Scott were supported by Wellcome Trust SRF Award WT074414MA to Sophie K. Scott. Hideki Takaso and Frank Eisner contributed equally to this study.

References

- Black, J. W.** (1951). The effect of delayed sidetone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, 16, 56–60.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D.** (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1, 635–640.
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J.** (2002). Speech production: Wernicke, Broca and beyond. *Brain*, 125, 1829–1838.
- Bloodstein, O., & Ratner, N. B.** (2008). *A handbook on stuttering* (6th ed.). Clifton Park, NY: Thomson Delmar Learning.
- Christoffels, I. K., Formisano, E., & Schiller, N. O.** (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, 28, 868–879.
- Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J.** (2008). Perceptual systems controlling speech production. *Journal of Neuroscience*, 28, 9969–9975.
- Eliades, S. J., & Wang, X.** (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, 89, 2194–2207.
- Eliades, S. J., & Wang, X.** (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex*, 15, 1510–1523.
- Eliades, S. J., & Wang, X.** (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, 453, 1102–1106.
- Fairbanks, G.** (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders*, 20, 333–346.
- Fletcher, H., Raff, G. M., & Parmley, F.** (1918). *Study of the effects of different sidetones in the telephone set* (Report No. 19412, Case No. 120622). Western Electrical Company.

- Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D.** (2007). *Statistical parametric mapping: The analysis of functional brain images*. New York, NY: Elsevier.
- Fu, K. M., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., & Schroeder, C. E.** (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience*, *23*, 7510–7515.
- Fukawa, T., Yoshioka, H., Ozawa, E., & Yoshida, S.** (1988). Difference of susceptibility to delayed auditory feedback between stutterers and nonstutterers. *Journal of Speech and Hearing Research*, *31*, 475–479.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A.** (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*, 280–301.
- Ham, R., Fucci, D., Cantrell, J., & Harris, D.** (1984). Residual effect of delayed auditory feedback on normal speaking rate and fluency. *Perception and Motor Skills*, *59*, 61–62.
- Hashimoto, Y., & Sakai, K. L.** (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping*, *20*, 22–28.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T.** (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*, 673–682.
- Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., & Ugurbil, K. A.** (2000). 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*, *287*, 156–160.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M.** (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*, 1125–1138.
- Howell, P.** (1990). Changes in voice level caused by several forms of altered auditory feedback in fluent speakers and stutterers. *Language and Speech*, *33*, 325–338.
- 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 Research*, *1*(1), 31–46.
- Jäncke, L.** (1991). The ‘audio-phonatoric coupling’ in stuttering and nonstuttering adults: Experimental contributions. In H. F. M. Peters, A. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 171–180). Amsterdam, the Netherlands: Elsevier Scientific.
- Langova, J., Moravek, M., Novak, A., & Petrik, M.** (1970). Experimental interference with auditory feedback. *Folia Phoniatrica*, *22*, 191–196.
- Lee, B. S.** (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, *22*, 824–826.
- Lee, B. S.** (1951). Artificial stutterer. *Journal of Speech and Hearing Disorders*, *16*, 63–65.
- Lombard, E.** (1911). Le signe de l’elevation de la voix [The sign of voice rise]. *Annales des Maladies de l’Oreille, du Larynx, du Nez et du Pharynx*, *37*, 101–119.
- Mackay, D.** (1968). 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*, *43*, 811–821.
- Natke, U.** (1999). Die Kontrolle der Phonationsdauer bei stotternden und nichtstotternden Personen: Einfluss der Rückmeldelautstärke und Adaptation [Control of phonation duration in stuttering and nonstuttering persons: Influence of feedback loudness and adaptation]. *Sprache - Stimme - Gehör*, *23*, 198–205.
- Norton, M.** (1952). *The borrowers*. New York, NY: Puffin Books.
- Okada, K., & Hickok, G.** (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, *98*, 112–117.
- Pfordresher, P. Q., & Benitez, B.** (2007). Temporal coordination between actions and sound during sequence production. *Human Movement Science*, *26*, 742–756.
- Rauschecker, J. P., & Scott, S. K.** (2009). Maps and streams in the auditory cortex: How work in non-human primates has contributed to our understanding of human speech processing. *Nature Neuroscience*, *12*, 718–724.
- Scott, S. K., Blank, S. C., Rosen, S., & Wise, R. J. S.** (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.
- Scott, S. K., & Johnsrude, I. S.** (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neuroscience*, *26*, 100–107.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J.** (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *Journal of the Acoustical Society of America*, *120*, 1075–1083.
- Siegel, G. M., Schork, E. J., Jr., Pick, H. L., Jr., & Garber, S. R.** (1982). Parameters of auditory feedback. *Journal of Speech and Hearing Research*, *25*, 473–475.
- Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E.** (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *Journal of Computational Neurology*, *502*, 894–923.
- Stager, S. V., Denman, D. W., & Ludlow, C. L.** (1997). Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions. *Journal of Speech, Language, and Hearing Research*, *40*, 832–847.
- Stager, S. V., & Ludlow, C. L.** (1993). Speech production changes under fluency-evoking conditions in nonstuttering speakers. *Journal of Speech and Hearing Research*, *36*, 245–253.
- Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K.** (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *Journal of the Acoustical Society of America*, *111*, 2237–2241.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H.** (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*, 1429–1443.
- Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S.** (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*, 499–503.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neuroscience*, *28*, 636–643.

Wise, R. J. S., Greene, J., Büchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*, 1057–1061.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., & Warburton, E. (2001). Identifying separate neural subsystems within Wernickes area. *Brain*, *124*, 83–95.

Received January 26, 2009

Revision received May 8, 2009

Accepted October 13, 2009

DOI: 10.1044/1092-4388(2009/09-0009)

Contact author: Sophie K. Scott, Institute of Cognitive Neuroscience, UCL, 17 Queen Square, London WC1N 3AR, United Kingdom. E-mail: sophie.scott@ucl.ac.uk.

**The Effect of Delayed Auditory Feedback on Activity in the Temporal Lobe
While Speaking: A Positron Emission Tomography Study**

Hideki Takaso, Frank Eisner, Richard J. S. Wise, and Sophie K. Scott
J Speech Lang Hear Res 2010;53;226-236; originally published online Nov 30, 2009;
DOI: 10.1044/1092-4388(2009/09-0009)

The references for this article include 11 HighWire-hosted articles which you can
access for free at: <http://jslhr.asha.org/cgi/content/full/53/2/226#BIBL>

This information is current as of May 4, 2010

This article, along with updated information and services, is
located on the World Wide Web at:
<http://jslhr.asha.org/cgi/content/full/53/2/226>

