Strength of German accent under altered auditory feedback

PETER HOWELL and KATHARINA DWORZYNSKI
University College London, London, England

Borden's (1979, 1980) hypothesis that speakers with vulnerable speech systems rely more heavily on feedback monitoring than do speakers with less vulnerable systems was investigated. The second language (L2) of a speaker is vulnerable, in comparison with the native language, so alteration to feedback should have a detrimental effect on it, according to this hypothesis. Here, we specifically examined whether altered auditory feedback has an effect on accent strength when speakers speak L2. There were three stages in the experiment. First, 6 German speakers who were fluent in English (their L2) were recorded under six conditions—normal listening, amplified voice level, voice shifted in frequency, delayed auditory feedback, and slowed and accelerated speech rate conditions. Second, judges were trained to rate accent strength. Training was assessed by whether it was successful in separating German speakers speaking English from native English speakers, also speaking English. In the final stage, the judges ranked recordings of each speaker from the first stage as to increasing strength of German accent. The results show that accents were more pronounced under frequency-shifted and delayed auditory feedback conditions than under normal or amplified feedback conditions. Control tests were done to ensure that listeners were judging accent, rather than fluency changes caused by altered auditory feedback. The findings are discussed in terms of Borden's hypothesis and other accounts about why altered auditory feedback disrupts speech control.

One model of how the brain controls speech is that auditory output is continuously monitored and used for making ongoing corrections when speech is in error. According to such accounts, the brain receives sensory information about speech output, and this is used in a manner analogous to the way a thermostat controls a central heating system (Fairbanks, 1955). The evidence for accounts like these was critically reviewed by Borden (1979, 1980). She raised two compelling arguments as to why auditory feedback cannot be involved in ongoing control of speech production. First, speakers who are adventitiously deafened late in life still speak. Second, speech sounds are produced rapidly, so the outcomes of perceptual analyses for on-line regulation of speech would not be available in time to use for correcting speech. She considered, therefore, that *established* speech is under open-loop control (operates without external feedback monitoring).

In her examination of the literature on other speaker groups, Borden (1979, 1980) considered that children with developing speech skills, people who are learning or have learned a second language, and dysfluent speakers do monitor auditory feedback. The factor that these

speaker groups have in common is that their speech systems are *vulnerable*. The evidence interpreted as support for vulnerability is that alterations to the sound of the voice affect speech control more in these speakers than in fluent adult native speakers of the language. As a benchmark against which to assess the claims about the effects of altered auditory feedback (AAF) on vulnerable speaker groups, the effects on fluent adult speakers speaking their native language under AAF are summarized briefly. There are three main ways that the auditory properties of the voice have been altered: Its intensity has been changed (amplified or attenuated), its time of arrival at the ear has been altered (i.e., speech heard after a delay), and its frequency composition has been modified (shifted up or down in frequency or had frequency bands filtered out). If intensity of the voice is changed before it is fed back to a listener, speakers offset about half the imposed change in feedback level (Lombard effect). When the voice is masked by noise, speakers increase their voice level to offset about half the decrease in perceived voice level (Fletcher effect; further details are given in Lane and Tranel's [1971] review). When fluent speakers speak under delayed auditory feedback (DAF), their speech becomes slurred, and they make speech errors (Lee, 1950). Speech rate is decreased, all forms of speech dysfluencies occur (for instance, vowels are drawled), and voice level and pitch increase (Fairbanks, 1955). Delaying the arrival time of speech leads speakers to treat the altered form of speech as if it were noise (it produces a Fletcher, not a Lombard, effect). Alteration of the response to the delayed sound from speech to noise would render it unusable for

This research was supported by a grant to the first author from the Wellcome Trust. Grateful thanks are due Bill Barry, who advised on the characteristics of German accent in English speech. Also, many thanks to East Carolina University for supplying the Digitech equipment used in this study. Correspondence should be addressed to P. Howell, Department of Psychology, University College London, Gower St., London WC1E 6BT, England (e-mail: p.howell@ucl.ac.uk).

feedback control. With an alteration to the frequency of feedback, as with a change in intensity, speakers make a response complementary to that arising from the experimental manipulation. Thus, if vocal feedback is shifted up in frequency, the speaker shifts voice pitch down, and vice versa (Elman, 1981; Garber & Moller, 1979).

The early explanations offered for the disruptions described were either, as was just mentioned, that speakers compensate for the alterations made by the experimenter (e.g., frequency alteration; Elman, 1981) or (a type of explanation applied principally to DAF) that the speaker continues in his or her attempt to use the AAF for controlling ongoing speech (Lee, 1950). Both these explanations are variants of feedback-monitoring views (in the first case, an appropriate compensatory response is made to offset the effect of the alteration, whereas in the second, the alteration itself has a characteristic effect on the output form of speech, as when drawing occurs). More recent explanations have accounted for the effects of AAF without recourse to feedback monitoring and correction processes. Thus, Lane and Tranel (1971) offered an interpretation of the effects of alteration to intensity of feedback in terms of the speaker's adjusting output for the listening requirements of his or her audience. Howell, Powell, and Khan (1983) accounted for the effects of DAF as being due to poorer performance when speech takes place along with an asynchronous rhythm (the delayed version of the voice in the DAF situation). Howell and Sackin (2000) and Howell, Rosen, Hannigan, and Rustin (2000) explained the effects of DAF and frequency-shifted feedback (FSF) as being due to degradation in performance when speaking along with concurrent separate perceptual streams. DAF would be one way of producing a second sound "stream" that is separated from the voice because of the temporal asynchrony. Spectral shifts (such as those that occur under FSF) also create a second sound source that is perceptually separate from other synchronous sounds (Bregman, 1990). Howell and Sackin and Howell et al. (2000) extend Howell et al.'s (1983) account of speech disruption under AAF to spectral signal disparities as well as temporal ones (DAF).

People who stutter are highly vulnerable to alterations to feedback in a unique way. DAF (Gibney, 1973; Soderberg, 1969; Webster, Schumacher, & Lubker, 1970) and FSF (Howell, El-Yaniv, & Powell, 1987; Kalinowski, Stuart, Sark, & Armson, 1996; Reed & Howell, 2001; Stuart, Kalinowski, Armson, Stenstrom, & Jones, 1996) markedly *improve* the fluency of these speakers. Early findings comparing DAF in fluent children and adults suggested that its effects were more apparent as age increased (Chase, Sutton, First, & Zubin, 1961). However, later experiments failed to replicate this pattern, and it is now generally accepted that younger children are more susceptible to DAF than are older children or adults (MacKay, 1968; Siegel, Fehst, Garber, & Pick, 1980; Siegel, Pick, & Garber, 1984). Several researchers have

investigated the effect of DAF on bilingual speakers (or polyglots), again with conflicting results. An early study by Rouse and Tucker (1966) observed more speech disruption when Americans read in their native language (L1) than when foreigners were required to read in English (their second language, L2). However, MacKay (1970) pointed out that language familiarity was confounded with subject group, so the differences between groups could not be unambiguously attributed to language familiarity. In MacKay's (1970) own study employing DAF, the speech of native German speakers who could speak fluent English was compared with that of native English speakers who could speak fluent German. They were asked to read sentences in their more and less familiar languages and in an unknown language. He found that giving DAF when the speaker was speaking the unknown language affected speech more than when DAF was given in L2 or L1 (L2 also was affected more than L1). These findings are supported by a more recent study on learners of Japanese and Spanish, who spoke under DAF (Kvavik, Katsuki-Nakamura, Siegel, & Pick, 1991). On the other hand, Fabro and Daro (1995) reported no difference in speech disruption between L1 and L2 in a study on simultaneous translators. These subjects, in contrast with the bilingual subjects in previous studies, had greater knowledge and higher speech rates in their L2. A feature of this study is that it highlights a possible confounding variable in the previous investigations: Different effects of DAF on L1 and L2 could simply be due to lower baseline fluency in L2 (Klein, Zatorre, Milner, Meyer, & Evans, 1994).

The starting point of our study was an anecdotal report of another situation in which speech appears vulnerable: Actors have difficulty maintaining an accent they adopt in auditoria where intensity and time (echo) disruptions occur. The timing, amplitude, and spectral changes to voice feedback in the auditoria are similar to those that occur in AAF experiments. The present study investigates this effect more formally and is specifically intended to determine whether a foreign accent is more apparent under altered feedback conditions like those required on stage. Brennan and Brennan (1981b) define foreign accent as "a manner of pronunciation different from standard speech with the grammatical, syntactical, and lexical levels consistent with the standard" (p. 488). If foreign accent is more perceptible under conditions of AAF, the speech would have a form like that used at an earlier stage of acquisition. In itself, this raises interesting questions about what speakers are doing to maintain the output form of the acquired language (see the Discussion section). Another factor that makes higher perceptibility of foreign accent under AAF worthy of study is that if these speakers have difficulty maintaining their fluency under these conditions, it would afford an interesting contrast with the well-known fluency-enhancing effects of AAF in people who stutter. The issue about what might be the common factor behind the improvement in people

who stutter and heightened susceptibility to AAF in vulnerable speech systems (here, L2 speakers) is also returned to in the Discussion section.

Several previous studies have focused on accent strength and the perception of accentual differences across speakers, although not under AAF (Brennan & Brennan, 1981a, 1981b; Flege, Bohn, & Jang, 1997; Flege & Fletcher, 1992; Flege, Munro, & MacKay, 1995; Magen, 1998; Ryan & Carranza, 1975; Thompson, 1991). Flege et al.'s (1995) study shows that native speakers can detect accents even in individuals who learned L2 as young children (5–8 years). This finding is particularly important for the present study, because most of the speakers recorded had spoken two or more languages from a very early age (see Table 1). The subtle distinctions involved in assessing an individual speaker's accent strength was deemed to be reliant on the sensitivity of native speakers to foreign accents in their language. The majority of the studies on accent strength have relied on judges' intuition to place speakers on a scale of native speech to strongest accent (Thompson, 1991). In Flege et al.'s (1995) study, for example, judges were required to position the lever on a response box at some point along a range defined by the labels "native speaker of English—no foreign accent" (top), "medium foreign accent" (middle), and "native speaker of Italian—strongest foreign accent" (bottom). Other studies have investigated individual, specific aspects of the accents (Brennan & Brennan, 1981b; Magen, 1998; Munro, 1993).

Studies like that of Flege et al. (1995), described above, obtained reliable and valid accent judgments even with untrained listeners. Here, however, throughout training and testing, subjects with some degree of phonetic training were employed. Such subjects were used so that they could make use of this analytical training. Minimum training was 2 h specifically on accent characteristics, apart from the instructions given during training. The characteristics that were assessed in training were chosen so as to provide a structural approach to what the subjects should listen out for, but the listeners were not restricted in what they used in subsequent tests in any way. Since there appear to be no previous studies on judgments of German accents in English, these data are reported, although it should be borne in mind that there are major procedural differences between the test procedures employed to obtain them and those for accent testing under NAF and AAF.

After the subject had been trained to detect individual characteristics, they made holistic judgments about accent (Flege et al., 1995). The holistic judgments allowed the listeners to use whatever they found useful in deciding about accent. Even so, they were reminded of the characteristics they had learned about (see Thompson, 1991, for a balanced discussion of the respective values of analytical and holistic judgments). Our main reason for making holistic judgments, besides its being what is done in other research in this area, is that judging an accent might be considered a decision about *integral* (holistic)

rather than *separable* (analytical) speech dimensions that are judged differently (Garner, 1974). Another procedural decision was to judge speech samples from a speaker against other samples he or she produced, not against samples from other speakers. Magen (1998) criticized comparing accents across speakers (see also MacKay, 1970). Procedures appropriate for testing individual subjects are needed so individual speech characteristics do not affect accent judgments. The problems involved in comparing accents across speakers with different individual characteristics would be exacerbated in the present study, since it is known that people vary in their susceptibility to the effects of AAF (Howell & Archer, 1984). Taking Magen's point, then, assessments in the main test for accent differences caused by AAF were made across speech samples for each individual speaker. Another important point raised by Magen's study concerns how to maximize perceptual differences between conditions. She used 7-point ratings and emphasized to her subjects that they should use all of the scale values when judging samples, in order to give some operating space to establish differences between judged characteristics of accent. In our study, the listeners put the samples from each speaker in rank order (no ties allowed). This achieved something similar to Magen's instructions to space samples along the entire rating continuum. The changes in procedures, between training and testing, from analytical to holistic judgments and from rating to ranking meant that testing and training could not be compared directly. As was said above, the training was intended to provide a structural approach to accent decisions, rather than to teach the judges to identify particular characteristics in samples *per se*. In this regard, the procedure paralleled, in some respects, what occurs in phonetic training. Analytical phonetic skills are taught as a basis for assessing fluency. Once acquired, they are then usually used without performing an explicit phonetic analysis.

Although reservations have just been expressed about comparing samples across speakers, it was necessary to do this to make one check. It is conceivable that holistic ranking judgments across different feedback conditions could be based on a ranking of fluency differences between those speaking conditions, since AAF affects speech fluency. If listeners maintain accent groupings (ranking German speakers of English at one end and English speakers at the other) for all speaking conditions, this possibility is less likely. Speakers of both languages have their fluency affected by AAF (MacKay, 1970; Rouse & Tucker, 1966). Thus, if fluency, not accent, is being judged, the division along the accent group dimension should be poorer in AAF conditions than in NAF conditions.

According to Borden's hypothesis, that auditory feedback is employed particularly in vulnerable speech systems, such as when speaking L2, it was expected that the accent would be more pronounced in altered amplitude, FSF, and DAF conditions, as compared with a normal auditory feedback (NAF) baseline. Speakers were also required to vary speech rate (under NAF) in order to check

Table 1
Language Usage Details From the 6 German Speakers Recorded Speaking English Under AAF

Characteristic	Speaker					
	1	2	3	4	5	6
Age	29.6	28.3	52.2	21.5	28.0	26.6
Gender	F	F	F	M	M	M
Age of L2 acquisition	Exposed to English from birth	10	10	15	6	13
Age of arrival in L2 country	UK since age 21; lived in the US	UK since age 20	UK since age 45	UK since age 17	UK since age 22	UK since age 22
Other languages spoken (excluding German & English)	French, Italian	French	Spanish	Romanian, French, Spanish	French, Czech, Spanish	Italian
Nationality of parents (other language of parents)	Father, German (Russian, French, Turkish, English); mother, American (German, Russian)	Both parents German (English, French) (emigrated to Germany before subject's birth; German, French, English)	Both parents German (French)	Both parents Romanian (French, English, Spanish)	Both parents German (English, French)	Both parents German
Percentage of use of English (Educ/Prof)	95	95	100	95	100	95
Percentage of use of English (Soc/Priv)	65	55	35	50	50	30

Note—Percentage of time spent speaking English is given separately for educational/professional and social/private life. German made up the complementary percentage in every case except Speaker 4, for whom German was 40% and Romanian 10%.

whether rate changes associated with AAF mediated any effects found (Fabro & Daro, 1995).

METHOD

Subjects

The subjects were employed for two types of task. They had to either speak in various test conditions or assess the speech from the recordings. Two sets of speech recordings were made. One was for training the judges, and the other provided the test material that was assessed. The details about the subjects who participated in these stages will now be given.

Speakers for the training and experimental conditions. The set of recordings for training the judges was from 4 German speakers fluent in English and 4 native British English speakers (2 male and 2 female in each language group). None of these took part in the recordings for the experimental conditions. The German speakers had a range of strength of accent (from 1 subject who had entered the United Kingdom 3 weeks before being recorded, to speakers who had spoken German and English in nearly equal amounts from an early age). Median age was 24.1 years for the German speakers and 25.4 years for the native English speakers. Two of the German speakers in the training phase of the experiment had one English- and one German-speaking parent and reported speaking both languages to an equal extent (since they lived and studied in England, English was used more frequently at the time of the test, whereas the opposite applied during their school years). Of the remaining 2, 1 had been in England for roughly 4 years in full-time higher education, had learned English at school in Germany (starting at an age of 12 years), spoke to most of his relatives in German, and conducted his education in English. The 4th speaker had entered England 3 weeks previously and indicated in the questionnaire that she predominantly spoke German in both professional and social situations (85% German, 15% English).

In the experimental conditions, 6 (different) German subjects were recorded (3 male and 3 female). Their ages ranged from 21.4

to 52.2 years (individual ages are given in Table 1). Again, all the speakers had German as their mother tongue. The age at which they started to acquire English varied, and details of this and other aspects of language usage (e.g., other languages spoken), collected by questionnaire, are also summarized in Table 1. A breakdown of percentage of use of English for the experimental group is given for two speaking environments (educational/professional and social/private) in Table 1.

Six British English speakers (3 male and 3 female) were recorded to serve as a control group. Their ages ranged from 18.0 to 21.5 years (median age was 20.0 years). All of these speakers had English as their first language. Two of them spoke no further language apart from English; 2 spoke French to university entrance standard (as well as Italian and German to elementary standard). The other 2 spoke fluent Tamil or French, and another, good Spanish. All were born in England and attended English-speaking schools, and English was spoken solely in their home environments.

Judges. The judges were 8 native English speakers (4 male and 4 female) ranging in age from 20 to 50 years. The subjects reported whether they spoke any other languages. Five spoke no other language beyond elementary school level, and the 3 remaining each had one other language (Welsh, Creole, and French). All had some phonetic training, varying from extensive experience in transcribing spontaneous speech to a minimum of 2 h on accent characteristics, as part of a course they had attended. All the judges were given additional training to assess phonetic characteristics that would distinguish German from English speakers when both groups of speakers spoke English (i.e., the training recordings). They then used aspects of these forms of assessment to rank separately the AAF and the control recordings of each speaker as to accent strength.

Procedure

Recordings for the training and experimental conditions. The procedures employed when making the training recordings and the recordings in the experimental conditions were similar. The procedure is described in detail for the experimental recordings, and then the differences in the training recordings are described. All the

speakers were recorded individually in an Amplisilence sound-attenuated booth. Each subject spoke under three feedback conditions (amplified feedback [AMP], FSF, and DAF), a normal auditory feedback condition (NAF), and two variations in rate (slow/fast), also while hearing NAF. Auditory feedback was altered using commercially available hardware (Digitech model 400). In all the speaking conditions, the subjects wore stereo headphones (model RS 250-924 headset). Two Sennheiser K6 microphones were positioned in front of the subjects, approximately 6 in. from their lips in direct line with the mouth. One was connected to the feedback device; the other was connected to a DAT recorder. In the DAF condition, the delay was 200 msec, which is maximally disruptive according to MacKay (1970). In the FSF condition, the frequency was shifted down by half an octave. When speaking under the amplified condition, the amplitude was increased by 16 dB. Under NAF (at normal, fast, and slow rates), speech was relayed back over headphones with no amplification or temporal or frequency distortion.

In each experimental condition, the subjects read a different stretch of speech. The passage of speech used was from a Thomas Hardy novel that was unfamiliar to all the speakers (Hardy, 1983, pp. 58–60). This was a descriptive passage with no conversation similar to those employed in other studies (Flege et al., 1995; MacKay, 1970). The passage was divided into six sections of approximately equal length (ranging between 123 and 160 words). The stretches employed were checked to ensure that they contained instances of all the consonantal and vowel phones the listeners had been trained on. The stretches allowed the complete range of prosodic dimensions that distinguishes English from German to be reflected, although the points at which they occurred could not be controlled, since their occurrence is to some extent a matter of personal style.

The subjects in the experimental group were told before the experiment started that they would be required to speak under several conditions in which they would hear their voices distorted. The subject was seated in front of the microphones. He or she was asked to put on the headphones and read the first sentence of the instruction sheet under NAF in order to check recording levels. The subject was then instructed about which condition he or she would receive next, so that the alteration (where applicable) would not produce a surprise response that affected his or her speech. The subjects were then given the stimulus material. The experimenter set the equipment and signaled the subject to read the selected text. Between conditions, the feedback was reset to NAF to instruct the subject about the next condition. For the slow and fast rate conditions, the subjects were asked to read faster/slower than usual by about 20% (actual rates produced were 20.7% slower and 23.1% faster; corresponding rates for the English controls were 15.4% and 16.9%). These rates were not as fast or as slow as the speakers could go but were noticeably faster or slower than they would ordinarily speak. Rate variation was limited in extent, to prevent misreadings from occurring. The six sections of the Hardy story were assigned at random to AAF and control conditions for the 1st subject and then distributed so that each section occurred once for each of the experimental conditions across subjects. The order in which the subjects did the experimental conditions was counterbalanced across subjects by Latin square.

Recording of the English and German speakers who were used for training the raters was obtained as in the manner described for NAF. A story was constructed for this purpose that included at least two examples of each of the consonant and vowel characteristics associated with strength of accent (described next). Note that, as earlier, it was not possible to control for prosodic factors.

Assessment of strength of accent in German and English speakers and in the experimental conditions. The judges who rated the speech were first trained to listen for specified characteristics of German accent. They were aware that they could use this training later to assess variations in accent strength by German speakers speaking English in the experimental conditions. Both these assessments took place individually in the Amplisilence sound-attenuated booth.

For training, a set of characteristics identified as likely to aid identification of a German accent was prepared. The characteristics are given in Table 2. The subjects were allowed as much time as necessary so that they were completely happy with these concepts. Examples were demonstrated by a native German speaker who was fluent in English (K.D.).

The judges listened to the tape to make each judgment about each characteristic individually. While making a judgment about each sample, they replayed the complete recording from DAT tape through a Fostex 6301B amplifier that was then heard over an RS 250-924 headset. The judges were allowed to rewind and relisten to the sample as often as they required. They made judgments about the set of characteristics in the same order for each speaker, and each characteristic was assessed individually. When they were satisfied with their decision about a characteristic, they recorded their response. They then repositioned the tape for replay in the same manner for the next characteristic until all the characteristics had been completed. Summary statements based on the descriptions given above were generated for each characteristic, and each judge gave a number between one (*strongly agree*) and seven (*strongly disagree*) on the speech of all 8 (4 English, 4 German) training speakers (Likert, 1932). Seven indicates a perfect English pronunciation in all cases. When they had completed their assessment of a speaker, they went on to assess the next. The order in which they assessed the speakers was counterbalanced across judges, using a Latin square.

After training, the judges assessed the experimental recordings. The first step in the procedure for this was to select samples 20 sec long from each recording made under AAF and control conditions. These were checked to ensure that they contained all the phones identified in training. The samples were transferred from DAT tape onto computer. The samples were selected and replayed from the PC. They were played to the judges via the Fostex amplifier and RS headset.

The judges were tested individually, seated in front of the computer. They were informed that they would hear samples of 6 different speakers. It was emphasized that they should use aspects they had rated in the previous session and any others they found useful (a reminder was given of the various characteristics). They were told that they would hear six samples of 1 individual speaker (the speech samples used were randomized across judges). Each sample was played once each. After each sample, they were allowed to make notes about its characteristic that would help in judging its accent strength. The time allowed for this was under the judge's control. When they had completed their note taking about the selected sample, they pressed the return key of the PC and heard the next. They passed through all six samples of the selected speaker, hearing each 20-sec sample once. After they had heard all the samples from a speaker, they rank ordered them according to increasing accent strength and entered the rank of each sample into the PC (ties were not allowed). They then heard the samples in the order they had specified. They indicated whether they agreed with their preliminary ranking by pressing the "y" or the "n" key on the keyboard. If the answer was "no," they were allowed one further opportunity to change the order. After that, samples of the next speaker were presented. The experimenter made sure that each judge understood the procedure and stayed with them while they assessed the 1st speaker, in case questions arose. The experimenter then left them on their own to rate the other 5 speakers according to the same procedure. The judges were not told if and what form of feedback was administered at the time of recording on any sample.

All the listeners were recalled to judge German speaker samples against English control samples with regard to accent strength. Before they did this, they had a refresher session on the accent attributes that involved judging selected training samples. Samples from 6 speakers were chosen for each judge (always consisting of 3 English and 3 German speakers). Each listener heard samples only from 6 speakers, since they reported being uneasy about ranking more than this. All 6 English and all 6 German speakers were heard by an equal number of listeners. The samples from English speak-

Table 2
Training Characteristics Used to Identify German Accent

Target		Result		Remark
Sound	Example	Sound	Example	
Consonant Characteristics				
“th”	<i>other</i> or <i>thing</i>	“s” or “z”	<i>see</i> or <i>lazy</i>	Vice versa
/w/	<i>wine</i>	/v/	<i>vine</i>	
/r/	<i>retrograde</i>	slightly trilled	<i>burrr</i> or <i>purrr</i>	
Final Consonants				Final consonants that are voiced in English tend to be produced as voiceless. This makes word pairs like <i>le<u>ave</u>-leaf</i> or <i>pea<u>s</u>-pea<u>ce</u></i> difficult to distinguish.*
Vowel Characteristics				
Diphthongal Vowels				
/eɪ/ and /oʊ/	<i>ei<u>ght</u>/wei<u>gh</u></i> and <i>so<u>ul</u>/tho<u>ugh</u></i>	/e:/ and /o:/	<i>fe<u>ll</u></i> or <i>gl<u>ow</u></i>	Diphthongal vowels (where the tongue glides from one vowel position to another), monophthongalized (tongue remains stationary).
/i:/ and /u:/	<i>fee<u>l</u>/be<u>ad</u></i> and <i>two<u>o</u>/sh<u>oe</u></i>			Pronounced too “pure,” i.e., without a slight diphthongalization, which occurs in English.
/æ/ The “u” vowel	<i>pa<u>ck</u>/ba<u>d</u></i> <i>hu<u>u</u>/du<u>ll</u></i>	“e” open central vowel /a:/	<i>se<u>t</u>/ge<u>t</u></i> <i>ca<u>rt</u>/ba<u>rn</u></i>	Similar to the Cockney accent.
The “o” vowel	<i>ho<u>u</u>/do<u>g</u></i>			Spoken too rounded and too close to the vowel sound in words like <i>sa<u>w</u></i> .
The schwa	<i>A<u>l</u>one/mo<u>th</u>er</i>			Too open like /a:/, as in <i>la<u>rge</u></i> , or too close, as in <i>he<u>r</u>/se<u>r</u>ve</i> .
Prosodic Characteristics				
Target	Remark			
Function words	too “strong” and they are unreduced, i.e., a lack of “weak forms” (such as <i>and</i> shortened to /nd/ or <i>have</i> to /ʒv/ in connected speech), as in “ <i>to have and to hold</i> ” becoming “ <i>to /ʒv/ /nd /to hold</i> .”			
Stress on vowel onsets	overexaggerated.			
Alternate accents	It is common in German to alternate accents in a more exaggerated way than in English from low to high, i.e., the second syllable is stressed with higher pitch as in <i>Zam<u>be</u>zi</i> . This can indicate a German accent.			

*Note that, although all final stops are pronounced as if voiceless in German, this would not necessarily apply to English spoken by a German speaker. The extent to which a speaker is able to voice final stops appropriately, then, may constitute an important cue to German accent.

ers were also counterbalanced with respect to which German speakers samples they were paired with across listeners, to ensure that arbitrary pairings did not affect the results. The samples for a listener's 6 speakers were ranked with respect to increasing accent strength for each speaking condition (speaking conditions were ranked independently).

RESULTS

The scores on the training characteristics were checked to establish whether German speakers speaking English can be separated from English speakers, also speaking English. The means, standard deviations, and

95% confidence intervals of the scores of each characteristic were calculated for all the judges for each speaker. The means and 95% confidence intervals for each characteristic are given on the left of Figure 1 for the German speakers and on the right for the English speakers. In the figure, the characteristics judged are given in the same order as in Table 2 (i.e., with consonant characteristics to the left, vowel characteristics in the center, and prosodic characteristics on the right).

Comparison of the English speech of the native German speakers with the native English speakers shows that it was judged more foreign (lower numbers) than the

Figure 1. Mean Likert scale scores (ranging between one and seven) and their associated 95 % confidence intervals for each characteristic judged for the native German (left) and English groups (right). Low values indicate that the speech is judged "more foreign." The characteristic judged is arranged along the abscissa in the same order as that enumerated in the Method section.

corresponding speech of the native English speakers (higher numbers). The first four characteristics (those associated with consonants) are less variable for the English speaker group than the other characteristics. This is partly attributable to many of the samples' receiving a maximum score for the characteristic for these speakers. Inspection of the data showed that none of these characteristics judged in the speech of any of the native German speakers had a mean score above five. The greater variability in German speakers is due in part to their being given scores away from the maximum and to the speakers having a wide range of accent strengths. Finally, it can be seen that there was hardly any overlap in the confidence intervals between the two speaker groups. Mann-Whitney *U* tests were performed on each of the 13 characteristics (a nonparametric test was used, since the variances differed between speaker groups). This was significant at a $p < .001$ in all cases, after Bonferroni corrections for multiple tests were made.

Next, a discriminant function analysis was performed to see what characteristics from the questionnaire an optimized classifier would use to separate speakers into their respective groups. The target classification was specification of whether the speaker was English or German, and the predictor variables were the scores of the characteristic for each of the raters. The data show that the four consonant characteristics alone were successful at classifying 92.2% of cases. The misclassified cases were spread

across different judges. The misclassified cases usually were those from the more fluent speakers, in particular those who had parents one of whom was a native English and the other a native German speaker.

The previous analysis indicates that the characteristics that the judges had been trained to use were successful at separating German speakers from English speakers when both groups were speaking English. Now that these characteristics had been learned, were the judges able to rank order the accent strengths of German speakers speaking English in the experimental conditions? The first question examined is whether the listeners ranked the accent strength in all six experimental conditions in a systematic fashion. The rankings given to each sample were analyzed to see whether there was a significant concordance (agreement) across assessors, using Kendall's *W*. The mean ranks for each of the six conditions are given in the top section of Table 3. Kendall's *W*, a chi-square on this, the degrees of freedom, and associated probability are given in the bottom section of Table 3. It can be seen that the concordances were significant for 4 of the 6 speakers, and a 5th one was right on the margin of significance ($p = .051$).

The present analysis includes the two rate conditions intended as controls. These recordings were made to ascertain whether rate variations that occur when feedback is altered make the speech sound odd. If rate-altered speech sounds odd, this might lead it to be judged as "for-

Table 3
Mean Ranks for the Three Altered Feedback
and Three Control Conditions for Each Speaker
(Top Section) and Kendall's W Statistics (Bottom Section)

	Speaker					
	1	2	3	4	5	6
Condition						
NAF	2.00	2.50	2.75	2.00	2.00	2.88
Slow	2.25	2.88	3.38	3.75	3.25	4.00
Fast	3.50	2.88	3.63	4.00	2.88	5.38
AMP	3.50	2.00	2.13	2.75	3.50	2.63
FSF	4.00	4.88	4.75	4.50	4.00	2.63
DAF	5.75	5.88	4.38	4.00	5.38	3.50
Test statistic						
Kendall's W	.521	.661	.275	.250	.370	.325
Chi-square	20.857	26.429	11.000	10.000	14.786	13.000
df	5	5	5	5	5	5
Asymp. sig.	.001	.001	.051	.075	.011	.023

Note—The higher the rank, the more pronounced the accent. NAF, normal auditory feedback; Amp, amplified; FSF, frequency-shifted feedback; DAF, delayed auditory feedback.

eign," rather than the accent characteristics' doing so themselves. The rank ordering of the rate conditions shows that except for Speaker 6, rate conditions were not judged as producing a strong German accent in the speech, as does FSF and DAF. This, in turn, suggests that rate variation is not the basis of the foreign-sounding nature of the FSF and DAF conditions. Speaker 6's natural English speech is spoken fast, and the subjects reported that it was difficult to judge because of its rate of delivery. This is supported by the data in Table 3, where all the conditions except the rate conditions had close mean rankings (range, 2.63–3.50), whereas accelerating (5.38) or decelerating (4.00) rates made the speech more noticeably foreign. Further inspection of Table 3 shows that both fast and slow speaking rates are ranked higher (stronger accent) than normal speaking rates for all 6 speakers. To see whether deviation in either direction from the mean rate affects accent relative to the NAF and the AAF conditions, mean slow and fast speech rate ranks were obtained for each listener. This nonintegrated ranking was used to rerank the, now five, speech conditions. Pairwise tests showed no significant differences between the amalgamated fast and slow rate conditions and any of the remaining conditions. DAF produces a Fletcher effect (Howell, 1990). The AMP condition (although it is a form of AAF in its own right) can also serve to assess whether the Fletcher effect under DAF mediates accent judgment (this might also apply to FSF, although Howell reports that this form of feedback produces a small Lombard effect). As with the two rate conditions, AMP did not receive higher foreign accent rankings than did either DAF or FSF. Taken together with the findings of rate, neither amplitude nor rate changes associated with AAF conditions appear to mediate the perceived foreignness.

In the next analyses, the two rate conditions were dropped because they varied somewhat across speakers. The ranks of the remaining four conditions were adjusted accordingly. These four sets of ranks were analyzed in a

manner similar to all six experimental conditions (i.e., including the two rate control conditions). The data are given in Table 4 in a manner equivalent to Table 3. The rankings of the conditions then present a clearer picture. The overall pattern across speakers shows that the FSF and/or the DAF condition gets ranked higher than the amplified and NAF conditions. The one speaker who failed to produce significant agreement across assessors was Speaker 6. This is the speaker whose accent strength was highly affected when speech rate was varied, as was discussed above.

Although Table 4 shows that the judges agreed about the rank order of DAF and FSF versus AMP and NAF, for 5 of the speakers, so far no statistical analyses have been reported that show whether there is any consistency in the way the conditions were ranked across speakers. To investigate this, the data were collapsed across all speakers (including the one who showed no significant differences in the rankings given these four conditions). This created a 4 (conditions) $\times$ 48 (6 speakers $\times$ 8 listeners) repeated measures table. Kendall's W value on these data was .337, which was significant by chi-square (chi-square = 48.475, $df = 3$, $p < .001$). The significant agreement across the listeners arose because NAF and AMP were given low rankings (1.75 and 2.02, respectively, indicating that they were considered close to native speech), whereas both FSF and DAF had a more pronounced accent and were, accordingly, ranked considerably higher (2.85 and 3.38). The pairs NAF–FSF, NAF–DAF, AMP–FSF, and AMP–DAF were all significantly different (critical difference, 33.37; differences obtained, 54, 74, 43, and 63, respectively; $p < .05$). The NAF–AMP and FSF–DAF differences were not significant (11 and 20, respectively).

A similar analysis was carried out just across speakers (i.e., a 4 $\times$ 6 repeated measures table). This showed the same pattern of results (Kendall's W = .0700, chi-square = 12.6, $df = 3$, $p < .001$). Pairwise tests revealed significant differences between the condition pairs NAF–DAF and

Table 4
Mean Ranks for the Three Altered Feedback and
NAF Control Conditions for Each Speaker (Top Section)
and Kendall's W Statistics (Bottom Section)

	Speaker					
	1	2	3	4	5	6
Condition						
NAF	1.25	1.75	2.00	1.50	1.50	2.44
AMP	2.38	1.38	1.75	2.13	2.25	2.25
FSF	2.63	3.00	3.25	3.38	2.63	2.25
DAF	3.75	3.88	3.00	3.00	3.63	3.00
Test statistic						
Kendall's W	.631	.794	.331	.431	.469	.075
Chi-square	15.150	19.050	7.950	10.350	11.250	1.800
df	3	3	3	3	3	3
Asymp. sig.	.002	.001	.050	.016	.010	.615

Note—The higher the rank, the more pronounced the accent. Slow and fast rate control conditions have been dropped from this analysis, and the rankings have been adjusted to between one and four. NAF, normal auditory feedback; AMP, amplified; FSF, frequency-shifted feedback; DAF, delayed auditory feedback.

AMP-DAF (critical difference, 11.8; obtained differences, 12 and 13; $p < .05$). Bearing in mind the loss of power when judges are excluded, it may be concluded that the pattern of results is consistent across speakers.

Overall, then, there is a tendency for higher accent rankings to be given FSF and DAF. Thus, the English speech of the German speakers sounds as if it has a more marked accent when recorded under DAF and FSF, with speech under FSF being ranked as having less of an accent. Although amplification is a form of AAF, it did not have an effect on accent rankings, and it was usually ranked similar to the NAF condition.

In the introduction, the well-known findings that speaking while alterations to auditory feedback are made leads to the speech's having nonfluent characteristics was discussed. It is also possible that when speech recorded under AAF is listened to, the speech is not judged as to its accentedness, but as to its fluency. Although the interpretation that speech becomes more accented under AAF is favored, it may be that speech becomes less fluent under AAF and that listeners give a fluency, rather than an accent, judgment. Testing this is not straightforward. For instance, it is unlikely that sensible responses would be given by listeners asked to judge samples of English speakers' speech under AAF and NAF conditions by using a set of characteristics that were drawn to their attention as potential characteristics useful for identifying a German accent. For both the test German and the English speakers, a count was made of the number of pauses greater than 500 msec and false starts (phrase, word, and part word repetition) for the samples in all the speaking conditions. These were totaled for each speaker. For each speaking condition, the samples were divided into above (three samples) or below (another three samples) the mean. The samples classified in this way were used as one dimension of a 2×2 contingency table, where the second classificatory dimension was whether the speaker was German or English. Fisher-Exact tests showed that there was no association between count category and native language for any speaking condition. The category of pauses and false starts provides an indication of baseline fluency and, in the case of AAF conditions, how affected a speaker is by the particular form of AAF. The nonsignificant differences across NAF and AAF conditions show that there is no evidence that the extent to which fluency is affected under AAF conditions does not depend on the speaker's native language; otherwise, a significant association would have occurred.

The listeners also provided a ranking of accent about these speakers. Three German and 3 English speakers were judged by each listener. For each speaking condition, the data from each listener were split around the mean; a rank less than three was taken as an indication of a German accent, and a rank greater than three was taken as an indication of an English accent. The data were put in a 2×2 contingency table (one contingency was actual accent, the second was accent according to the earlier criterion). Chi-square tests (1 *df*) were significant ($p < .05$, in all

cases). This shows that the accent judgments separated English from German speakers even in the AAF conditions. Since the earlier analysis showed that speakers within each language group could not be differentiated by measuring the extent to which their fluency was affected in NAF and AAF conditions, accent judgments appear to have been independent of fluency judgments.

At the end of the additional testing, we also included a partial evaluation of what attributes had been used to judge accent. All the judges indicated that they had used the "th" and /w/ consonants frequently. Five of the 8 raters also had employed the /r/ consonant. Of the vowel characteristics, articulation of "u" (for example in "hut"—characteristic eight) was highlighted by all the raters. These, admittedly crude, indices suggest that subjects appear to have known what attributes were valid indicators about accent (revealed by the consistency with which they indicated using them and what appeared to be useful in the analysis of the training results). This occurred even though the results of what characteristics were useful for the listeners were not known to them.

DISCUSSION

The ratings given the accent characteristics in the training session can differentiate between the two groups who had either German or English as their L1. The judges were not explicitly required to use what they had learned in training during the test conditions. They were, however, reminded of the characteristics they had been taught that distinguished English with a German accent from a native English accent (see other reasons for considering that the accent characteristics were used for these decisions, given below).

The first test analyses were performed separately on each speaker so that emergence of foreign accent under AAF was judged relative to that speaker's most fluent accent—usually NAF. This procedure meets Magen's (1998) directive to perform within-speaker tests. Despite the heterogeneity in accents between the speakers, accents in the AAF conditions were judged consistently stronger than in normal feedback conditions for all the speakers. When slow and fast speech rates were examined separately or after they were combined into a composite measure, they were not significantly different from NAF or any of the AAF conditions, so rate does not appear to be a major mediator of perceived accent. In contrast, most speakers were judged to have a noticeable accent when speaking under FSF and DAF. Table 4 suggests that there is also some such effect of AMP, since this was third-most highly ranked for half of the speakers (high ranks indicate strong German accent). FSF produced a less marked accent than did DAF for 5 of the 6 speakers, which suggests that it is less disruptive than DAF to vulnerable speech systems. Note that the difference in ranking between FSF and DAF was not significant in the analysis over four conditions across speakers and judges. The reason for this may be because the Digitech equipment pro-

duces a slight and variable delay when frequency-shifting speech (mean difference between delayed and direct sound measured on 10 instances was 2.5 msec, with a standard deviation of 0.9 msec). This is probably because a sample that extends in time is needed in order to do a frequency shift using Fourier analysis. The slight delay under FSF makes it somewhat like DAF and may contribute to the nonsignificant difference. This can be checked in future work using speed-changing techniques (Howell et al., 1987) that produce negligible, nonvariable delays (on the order of 0.001 sec for a 10-kHz sampling rate). Given this proviso about the extent to which the frequency shifting condition is contaminated by delay and the fact that, despite this, FSF is consistently ranked as having less of an accent than DAF for 5 of the 6 subjects, the guarded conclusion is that FSF affects accent less than does DAF. This conclusion also fits with other data that show that FSF is less disruptive to speech production. For example, a slight Lombard effect occurs with FSF, whereas DAF results in a pronounced Fletcher effect (Howell, 1990). A further fact of note is that during these tests, half of the judges reported spontaneously that it was easier to distinguish accent strength within the same speaker than it was to distinguish between speakers with different native languages. This supports the impression that the differences in accent brought about in the experimental conditions were quite marked. All these reasons lead us to suppose that AAF conditions cause a speaker's accent to be more noticeable.

The second test examined whether the ranking judgments made about individual speakers' several speech conditions, which should have been based on accentual characteristics, were really fluency judgments. In these tests, rankings across speakers were obtained for each speaking condition. The listeners were again reminded to do this by accent. Tests like this necessitate the comparison across speakers that Magen (1998) advises against. The tests were performed like this because we did not want to require the subjects to use characteristics they had been told about as features of German accent when judging altered and nonaltered feedback samples from English speakers. It was first shown that fluency characteristics (pauses and false starts) were not associated with language groups. Results on accent rankings by the listeners used in the first test showed German speakers most accented and English speakers least accented for *all* the speaking conditions. If these judgments had been made on the basis of the fluency characteristics identified previously, disruption of fluency in AAF conditions should have affected both speaker groups and the speaker group ordering would have broken down in AAF conditions. Since this did not occur, this gives some support to the view that accent was indeed being judged. In addition, informal debriefing of the listeners indicated that they all considered the consonant characteristics most useful in deciding about accent, whereas vowels are most affected by AAF conditions, such as DAF (Fairbanks, 1955). Since these characteristics were the ones found to be most reliable in training (the listeners were not aware of this),

once again these results indicate that listeners were making use of the accent training they were given. In future work, we plan to obtain ratings for a restricted number of accent characteristics over speaking conditions based on the findings on the training session reported here. Rating characteristics need to be restricted so that the amount of time listeners spend judging the several (six) speech conditions allows a manageable procedure in terms of time for listener tests. Next, studies will be considered that suggest a likely central nervous system (CNS) locus of accent differences.

Multilingual Speakers

Studies using imaging techniques suggest that at least some distinctive accent characteristics may arise during speech execution. These studies have sought to identify what components of a task activate different CNS structures when L1 or L2 is spoken. A PET scan study by Klein et al. (1994) was designed to establish whether speakers use the same areas of the brain when formulating speech in L1 and L2 or whether different areas are reserved for use in the two languages. Different activity occurred in the left putamen when PET scans were compared between conditions involving repetition of words in L1 or in L2. Klein et al. recognized that these activation differences could either arise because this area is only used for processing one of the languages or be due to variation in facility in using different speech responses. To test between these alternatives, they presented words in L1 that had to be translated into L2 to make a response, or vice versa. When scans were compared for tasks involving the same spoken responses when the words were presented in different languages (e.g., where scans from a task in which words in L2 were translated into L1 were compared with ones where words were presented and repeated in L1), no difference in focus of activation occurred in any CNS site. The latter tasks require different cognitive-linguistic processing but the same response. Since activity in the left putamen was associated with the speaker's response, rather than with what cognitive-linguistic processing was being performed, the authors concluded that this activity was due to differences between speech responses between L2 and L1. If results from this low-level language-processing task apply to higher level language tasks as well, there would be no support for the hypothesis that a language learned later in life is subserved by different higher level neural processes from those employed when the speaker uses his or her first language.

The previous findings suggest that the response forms of L1 and L2 differ. If this is so, L1/L2 speakers would then have more complex speech response decisions than monolingual speakers, because the appropriate response from two forms has to be selected. The data of Speaker 1 are of interest for this proposal. She was exposed to English and German in roughly equal amounts from birth so, by any definition, she acquired English and German before the period that is critical to be regarded as *bilingual* (usually taken as an age less than six). Moreover, this

subject continued to speak English and German about the same amount within the home environment throughout her life and lived for long periods in countries that used each language throughout her critical period years. Consequently, she has used the requisite language for long stretches outside as well as inside the home. Informal listening to samples of her speech in ordinary listening conditions by native speakers of the respective languages showed that the two speech forms were indistinguishable from the speech of native speakers of each language. *Bilingual* was not a requirement of the experiment. Not surprisingly, therefore, none of the other speakers met the criterion of learning English as an L2 before the end of the critical period, and all had a perceptible German accent (albeit, to different degrees) when speaking English in normal listening conditions. Despite these differences between Speaker 1 and the rest, she showed a clear pattern similar to the idealized form described earlier (accent varying from most noticeable under DAF, through FSF and AMP to NAF, where accent was least noticeable; sample extracts are demonstrated on <http://www.speech.psychol.ucl.ac.uk/index.html>). Since the Klein et al. (1994) results suggest that her balanced lexical knowledge and syntactic usage in the two languages would not be processed by different parts of the brain, accent differences would not arise at these levels under AAF. The reason for this is that if there was any effect of AAF, each language form would be affected identically. Our suggestion as to why even our bilingual speaks with an accent under AAF is that all L1/L2 speakers have multiple speech output forms representing the pronunciation variants of the languages known. He or she has a larger choice of options and so makes more demands when selecting speech responses on the basis of language-appropriate phonotactic constraints (e.g., /pf/ is permitted in German, but not in English), when choosing a phone or allophonic variant that can occur in only one of the languages, and so forth. Multiple forms (L1 and L2) make response decisions more difficult to prepare than when a speaker has a single dominant L1, which is why Klein et al. found language-related activity. Next, we will consider how AAF might disrupt response processes in different ways in two forms of vulnerable speech (Borden, 1980).

Effects of Feedback on Vulnerable Speech

The present study was motivated by Borden's (1979, 1980) vulnerability hypothesis. Recall that she addressed why AAF has a big effect on speakers who have vulnerable speech systems. Her explanation for the effects of AAF on an acquired language was that the latter needs closer monitoring. Adults learning a new language use feedback, but lacking established links between feedback and the new production programs, their new speech gestures are likely to be modifications of their old systems (Borden, 1980). In this respect, alterations to feedback should have a direct effect on accent by disrupting the links being established.

The view that monitoring and correcting speech on the basis of auditory feedback occurs only during the time

when language is being acquired (as well as a limited number of other specified situations) has appeal, since it fits with introspections about language use. It offers an account of how a child matches what it hears with what it needs to say. At the same time, it explains the apparent inconsistency that intuitively it does not feel that once speech is acquired, it is listened to and corrections made on this basis. (These impressions are supported by the points Borden [1979, 1980] discusses about the rapid rate at which speech can be produced and about the fact that deaf speakers can produce speech when auditory feedback is not available [Lane, Wozniak, Matthies, Svirsky, & Perkell, 1995]). Finally, in the field of skill acquisition in general, the ideas fit with the view that automatization of a skill involves a shift from a controlled to an automatic process with fewer attentional demands (Posner, Inhoff, Friedrich, & Cohen, 1987) and that, for speech to become automatized, auditory feedback is required (Lieberman, 1984). Although Borden's proposal limits the scope of feedback monitoring, it runs into problems when attempting to account for stuttered speech control and how it is ameliorated by AAF (another type of vulnerable speech). Feedback explanations of the anomalous effect of AAF in people who stutter (i.e., their speech improves) have proposed that the altered feedback corrects some structure involved in monitoring that works defectively (e.g., Cherry & Sayers, 1956). However, empirical work has consistently failed to find such defects at the sites proposed (Howell, Marchbanks, & El-Yaniv, 1986; Howell & Powell, 1984; see, however, the recent work by De Nil, Kroll, Kapur, & Houle, in press, discussed below).

Accepting the evidence that L2 speakers have a problem in selecting an appropriate output form under AAF, an account of how the disruption to the links with speech output forms is needed, whatever the ultimate resolution of the issue of whether feedback is used for monitoring speech in circumscribed circumstances. One proposal that could be adapted to account for how disruption to these links occurs is available. Howell et al. (2000) argued that manipulations like DAF alter the situation from one task (speaking in synchrony with production) to two tasks (production in addition to hearing an asynchronous sound). This (and indeed, other) dual-task situations increase processing demands. What has to be considered is how increased processing demand can affect speech output in two different ways—leading to a disruption of fluency in children and L2 speakers acquiring language but to an improvement in fluency in speakers who stutter.

Here, we discuss two tentative possibilities. First, disruption could occur to the process that controls how the links are set up to access output forms. In the case of L1/L2 speakers, the control (attentional) mechanism that accesses the output forms could have its operation impaired by the higher number of output forms that are available, or the higher number of output forms per se could disrupt performance.

Articulation in fluent speakers is an automatic process (Posner et al., 1987) that requires few attentional resources. Articulation by nonnative speakers is relatively demand-

ing, requiring controlled attention, and is therefore more likely to show dual-task effects when paired with altered feedback (or for that matter, any demanding stimulus). In contrast, it is possible that one aspect of the speech production deficit in people who stutter may be overreliance on attentional control. One piece of evidence for this could be the heightened activity in the anterior cingulate observed in these individuals (De Nil et al., in press). Altered feedback may then serve to release people from the inhibitory effects of controlled processing, releasing the relatively automatized aspects of speech output. This explanation draws on attentional inhibition and PET results as support. One drawback is that PET scan data cannot establish whether increased activation is excitatory or inhibitory.

According to the second account, demand increases owing to the extra number of stimulus or response alternatives available. It is proposed that speakers operate at some point on a speed-accuracy continuum. In taxing circumstances (e.g., under AAF when the number of stimuli increase or when an extra response—dual task—is required), if speakers attempt to speak too rapidly, errors will ensue (Blackmer & Mitton, 1991; Howell & Sackin, 2000). Alternatively, they can avoid making errors by regulating speed to a slower rate; hence, the term *tradeoff*. Speakers can elect where to operate on the continuum, and they will use such factors as the rate of communication they want to maintain to determine their choice. It is appropriate for a child or a speaker of a foreign language to adopt a comparatively slow rate, and if they do this, they will mainly avoid dysfluency. Elsewhere, evidence has been presented that people who stutter go too fast, and in these stretches, stuttering rate is high (Howell, Au-Yeung, & Pilgrim, 1999). They are, effectively, operating at different points on the tradeoff continuum.

A speaker can choose to slow speech as much as is commensurate with maintaining communication. What happens if the speaker is overtaxed, as when speaking a less familiar second language? The speaker can adjust the rate lower up to a point, but eventually he or she will reach the limit of adjustment to a slow rate that he or she considers permissible while maintaining communication. Up to this point, fluency failure could be avoided by slowing speech, but no further rate adjustment will be possible. If a higher rate than this is maintained, it will lead to fluency breakdown in the regions where linguistic demand is high (Au-Yeung, Howell, & Pilgrim, 1998), since there is evidence that speakers attempt to produce speech before planning is complete in these circumstances (Blackmer & Mitton, 1991; Howell & Sackin, 2000). If this strategy is maintained, it would also be expected that there would be a high chance of errors persisting into later life (i.e., a heightened chance of long-term stuttering). On the basis of this account, it would be expected that age, L2 acquisition, and whether a speaker stutters or not would interact. The current results suggest that asking speakers to adjust L2 rate 20% or so around

their chosen rate appears to be achievable without breakdown in fluent speakers.

Speech rate may need to be regulated independent of linguistic processes (Howell, Au-Yeung, & Rustin, 1997; Howell et al., 2000; Howell & Sackin, 2000). Ivry (1997) has identified the lateral part of the cerebellum as a structure that does this. The timing mechanism acts as a *governor* for speech rate. *Governor* is used here in the engineering sense of a self-acting contrivance for allowing rate regulation and differs from a feedback regulator in that it is an integral part of the ongoing operation and can restrict its operation to situations in which some specified condition is in effect (e.g., pressure release in a safety valve). The critical attribute of a governor is that it is an integral part of the machine and not a device that stands separate and independently modifies operations. If the timing mechanism is general in its operation, as well as independent, it can be employed in a range of tasks that require rate control. Tasks that take place concurrently (like listening to DAF or FSF while speaking) and that both require the lateral cerebellar timing module would make high demands on the governor that would affect rate of execution of the tasks. These demands could lead to speech's being modulated at a slower rate, since it is not devoted to speech control alone. Since the cerebellar timing mechanism is independent, this rate modulation occurs irrespective of what linguistic processing is taking place (the latter is done by another module or modules). In this way, the extra time allowed because of DAF-slowness allows the linguistic planning done by these separate modules to go on unaffected and be completed in the time required, even with complex words, so the plan and the executed form correspond (Howell et al., 2000). This suggests that any secondary task requiring timing regulation will be effective in maintaining fluency as long as rate is slowed locally, as happens with FSF (Howell & Sackin, 2000). The question that remains for treatment is how to use the fluency-enhancing procedures so a speaker can learn how to maintain local control of speech rate (Reed & Howell, 2001).

REFERENCES

- AU-YEUNG, J., HOWELL, P., & PILGRIM, L. (1998). Phonological words and stuttering on function words. *Journal of the Acoustical Society of America*, **41**, 1019-1030.
- BLACKMER, E. R., & MITTON, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, **39**, 173-194.
- BORDEN, G. (1979). An interpretation of research on feedback interruption in speech. *Brain & Language*, **7**, 307-319.
- BORDEN, G. (1980). Use of feedback in establishing and developing speech. *Speech & Language*, **3**, 223-242.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- BRENNAN, E. M., & BRENNAN, J. S. (1981a). Accent scaling and language attitudes: Reactions to Mexican-American English speech. *Language & Speech*, **24**, 207-221.
- BRENNAN, E. M., & BRENNAN, J. S. (1981b). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research*, **10**, 487-501.

- CHASE, R., SUTTON, S., FIRST, D., & ZUBIN, J. (1961). A developmental study of changes in behaviour under delayed auditory feedback. *Journal of Genetic Psychology*, **99**, 101-112.
- CHERRY, C., & SAYERS, B. (1956). Experiments on the total inhibition of stammering by external controls and some clinical results. *Journal of Psychosomatic Research*, **1**, 233-246.
- DE NIL, L., KROLL, R. M., KAPUR, S., & HOULE, S. (in press). A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults. *Journal of Speech, Language & Hearing Research*.
- ELMAN, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal production. *Journal of the Acoustical Society of America*, **70**, 45-50.
- FABRO, F., & DARO, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain & Language*, **48**, 309-319.
- FAIRBANKS, G. (1955). Selected vocal effects of delayed auditory feedback. *Journal of Speech & Hearing Disorders*, **20**, 333-345.
- FLEGE, J., BOHN, O., & JANG, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, **25**, 437-470.
- FLEGE, J., & FLETCHER, K. (1992). Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America*, **91**, 370-389.
- FLEGE, J., MUNRO, M., & MACKEY, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, **97**, 3125-3134.
- GARBER, S., & MOLLER, K. (1979). The effects of feedback filtering on nasalization in normal and hypernasal speakers. *Journal of Speech & Hearing Research*, **22**, 321-333.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GIBNEY, N. (1973). Delayed auditory feedback: Changes in the volume intensity, and the delay interval as variables affecting the fluency of stutterers' speech. *British Journal of Psychology*, **64**, 55-63.
- HARDY, T. (1983). *The return of the native* (5th ed.). Harmondsworth, U.K.: Penguin.
- HOWELL, P. (1990). Changes in voice level caused by several forms of altered feedback in normal speakers and stutterers. *Language & Speech*, **33**, 325-338.
- HOWELL, P., & ARCHER, A. (1984). Susceptibility to the effects of delayed auditory feedback. *Perception & Psychophysics*, **36**, 296-302.
- HOWELL, P., AU-YEUNG, J., & PILGRIM, L. (1999). Utterance rate and linguistic properties as determinants of speech dysfluency in children who stutter. *Journal of the Acoustical Society of America*, **105**, 481-490.
- HOWELL, P., AU-YEUNG, J., & RUSTIN, L. (1997). Stuttering as an automatization deficit. In W. Hulstijn, H. F. M. Peters, & P. H. H. M. van Lieshout (Eds.), *Speech production: Motor control, brain research and fluency disorders* (pp. 573-578). Amsterdam: Elsevier.
- HOWELL, P., EL-YANIV, N., & POWELL, D. (1987). Factors affecting fluency in stutterers when speaking under altered auditory feedback. In H. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 361-369). New York: Springer-Verlag.
- HOWELL, P., MARCHBANKS, R. J., & EL-YANIV, N. (1986). Middle ear muscle activity during vocalisation in normal speakers and stutterers. *Acta Oto-Laryngologica*, **102**, 396-402.
- HOWELL, P., & POWELL, D. J. (1984). Hearing your voice through bone and air: Implications for explanations of stuttering behaviour from studies of normal speakers. *Journal of Fluency Disorders*, **9**, 247-264.
- HOWELL, P., POWELL, D. J., & KHAN, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 772-784.
- HOWELL, P., ROSEN, S., HANNIGAN, G., & RUSTIN, L. (2000). Backward masking performance in children who stutter and its relation to dysfluency rate. *Perceptual & Motor Skills*, **90**, 355-363.
- HOWELL, P., & SACKIN, S. (2000). Speech rate manipulation and its effects on fluency reversal in children who stutter. *Journal of Developmental & Physical Disabilities*, **12**, 291-315.
- IVRY, R. (1997). Cerebellar timing systems. In J. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 555-573). San Diego: Academic Press.
- KALINOWSKI, J., STUART, A., SARK, S., & ARMSON, J. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. *European Journal of Disorders of Communication*, **31**, 259-269.
- KLEIN, D., ZATORRE, R. J., MILNER, B., MEYER, E., & EVANS, A. C. (1994). Left putaminal activation when speaking a second language: Evidence from PET. *NeuroReport*, **5**, 2295-2297.
- KVAVIK, K., KATSUKI-NAKAMURA, J., SIEGEL, G., & PICK, H., JR. (1991). Delayed auditory feedback effects on learners of Japanese and Spanish. *Folia Phoniatrica*, **43**, 282-290.
- LANE, H., & TRANEL, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech & Hearing Research*, **14**, 677-709.
- LANE, H., WOZNIAK, J., MATTHIES, M., SVIRSKY, M., & PERKELL, J. (1995). Phonemic resetting versus postural adjustment in the speech of cochlear implant users: An explanation of voice-onset time. *Journal of the Acoustical Society of America*, **98**, 3096-3106.
- LEE, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, **22**, 824-826.
- LIEBERMAN, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.
- LIKERT, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology* (Whole No. 140).
- MACKEY, D. (1968). Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *Journal of the Acoustical Society of America*, **43**, 811-821.
- MACKEY, D. (1970). How does language familiarity influence stuttering under delayed auditory feedback? *Perceptual & Motor Skills*, **30**, 655-669.
- MAGEN, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, **26**, 381-400.
- MUNRO, M. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language & Speech*, **36**, 39-66.
- POSNER, M. I., INHOFF, A. W., FRIEDRICH, F. J., & COHEN, A. (1987). Isolating attentional systems: A cognitive-anatomical analysis. *Psychobiology*, **15**, 107-121.
- REED, P., & HOWELL, P. (2001). Presentation of frequency-shifted feedback to sustain long-term improvements in fluency in people who stutter. *European Journal of Analysis of Behavior*, **1**, 89-106.
- ROUSE, R., & TUCKER, G. (1966). An effect of delayed auditory feedback on speech in American and foreign students. *Journal of Speech & Hearing Research*, **9**, 456-460.
- RYAN, E. B., & CARRANZA, M. A. (1975). Evaluating reactions of adolescents towards speakers of standard English and Mexican American accented English. *Journal of Personality & Social Psychology*, **31**, 855-863.
- SIEGEL, G., FEHST, C., GARBER, S., & PICK, H., JR. (1980). Delayed auditory feedback and children. *Journal of Speech & Hearing Research*, **23**, 802-813.
- SIEGEL, G., PICK, H., JR., & GARBER, S. (1984). Auditory feedback and speech development. *Advances in Child Development & Behaviour*, **18**, 49-79.
- SODERBERG, G. (1969). Delayed auditory feedback and the speech of stutterers: A review of studies. *Journal of Speech & Hearing Research*, **34**, 20-29.
- STUART, A., KALINOWSKI, J., ARMSON, J., STENSTROM, R., & JONES, K. (1996). Stuttering reduction under frequency-altered feedback of plus and minus one-half and one-quarter octaves at two speech rates. *Journal of Speech & Hearing Research*, **39**, 396-401.
- THOMPSON, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, **41**, 177-204.
- WEBSTER, R., SCHUMACHER, S., & LUBKER, B. (1970). Changes in stuttering frequency as a function of various intervals of delayed auditory feedback. *Journal of Abnormal Psychology*, **75**, 45-49.