Tolerable Hearing Aid Delays. II. Estimation of Limits Imposed During Speech Production

Michael A. Stone and Brian C. J. Moore

Objective: **We used real-time processing in a wearable digital hearing aid to examine the effect of processing delay on normal-hearing participants while speaking. Objective and subjective data were recorded so as to permit analysis of both the production and perception of speech read aloud from a script. We also asked participants to rate the disturbance of the echo introduced by the delay.**

Design: **Thirty-two (16M, 16F) participants were fitted binaurally with behind-the-ear (BTE) aids connected to a digital processor. A 4 mm Libby horn, surrounded by an expanding foam earplug, conducted processed sound into each ear canal. The processor provided either linear processing or three-channel, fast-acting wide dynamic range compression, independently to each ear. Insertion gains were set, using a KEMAR manikin, to be 0 dB over a wide frequency range, for frontally presented speech with a free field level of 65 dB SPL. Additionally, the aids introduced one of four selectable delays (7 to 43 msec) between the BTE microphone and receiver. After a short period of acclimatization, each participant read 16 prose passages of about 500 words in length in each of two similar-sized rooms with markedly different acoustics: reverberant and nonreverberant. For each passage, a subjective rating of the level of disturbance of the perceived echo was recorded, as well as simultaneous recordings from a microphone and a Laryngograph, which directly records glottal pulses.**

Results: **Disturbance ratings generally increased monotonically with increasing delay. Averaged results show that a delay between 25 and 30 msec is rated as "disturbing." Measures were also taken of word production rate, speech level and range of level as well as fundamental frequency and range of fundamental frequency. For these measures of speech production, there was no significant effect until the delay exceeded 30 msec. There was little effect of acoustic environment or aid processing (linear or compression).**

Conclusions: **The acceptability of delays introduced by digital hearing aids is primarily determined by aspects of the perception of self-generated speech. Speech production, on average, is hardly affected unless the processing delay exceeds 30 msec. The permissible limit of 20 to 30 msec is smaller than the delays at which audio-visual integration is disrupted.**

(Ear & Hearing 2002;23;325–338)

Department of Experimental Psychology, University of Cambridge, Cambridge, England.

DOI: 10.1097/01.AUD.0000027431.38251.D0

In the Introduction to Stone and Moore (1999) we described in detail how there are two principal paths by which one hears one's own voice when speaking. The first is a "direct" path via solid structures of the head to the meatus and then through the middle ear to the cochlea. The second is an "indirect" path via air conduction, sound being radiated from the lips to the auditory meatus and then though the middle ear to the cochlea. The sum of the signals from the two paths drives the cochlea, and leads to the perception of sound. The addition of a signal to a delayed version of itself is equivalent to "comb-filtering"; the spectrum acquires peaks separated by notches or nulls. The comb filtering gives rise to a certain timbre. For a nonoccluded ear, the delay between the two paths is about 0.7 msec (Stromsta, Reference Note 1), so we normally hear our own voice with some degree of comb filtering. The resulting timbre is taken as our "normal" voice, which is why recordings of our own voice, which are only of the airborne sound, appear different to us when we listen to them.

When a hearing aid is inserted into the meatus, the obstruction reduces the level of the unamplified sound carried via the air conduction path. The total air-conducted sound then comes from any leakage around the aid or mold (and through any vent) and from the hearing aid, which introduces both amplification and a delay. The perceived timbre of the user's voice is then dependent on: (1) the "occlusion effect" produced by the meatal obstruction (Killion, Wilber, & Gudmundsen, 1988; von Békésy, 1960; Zwislocki, 1953); (2) the frequency-dependent amplification used in the aid; (3) the delay introduced by the aid; (4) the amount of leakage though the sound delivery system in the meatus; and (5) the user's hearing loss. Factors (2) and (5) both reduce the relative contribution from the direct bone-conduction path.

Current analog hearing aids rarely introduce delays greater than about 1 msec. Such short delays are unlikely to cause problems except possibly for sound localization when a user is fitted unilaterally.

In contrast, aids incorporating digital processing in the signal path introduce longer delays. The magnitudes of these delays depend on the nature of the algorithm(s) used in the digital processor, ranging from about 1 to about 10 msec in current commercial hearing aids, but potentially being longer in the future. As the magnitude of the delay changes, so does the perceived effect: with increasing delay, the percept changes from coloration (timbre), to a noticeable echo.

Alterations in the level and delay of the air-conduction path can have many effects on the aid user. These include: (1) disruption of voice production, which may be produced both by conflicts between the auditory and proprioceptive systems and by disturbance of auditory self-monitoring; (2) disruption of audio-visual communication (speech reading); and (3) subjective disturbance when the aid user talks. The aspect that first becomes disturbing will determine the maximum acceptable delay, as delay is increased.

Our previous report (Stone & Moore 1999), concentrated solely on effect (3) above. In that experiment, we asked normal-hearing listeners to rate the quality of speech samples processed as if they were wearing a hearing aid and had a hearing loss. The source materials were in-ear and above-ear recordings of two males reading from a script. Off-line processing simulated both a high-quality hearing aid, and the recruitment aspects of a hearing loss. The parameters varied were the magnitude of the hearing aid delay and the degree of simulated hearing loss. Because we used off-line processing, it was not possible for the listener to produce and monitor his/her own speech. The delay was rated as "disturbing" when it reached 20 ms.

Agnew and Thornton (2000) also studied the subjective effects of delay. Their participants listened to their own voices through a hearing aid that introduced a variable delay. Participants adjusted the gain of the aid to place the sound at their most comfortable level; this level was not measured. The participants varied the length of the delay and determined the amounts that were noticeable and objectionable when compared with no delay. The mean time delay which was judged to be objectionable was 14 msec. This limit is somewhat lower than the limit found by Stone and Moore (1999). The modest difference may be due to several factors: (1) the participants of Agnew & Thornton could directly compare conditions with and without delay, which might increase the salience of the delayed sound; (2) their participants were engineers working for a hearing aid company, who may be more critical listeners than is typical; (3) Stone & Moore (1999) used a simulation of hearing loss, whereas Agnew and Thornton (2000) did not.

In the present experiment, we extended our earlier work to concentrate more on the effect of delay on speech production. The effect of delaying the acoustic feedback of sound, and more specifically, speech, to the producer's ear was extensively studied in the 1950s under the title "Delayed Auditory Feedback" (DAF) (e.g., Black, 1951; Butler & Galloway, 1957; Fairbanks, 1955; Lee, 1950; and a summary in Yates, 1963). These researchers found various effects as the delay between source and ear was increased, such as:

(1) Increase in fundamental frequency, *f0* (Fairbanks, 1955)
(2) Increase in production level (Black, 1951; Fairbanks, 1955; Spilka, 1954)
(3) Decrease in production rate (Fairbanks, 1955) and
(4) Disfluency of various forms such as increased speech errors (Butler & Galloway, 1957).

However, these effects did not occur for all participants. The magnitude of some of these effects increased with increasing delay up to a delay of about 200 msec, and thereafter reached an asymptote or decreased. However, Black (1951) reported that with increasing delay there was an increase in word production rate accompanied by more stumblings, leading to little or no change in overall rate. Bachrach (1964) investigated the effect of delays between 0.2 and 0.8 sec, and reported that, with increasing delay, men speak more slowly and decrease their production intensity, while women speak more quickly and increase their intensity.

In these earlier studies, experimental conditions were not realistic (and nowadays, would be unethical) for a hearing aid application, especially for mild to moderate hearing losses. Typically, elevated sound pressure levels were used, ranging from 85 to 90 dB SPL (Black, 1951; Fairbanks, 1955) to over 110 dB SPL (Spilka, 1954). The intention in these experiments was to mask any residual air- and bone-conducted sound, so as to investigate the effect of the delayed sound alone on speech production. Tiffany and Hanley (1952) reported that the effect of DAF on speech production increased with increasing replay level. Similarly, Butler and Galloway (1957) reported increasing error rates with increasing replay level. Also, as most investigators used delays giving large effects, there appears to have been little investigation of the delay range 0 to 40 msec, which is more relevant for a hearing aid application; the study of Black (1951) is an exception.

MacKay (1968) found that the effect of DAF on inducing stuttering showed a "breakpoint" around

100 msec: below this there was little stuttering, and above it stuttering increased markedly. Black (1951) measured speech production rates under DAF with time delays varying in 30 msec steps from 0 to 300 msec. The greatest increment in interference occurred between delays of 30 and 60 msec. In slight contradiction, figures in Lee (1950) and, more recently, Howell and Powell (1987), imply that the reading rate between 0 msec and their first recorded value of delay decreased linearly as delay was increased. Overall, the data in the literature do not provide any clear indication of the likely effect on speech production of hearing aid-produced delays in the range 0 to 40 ms. However, the data do lead to the expectation that there will be some disruption of speech production, perhaps with a slowing of speech rate.

Further factors which could affect speech production when wearing hearing aids arise from:

(1) The acoustic environment: the amount of reverberation may affect the perception of hearing aid-induced delay. Black (1950) investigated speaking rates (without DAF) in rooms of varying volume, shape and reverberation time, and found that people read more slowly in large rooms with a "live" acoustic, than in small rooms with a "dead" acoustic. Also, he found a significant increase in reading rate over about the first 15 sec of reading as the participant became acclimatized to the acoustic environment.

(2) The characteristics of the hearing aid: the timbre of the perceived sound will be dependent on frequency response and the form of any automatic gain control. Fast-acting compression amplifies low-level portions of the input, reducing brief dips in energy, which may be perceived as an effect similar to that of reverberation. Howell and Powell (1987) presented evidence suggesting that speech production rates were linked to the time interval between the produced sounds and the perceived ending of the delayed feedback linked to those sounds. Fast-acting automatic gain control would increase the audibility of decaying portions of speech sounds, delaying the perceived endings, and this might lead to a slowing of speech production.

The object of the present experiment was to obtain both objective and subjective measures of the effect of hearing aid-induced delay on speech production and perception; we aimed to define an upper limit to the permissible processing delay. Parameters varied were the magnitude of the delay, the acoustic environment and the hearing aid processing.

TABLE 1. Parameters for the compression processing within the digital aids.

Channel	Bandwidth	Attack, Release Times (msec)	Free Field Compression Threshold (dB SPL)
1	DC to 1220 Hz	10, 60	32
2	1220–2700	5, 40	32
3	2700–6500	2, 20	32

Materials and Methods

Equipment

The experimental aids were "Audallions" made by Audiologic. Each aid consisted of two behind-the-ear earpieces, connected by wires to a small chest-worn processor, containing a Motorola DSP56009 DSP chip. The signals picked up by the microphones in the earpieces were fed to the analog-to-digital converters (ADCs, 13-bit resolution clocked at 15.625 kHz) in the chest-worn part. Analog circuitry within the aids performed high-frequency emphasis prior to the ADC, to flatten the speech spectrum and give better coding of the higher-frequency portions of the input. The ADC performance allowed useful amplification to be applied up to at least 6.5 kHz and gave resolution below the noise floor introduced by the microphone. The earpieces contained analog compression limiters with a compression threshold (at 500 Hz) of 90 to 95 dB SPL (input level). These limiters prevented overloading of the ADCs. To avoid activation of the limiters, some participants with intense vocalizations were rejected. After digital signal processing, the internal digital-to-analog converter (DAC, 16-bit resolution) was used to produce an analog voltage which was fed back to the amplifier and receiver in each earpiece. In this experiment, the units were initially set to give their near-minimum processing delay of 7 msec: a programmable delay line was then used to provide extra delay, giving total delays of 7, 18, 30, or 43 msec. All processing was performed in the frequency domain using the overlap-add method (Allen, 1977). This used forward and inverse fast-Fourier transforms (FFT) of segments of 32 time samples, overlapped by 50%. Two units were used during the experiment, one configured for linear processing, and one configured for three channels of fast-acting wide dynamic range compression, each channel having a compression ratio of two. Parameters for the compression processing are shown in Table 1.

The compression processing employed by Stone and Moore (1999) used symmetric attack and release times. In order to prevent large overshoots in output level at the onset of sounds, the audio signal passed through a compensating delay before the gain signal was applied. In this real-time implementation, such a system would mean that the minimum delay testable would be around 12 msec. So as to be able to test a sub-10 msec overall delay, the compression processing employed in this experiment used a first-order exponential attack and decay weighting of the channel magnitude. Fast attack times reduced the overshoot at signal onsets. With only a first-order weighting of the channel signal, fast attack times require longer release times, otherwise the gain control signal fluctuates at a rate corresponding to the fundamental frequency, which would give poor sound quality.

Equalization was calculated to convert the non-flat frequency response of the microphone in a BTE position to that corresponding to an equivalent free field representation. The equalization was based on measurements obtained by fitting the aid and earpiece on a KEMAR manikin. A separate equalization was calculated to correct the receiver response so that it produced a flat insertion response on the KEMAR. For both equalizations, the short analysis time window produced moderate constraints on the accuracy of the required equalization, and also the choice of bandwidth for channel 1 in the compression aid. Combined with the processing unit, each BTE aid provided 0 dB insertion gain over a wide frequency range for frontally presented, free field speech-shaped noise (Byrne et al., 1994) with a root-mean-square (RMS) level of 65 dB SPL. The overall effect of the equalization was to produce "normal" loudness for a normal-hearing listener. An identical rationale is found in some hearing aid gain prescription formulae for hearing-impaired listeners (loudness restoration), e.g., Moore (2000).

Binaural fittings were used throughout the experiment, and the processing was independent for each ear. Delivery to the ear canal was via 42 mm tube-length "Libby" horns housed in unvented expanding foam earplugs (available from GNReSound). Prior to insertion of the earpieces, we used a light film of petroleum jelly and precompression tubes to achieve a tight roll-up of the foam. It was then usually possible to achieve deep insertions to beyond the second bend in the meatus. This served two purposes: it virtually eliminated the leakage path around the earpiece and it reduced the occlusion effect (a possible source of confusion with the processor-added echo). The reduction of the occlusion effect was desirable to make the listening experience more like that of a hearing-impaired aid user. The perception of the occlusion effect is much reduced with increasing low-frequency hearing loss.

The attenuation provided by the earplug/tubing was measured using free field audiometry, first with the ear canal open and then with an inactive aid fitted to the ear. The testing was done using one ear only, the other ear being plugged throughout with a deeply fitting EAR® expanding foam plug. Four participants were used. The mean attenuation provided by the earplug/tubing was 12 dB at 64 Hz, 15 dB at 128 Hz and 18 dB at 256 Hz.

Each processing unit could store four programs, selected by a front-panel push-button, and confirmed in the listener's ear by a spoken number recorded in the Audallion. These four programs were used to vary the overall delay. To partially "blind" the participant, there was a permuted relation between the program number and the magnitude of the delay. This relation was the same for each participant. Different permutations were used on the linear-processing unit and on the compression-processing unit.

Participants and Method

Thirty-two participants (16M, 16F, aged 18 to 53 yr) were selected with pure-tone thresholds not greater than 15 dB HL at the standard audiometric frequencies between 125 and 8000 Hz (except for one participant, for whom the threshold was 20 dB HL at 8 kHz).

Participants attended two sessions, spaced about a week apart. In each session, they read two short acclimatization passages (see below for details) and then 16 separate passages selected from Lewis Carroll's books "Alice's Adventures in Wonderland" and "Through the Looking Glass." Each of the main passages was edited to be approximately 500 words in length and featured only narrative and monologue by "Alice." Each passage was printed in a large font on A4 paper and held vertically in a stand, between 60 and 90 centimeters in front of the participant, just below the horizontal from the participant's eye level. Participants were asked to read in a fluid, conversational style, preferably not dramatic. They were instructed to correct misspoken words, but not omissions or substitutions, provided that fluency of production was maintained. Drinking water was freely available.

Simultaneous two-channel recordings of each reading of the prose were made to digital audio tape (DAT) using either a Sony DTC 55-ES, or a Sony DTC1000-ES machine. One channel of the tape carried a Laryngograph signal. The Laryngograph is an "electroglottograph" which measures the changing resistance across the glottis during phonation and provides a clean signal related to vocal-fold activity, which made possible a more reliable extraction of fundamental frequency statistics, uncontaminated by room acoustics and background noise (Fourcin & Abberton, 1972). The other channel carried the A-weighted output of the omnidirectional microphone of a sound level meter (Realistic, Radio Shack, catalog number 42–3019). The microphone was placed at a fixed distance from the lips of the speaker (see later for details). This distance was maintained throughout the experiment using a knotted string attached below the microphone.

The acclimatization passages were about 60 sec of narrative prose ("Alice" and "Arthur the Rat," both supplied by the Department of Phonetics and Linguistics, University College, London). The first passage was read with the Laryngograph electrodes attached via a neckband but with the ears unoccluded. The foam earplugs were then inserted and allowed to expand. A processing unit was attached to the BTE units, and the BTEs attached to the earplug tubing. The participant was instructed to listen out for a single distinct echo and to provide a rating as to the magnitude of its disturbance. They were asked not to confuse the echo with room-derived reverberation. This task is potentially confusing because the sound produced by the aid is of a bright timbre, but arrives after the bone-conducted sound, which is of a duller timbre (Stone & Moore, 1999). This is the opposite of our everyday experience, where echoes are duller in timbre and follow in time the direct sound arriving at our ears. The range of the percept was demonstrated by asking the participant to select first the minimum delay and to read a few sentences of a dummy passage. The maximum delay was then selected and again a few sentences from the same dummy passage were read. This minimum-maximum comparison was repeated until it was clear that the participant understood the task. Occasionally, for some participants whose session started with the compression processing, it was necessary to temporarily use the linear processing unit to accentuate the effect of changing delay in this training period. Generally, the participant was only "unblinded" by knowing the two numbers of the settings for the minimum and maximum delay on the first processing unit in the session. The other processing unit, swapped to halfway through the session, had a different permutation between delay and setting number. Of the eight possible settings between the two units, the participant was "blind" to six of them (unless they remembered settings from the previous session, which was another reason for the delay between sessions). The second acclimatization passage was then read at the maximum delay setting.

Each session took place in one of two different acoustic environments, of similar volume, but different reverberation times:

(1) A small, carpeted, foam-lined double-walled booth (3.5 by 2.7 by 2.4 meters). Reverberation time (RT_{60}) was fairly constant at 50 msec, across the frequency span 0.25 to 4 kHz. In the octave around 0.125 kHz, it rose to 140 msec. Apart from the microphone and script stands, there was no other equipment in front of the participant, who sat at least 1 meter from the wall. This was done in order to keep early reflections to a very low level and thereby enhance perception of the processing delay. The fixed distance for the recording microphone was 1 meter.

(2) A small uncarpeted office (3.4 by 2.3 by 3.1 meters). Reverberation time (RT_{60}) was fairly constant around 390 msec, across the frequency span 0.1 to 4 kHz. This was the same room as used in Stone and Moore (1999), except that the double-component decay previously reported at 100 Hz had been eliminated. As well as the microphone and script stands, there were other pieces of rigid-faced equipment in front of and to the side of the participant, within 1 meter of the participant. Early reflections were therefore increased, which might be expected to reduce the perceptibility of the processing delay. The fixed distance for the recording microphone was 0.6 meters.

Background noise levels were <20 and <35 dBA in the booth and office environments respectively, except for some occasional low-frequency ventilation noise in the booth and structure-borne vibration in the office. Typically, these levels were less than 40 dB SPL, with virtually all the energy below 100 Hz. Although usually inaudible, the noise produced some complications in the analysis of the recordings, which will be referred to later.

Participants were tested in a counterbalanced order of room type, aid processing and delay magnitude. Each delay setting was tested twice for each aid in each environment. So, after the acclimatization passages and eight passages had been read, the processing unit was swapped (from linear to compression or vice-versa), without disturbing the earplugs.

From a reference word in each passage, the experimenter recorded the word count achieved through the passage at the end of 2 minutes. The reference point was usually achieved about 10 to 15 sec after reading commenced, about 40 words into the passage (the mean reading rate across the 32 participants was later found to be around 200 words per minute). The recording was stopped either when the participant finished reading the passage, or after about 3 minutes, so as to prevent fatigue. At the end of each passage, participants used a rating scale to judge the disturbance of the echo heard during the reading. The rating scale was represented by 1 for "Not at all disturbing," through 4 for "disturbing," and up to 7 for "highly disturbing." No labels were specified for intermediate points.

Data Processing

The experiment produced 1024 subjective ratings and 1024 recordings, corresponding to 32 participants, two acoustic environments, two types of processing, four degrees of delay, and two samples per condition.

The recordings were transferred as stereo analog from DAT to a PC soundcard using a 16-kHz sampling rate, taking care to match the tape dynamic range to that of the PC soundcard. A MATLAB script was used to process the microphone recordings. A high-pass finite-impulse-response (FIR) filter was used to reduce fan hum and low-frequency environmental noise by at least 40 dB. The corner frequency was chosen to be just below the lowest fundamental frequency used by the participant. For some of the male participants this required filters with a total impulse-response length of around 170 msec. The reverberation audible in the filtered recording was then a composite of that produced by the room reverberation and the effect of the filter impulse response. This did not seriously compromise the measures of the distribution of speech levels since the magnitude of the filter impulse response had decayed by 60 dB within about 20 msec on either side of the peak of that same impulse response. The envelope of this "clean" signal was extracted, and the RMS value calculated in successive, nonoverlapped, rectangular-shaped, 10 msec windows. A histogram was formed of the RMS levels for all the windows using bin widths of 0.4 to 0.5 dB. The choice of around 0.5 dB was a compromise between smoothing the histogram, thereby distorting its shape, and too much detail producing false features on the underlying shape. Typically, the histogram exhibited a main peak with a secondary peak at a lower energy level. The secondary peak was typically produced by participant-generated background noise (clothing and chair movements) and microphone noise, and was less related to the speech. A threshold was chosen automatically at a level just above the valley between the two peaks, segmenting the levels into nonspeech or speech categories. Typically, this produced a rejection of 25% of the time windows. Two parameters were then extracted from the "speech" levels:

(1) The mean level of the speech.
(2) The mean and standard deviation of the distribution of speech levels.

The parameters in (2) were very sensitive to variations in the threshold chosen, so the 16 histograms from each participant in each environment were previewed to select an average threshold: the automatic-thresholding procedure could then be manually steered to the nearest bin in the histogram to the average threshold. Typically, the standard deviation of the thresholds chosen was less than 0.2 dB across the 16 passages produced by a single participant.

The Laryngograph traces were resampled from 16 to 32 kHz (in the digital domain) so as to provide better time resolution, especially for female voices. Using the tool "HQTx" in "Speech-Filing-System" (SFS ver 4.11, Sept 2000, available from *http://www.phon.ucl.ac.uk/resource/sfs*), the fundamental periods were extracted. From a histogram of the distribution of these periods, the software calculated the median *f0*, as well as the 10% and 90% points of the distribution of *f0* intervals. Occasionally, the software could not produce valid estimates, so the fundamental periods were estimated from the microphone recording using the tool "*pp*": this utility was employed on the high-pass filtered speech signal, again resampled to 32 kHz.

In summary, for each spoken passage, six measures were derived:

(1) Subjective rating.
(2) Word production rate.
(3) Mean speech level.
(4) Standard deviation of the thresholded histograms of speech level.
(5) Median *f0*.
(6) Range of *f0* between the 10% and 90% points of the distribution.

RESULTS

Subjective Ratings

The ratings are plotted in Figure 1 as means $\pm$1 SD. Long- and short-dashed lines indicate the results for linear processing and multi-band compression, respectively. The filled and open circles show mean scores for the 16 males and the 16 females, respectively. Since the results are shown separately for each processing type, there are two symbols for each gender group at each delay. The symbols have been displaced from the true delay values for clarity. This displacement hides the fact that the females gave higher average ratings than the males, by between 1/3 and 1/2 scale units.

A within-subjects analysis of variance (ANOVA)

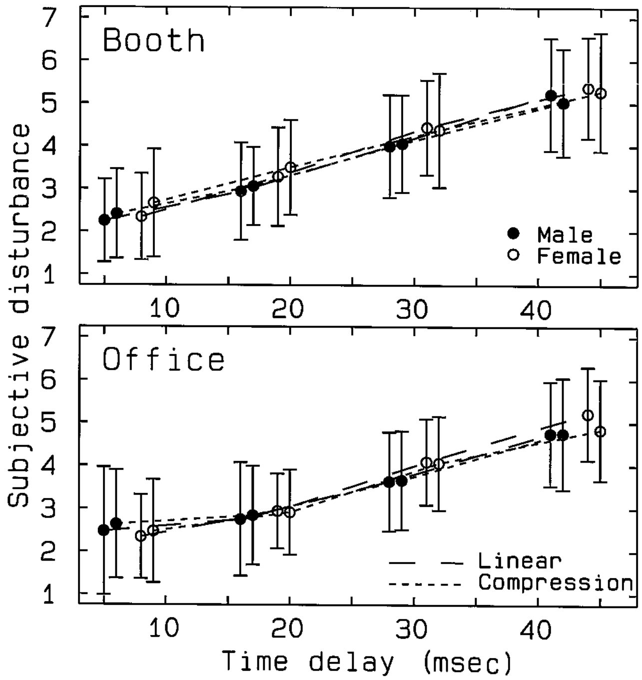

Figure 1. Mean subjective disturbance for each acoustic environment, plotted as a function of delay. The rating scale was represented by 1 for "Not at all disturbing," through 4 for "disturbing," and up to 7 for "highly disturbing." Error bars indicate ±1 SD across participants. Long- and short-dashed lines show results for linear and compression processing, respectively. Filled and open circles show results for male and female participants, respectively.

with factors acoustic environment, processing (linear or compression), gender and delay was performed on the subjective ratings. There was a significant effect of delay, $F(3,90) = 156.8$, $p < 0.001$. Rated disturbance increased monotonically with increasing delay in both environments: the average rate of change across both environments was 1 scale unit per delay setting, except between the two lowest delay settings, where the average change was 0.6 scale units.

There was a significant interaction of acoustic environment with delay: $F(3,90) = 3.72$, $p = 0.014$. The mean ratings were 0.34 scale units higher (worse) in the booth than in the office at the three higher values of delay. A two-tailed *t*-test showed these differences to be significant at $p < 0.05$. At the 7 msec delay, the difference of 0.06 scale units between the two environments was not significant ($t = 0.4$, 127 df).

In both acoustic environments, participants tended to give higher disturbance ratings to the compression processing than the linear processing at low delays, but lower ratings at the longer delays. However, the interaction of processing and delay just failed to reach significance: $F(3,90) = 2.62$, $p = 0.056$. There was no significant main effect of gender and no interaction of gender with other factors

Separate within-subjects ANOVAs were performed on the male and the female results, with factors acoustic environment, processing (linear or compression) and delay. For both gender groups, there was a significant effect of delay: $F(3,45) = 71.7$, $p < 0.001$, for the males and $F(3,45) = 86.0$, $p < 0.001$, for the females. The interaction of acoustic environment and delay was significant only for the male results, $F(3,45) = 4.78$, $p = 0.006$.

Between 18 and 43 msec, the slope was about 0.8 scale units per 10 msec of extra delay, nearly independent of processing or gender. Since the mean difference in ratings between the two environments was 0.34 scale units, we can predict that, for equal levels of subjective disturbance, the delay could be 4.2 msec longer in the office than in the booth.

It should be noted that the data in Figure 1 do not indicate that there is a "threshold" above which the delay becomes "disturbing." Rather the disturbance ratings increase progressively with increasing delay. Most participants required a delay exceeding 20 msec to give the intermediate rating of "disturbing," but for a few participants this rating occurred for the smallest delay of 7 msec.

Word Production Rates

The maximum effect of delay on word production rates was around 5%, but the variation of mean word rate across the 16 text passages used in each session for a given participant was nearly +5/−7%. So as to reduce across-list differences and provide a clearer picture of the effects of delay on word rate, the word rates were normalized for each passage according to the acoustic environment in which it was recorded. The normalization was performed by dividing by the average of the eight results obtained in both the linear and compression conditions at the minimum delay within that environment. The eight results are obtained from two participants, by two genders by two processing conditions. In the normalization, the relative rate between the two environments was preserved (this was only 0.7% different at the minimum delay). The effects of delay on the normalized speech production rates are shown in Figure 2. Again, long-dashed lines indicate the results for linear processing and dashed lines the results for fast compression. The solid circles show the results for males, and the open circles those for females. However, for this and subsequent plots, the results are given as means ±1 SE.

A within-subjects ANOVA with factors acoustic environment, processing, gender and delay was performed on the normalized word rates. There was a

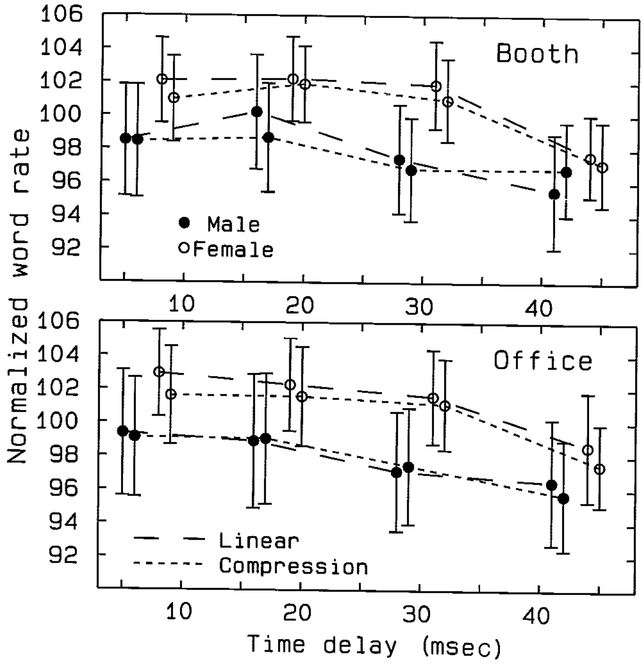

Figure 2. Normalized word production rates for each acoustic environment, plotted as a function of delay. Error bars indicate ±1 SE across participants. Otherwise, as Figure 1.

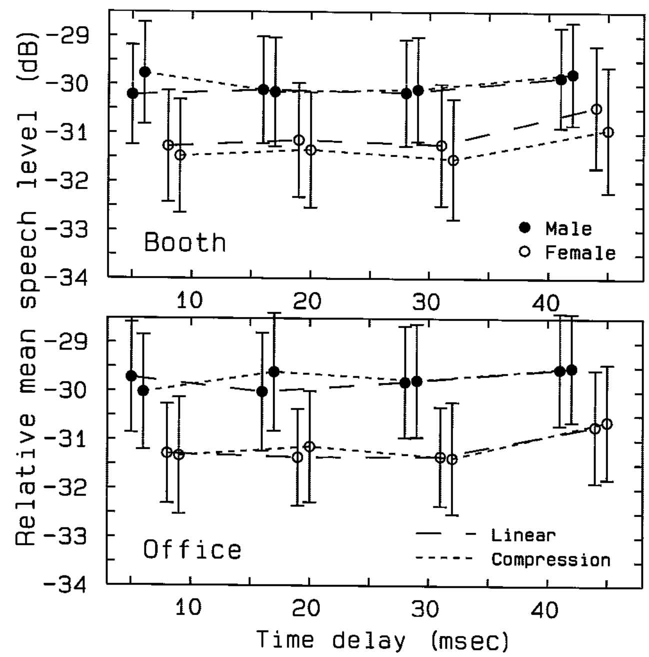

Figure 3. Mean speech level produced in each acoustic environment relative to digital maximum in a 10-msec segment, plotted as a function of delay. Otherwise, as Fig. 2.

significant effect of delay, $F(3,90) = 14.9$, $p < 0.001$. Mean word production rate decreased by 3% as the delay increased from 7 to 43 msec. However, the decline in word rate with increasing delay was only strongly significant when the delay increased from any of 7, 18 or 30 msec to 43 msec (two-tailed t-test, $t \geq 3.84$, 127 df, $p < 0.01$). There was a marginally significant difference ($p \approx 0.05$) between 18 and 30 msec, but not between 7 and 30 msec.

A separate ANOVA of the results from females, with factors acoustic environment, processing and delay, produced two significant effects. There was a significant effect of delay, $F(3,45)=10.1$, $p < 0.001$, and a significant effect of processing $F(1,15) = 4.57$, $p = 0.049$. Females spoke 0.8% faster, on average, with linear processing than with compression. A similar separate ANOVA of the male results again produced a significant effect of delay, $F(3,45)=5.9$, $p=0.002$. However, for males, there was no significant effect of processing: $F(1,15) = 0.16$, $p = 0.70$.

General Observations on the Measures of Speech Level and Fundamental Frequency

ANOVAs of both the speech and Laryngograph data produced several significant effects, but of small magnitude. Effects which were significant in the data across all participants were usually more significant in one gender group, but insignificant in the other gender group. The effect of the factors acoustic environment and processing type only appeared in measures of *f0*.

For the measures of speech level, one would expect to see a significant effect of acoustic environment due primarily to the difference in absolute levels on the tapes. The sound level recorded in the office was higher in level because of (a) the closeness of the microphone to the speaker, and (b) the recorded signal being a combination of both direct and reverberant sound. Theoretically, knowing the RT_{60}, it is possible to estimate the direct sound level. In this experiment, the absolute difference in production levels between the two environments was not the main interest: we were interested in the interaction with delay and possibly processing. Additionally, reverberation makes differences in the shapes of speech-level histograms inevitable. Consequently, the effect of acoustic environment, other than as an interaction effect, will be ignored in the summaries. For the measures of *f0*, there were effects of room acoustic which cannot be attributed to the recording technique, and so these will be reported.

Mean Level

Figure 3 shows the RMS level relative to digital maximum of the nonoverlapping 10-msec segments of speech. These segments were selected as exceeding the threshold (defined earlier) during the 2-minute timed segments of each passage. Long- and short-dashed lines show results for linear and compression processing, respectively. Filled and open circles show results for males and females,

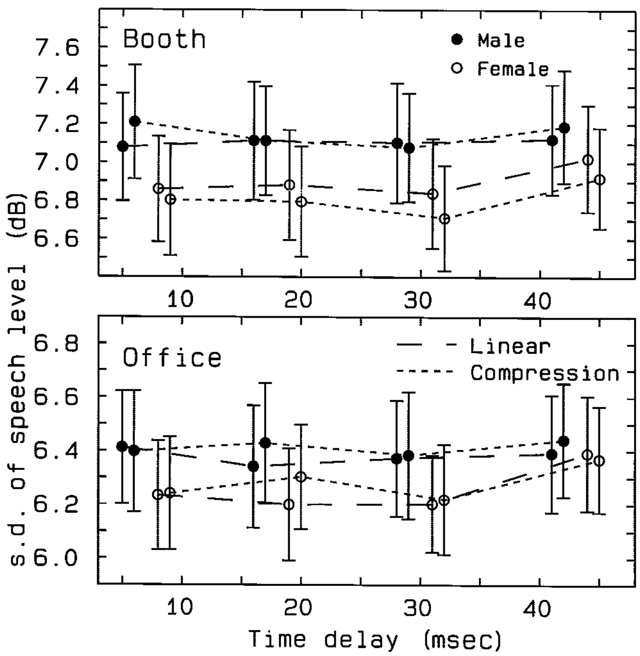

Figure 4. Standard deviation of the distribution of the level in 10-msec sections of speech produced in each acoustic environment, plotted as a function of delay. Otherwise, as Fig. 2.

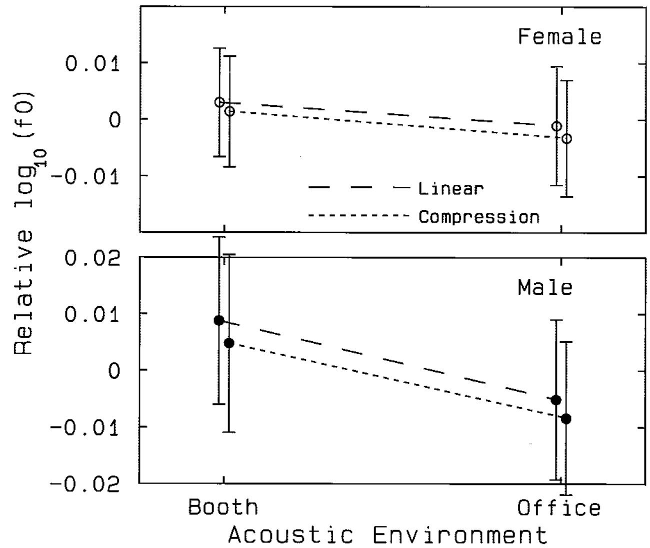

Figure 5. The effect of acoustic environment and aid processing on the fundamental frequency of the speech produced by either female participants (open circles, top panel), or male participants (filled circles, bottom panel). Data are collapsed across delay times. Error bars indicate ±1 SE across participants. Long-dashed lines show the results for linear processing and short-dashed lines show results for compression processing.

respectively. The effects of delay were small. A within-subjects ANOVA, with factors acoustic environment, processing, gender and delay, performed on the mean speech levels gave a significant effect of delay, $F(3,90) = 4.79$, $p = 0.004$. The average level was almost the same for the smallest three delays but increased by 0.4 dB for the 43-msec delay. There were no significant effects of gender.

Separate ANOVAs of the results for males and females, with factors acoustic environment, processing and delay, showed that the effect of delay was significant for the females: $F(3,45) = 7.1$, $p < 0.001$, and insignificant for the males. For females, the average level was almost the same for the smallest three delays but increased by 0.64 dB for the 43-msec delay.

The ANOVA of the male results showed a weak interaction of acoustic environment, processing and delay, $F(3,45) = 2.81$, $p = 0.05$. Between the 7- and 18-msec delays, there were very small and insignificant changes in the speech level under linear processing, but under compression processing, the speech level decreased in the booth by 0.4 dB, but increased in the office by 0.4 dB. This result will be touched on in the discussion.

Standard Deviation of Histograms of Speech Level

Figure 4 shows the SD of the histograms of all the nonoverlapping 10-msec segments used in the speech level measurements. Symbols and line types have the same meaning as before. Generally, the effects of delay were very small. A within-subjects ANOVA of the SD values, with factors acoustic environment, processing, gender and delay, gave a significant effect of delay, $F(3,90) = 3.03$, $p = 0.033$. Pairwise comparisons showed that changes from either 18 or 30 msec to 43 msec were significant, ($t \geq 2.09$, 255 df, $p \leq 0.05$), but the change from 7 to 43 msec was not ($t = 1.89$). The SD increased by 0.08 dB as the delay increased from 18 to 43 msec. The effect of delay was stronger in a separate ANOVA of the female results, with factors acoustic environment, processing and delay, $F(3,45) = 4.62$, $p = 0.007$, and was significant when comparing any of the means for 7, 18 or 30 msec with the mean for 43 msec ($t \geq 2.51$, 127 d.f., $p < 0.025$). On average, the SD increased by 0.15 dB as the delay increased from 7 to 43 msec.

Median f0

Since there was no significant effect of delay on the median *f0,* the data were collapsed across delay. Figure 5 shows the base-10 logarithm of the median *f0*, plotted as a function of acoustic environment. Since the effects are small, the results are plotted offset, such that the mean of the four points for each participant group is zero. The offset permits the size of the effect to be compared between participant groups. The offset for the females (open circles, top panel) is $2.29\log_{10}(f0)$ units, equivalent to a normalization to 195.8 Hz, and the offset for the males (filled circles, bottom panel) is $2.048\log_{10}(f0)$ units, equivalent to a normalization to 111.7 Hz. Thus, the ordinate scale is defined as "Relative." Long- and short-dashed lines show results for linear and compression processing, respectively. A within-subjects ANOVA with factors acoustic environment, processing, gender and delay was conducted, based on the $\log_{10}(f0)$ values. Logarithmic values were used to equalize the variability in *f0* between the two participant groups. The values were not normalized. The ANOVA gave a significant effect of processing, $F(1,30) = 8.1$, $p = 0.008$; the *f0* was always higher (on average by 0.65%) for linear processing than for compression processing.

There was also a significant effect of acoustic environment, $F(1,30) = 8.82$, $p = 0.006$. *f0* was 2.1% higher in the booth than in the office. Although there was a significant effect of gender, which would be expected, there were no interactions of gender with any of the other factors.

In separate ANOVAs of the male and female results, using the factors acoustic environment, processing and delay, the effect of processing was significant for the females, $F(1,15) = 5.55$, $p = 0.032$, but not for the males, $F(1,15) = 4.20$, $p = 0.058$. The difference for the females was 0.445%. The effect of acoustic environment was significant for the males $F(1,15) = 10.7$, $p = 0.005$, but not for the females. The difference for the males was 3.2%. There was no significant effect associated with delay in any of the ANOVAs.

Range of f0

Figure 6 shows the range of *f0* values as a function of delay for each gender. To normalize the range of *f0* across the two genders, the range was expressed as the logarithm of the ratio between the 10% and 90% frequencies of the cumulative probability histogram. Results were averaged for the two environments. Filled and open circles show the results for male and female participants, respectively.

A within-subjects ANOVA was conducted with factors acoustic environment, processing, gender and delay. Because of asymmetry in adjacent laryngeal pulses, the *f0* analysis program, "HQTx," can sometimes produce a bimodal distribution of *f0*, containing a peak centered near a value of half the modal *f0*, as well as at the modal *f0*. More commonly, this appears with female voices. The *f0* histogram then appears to have a misleadingly large range.

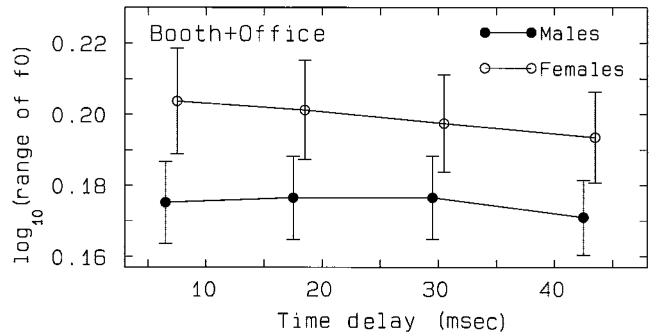

Figure 6. The range of *f0* produced by male participants (filled circles), or female participants (open circles), collapsed across type of processing. Error bars indicate ±1 SE across participants. Filled and open circles show results for male and female participants, respectively.

The value of range used in the ANOVA was limited to 2.52, equivalent to 4/3 of an octave. The choice of limit was a compromise between minimizing the number of true outliers and maximizing the inclusion of valid data from expressive speakers. A one-octave limit affected 7% of female data, 7/6-octaves affected 2.7% and 4/3 octave affected 1%. There was a significant effect of delay, $F(3,90) = 2.73$, $p = 0.049$. The range of *f0* at 7 and 18 msec was significantly greater than at 43 msec ($t \geq 2.35$, 255 df, $p < 0.025$). The reduction in range with increasing delay was $0.0072\log_{10}$ units, a 1.7% decrease.

Only one other significant effect was found. In a separate ANOVA of the male results, with factors acoustic environment, processing and delay, there was a 0.6% wider range of *f0* in the booth than in the office: $F(1,45) = 6.18$, $p = 0.025$.

DISCUSSION

The results presented here were obtained with normally hearing participants. Therefore, we should consider the extent to which the results would be applicable to hearing-impaired users of hearing aids. In our study, the aids were programmed so as to have 0 dB insertion gain for a 65 dB SPL speech-shaped noise input, so that the sounds transmitted via the aids would have had close to normal loudness for a speech input at a moderate level. Many hearing aid designers and fitters state as their rationale that the aids should restore loudness to "normal." If this were achieved, the experience of the impaired listener listening via the aids, in terms of loudness, would be similar to that of our normal-hearing listeners listening through our experimental aids. Hence, we would expect effects on speech production to be similar for the two cases. However, the experience of bone-conducted sound and leakage sound would have been different for our normal-hearing listeners

TABLE 2. Summary of principal statistically significant effects.

	Participant Group		
Measurement	Males	Females	All Participants
Subjective ratings of disturbance of echo	(a) Delay: $p < 0.001$	Delay: $p < 0.001$	(a) Delay: $p < 0.001$*
	(b) Interaction of environment and delay, $p = 0.006$		(b) Interaction of environment and delay, $p = 0.014$
Word production rate	Delay: $p = 0.002$	(a) Delay: $p < 0.001$ (b) Processing: $p = 0.049$	Delay: $p < 0.001$*
Mean vocal level	3-way interaction of environment, processing and delay, $p = 0.05$***	Delay $p < 0.001$*	Delay: $p = 0.004$*
SD of vocal level		Delay: $p = 0.007$*	Delay: $p = 0.03$**
Median *f0*	Environment: $p = 0.005$	Processing: $p = 0.032$	(a) Processing: $p = 0.006$ (b) Environment: $p = 0.008$
Range of *f0*	Environment: $p = 0.025$		Delay: $p = 0.049$*

** Only significant when comparing results at 43 msec to other values of delay.*
*** Only significant when comparing results at 43 msec to values at 18 and 30 msec.*
**** See main text for exact occurrence of significance.*

than for hearing-impaired listeners using hearing aids. The effects of this on perceived disturbance were discussed in Stone and Moore (1999). Briefly, the earlier results suggest that the disturbance decreases with increasing (simulated) low-frequency hearing loss.

The pattern of the disturbance ratings reported here is very similar to that reported in Stone and Moore (1999) for simulated mild to moderate sloping hearing losses. This suggests that the rating procedure is reliable, and supports our earlier finding that delays are rated as "disturbing" when they exceed about 20 ms.

The DAF literature reviewed in the Introduction led to the expectation that word production rates might be reduced by up to about 20% over the range of delays used here. However, as noted in the Introduction, earlier studies on DAF did not usually examine the effects of small delays. In fact, observed reductions were less than 5 to 6%. We have shown that the effects of delay on speech production are very small and are insignificant until the delay exceeds about 30 msec. Even at our maximum delay of 43 msec, the measured effects are small. These smaller-than-expected effects are probably due to the much lower levels of replay in our experiment compared to those reported in the DAF literature.

During data collection, some of the participants did appear to be distinctly affected by the delays; their speech production generally appeared to be labored. To assess whether these subjective impressions were real, a subset of 10 participants was selected; these were judged subjectively to be markedly affected by the delay during initial testing. Statistical analyses of the data from these participants did not reveal any significant effects other than those reported for the whole participant group. Thus, if there were any marked initial effects of the delay for these participants, the effects largely disappeared during the course of testing. We will return to this point later.

A summary of the main statistical findings from the ANOVAs is given in Table 2. The general impression is that effects of delay, the principal parameter under investigation, are more consistently seen among the females. For the males, the acoustic environment plays a greater role. Apart from the peculiar 3-way interaction found for the mean vocal level with the males, the significant effects mostly appear as the delay changes from 30 msec to 43 msec; it is only with the range of *f0* that a significant effect appears below 30 msec. Although the measured effects are small, there is a consistency of the delay range in which they appear. In the design of hearing aid algorithms, it appears possible to use delays up to about 30 msec before speech production is disrupted.

It is noteworthy that there were few significant effects of aid processing. One of our reasons for including two different types of processing was to mimic two different types of rationale that have been used for hearing aid fitting. One rationale is the restoration of loudness to "normal" (Moore, 2000). The linear processing used in this experiment (when applied to normal-hearing participants) could be regarded as being of that type. Another approach is to maximize audibility (Rankovic, 1991) or at least to reduce the variation of audibility with level. An extreme implementation would be to use a compression limiter with a moderate to low compression threshold. Quiet sounds would be well amplified to ensure that they remained audible and would therefore appear louder than "normal." At higher input levels, the compressor would reduce its gain, and sounds would eventually become quieter than "normal." This can be said to reduce the "naturalness" of real world sounds; loudness fluctuations over time would be smaller than normal. The compression processing employed in this experiment (when applied to normally hearing participants) is more representative of this approach. The lack of a significant effect of processing in our experiment indicates that the effect of delay should be broadly similar for different fitting rationales, provided that the fitting rationales do not produce gross deviations from normal overall loudness.

It could be argued that the very small changes in word production rate reported here could be partially explained by the "carry-over" effect reported by Black (1951). In his experiment, the durations of the phrases which he required participants to read were measured in pilot trials where the feedback signal heard by the participants was not delayed. During his main experiment, the mean phrase duration for his no-delay condition was 27% longer than in the pilot trials. Exposure to the delayed signal in other conditions appeared to have "carried over" into the no-delay condition, leading to slower production in that condition. Because of this, there was no significant change in phrase duration until the delay reached 60 msec. If phrase durations were measured relative to those determined in the pilot trials, the change in phrase duration would have been significant once the delay reached 30 msec. However, it is not clear that Black's pretrial measurements of the durations of the phrases were carried out under the same elevated replay levels as in his main test. Additionally, one might expect the carry-over to be influenced by the range of delays experienced by the participants during the test. The delays used here were much shorter than the 300-msec maximum delay of Black.

Although a direct measure of the possible carry-over effect was not structured into the test reported here, it is possible to look at mean production rates while the participants were wearing no hearing aids and had not experienced any delay. Referring back to the experimental method, before the participant wore the hearing aids, they read one passage called "Alice," while wearing the Laryngograph neckband. This passage does not appear in the Lewis Carroll books, although it has a similar vocabulary and grammar. Word production rates for males were 204.5 and 204.7 per minute in the booth and office, respectively. The corresponding rates for women were 206.0 and 202.9. Across all the Lewis Carroll-authored passages, the rates for the same conditions were 205.1 and 204.6, and 211.4 and 210.8; these are within about 3% of the rates for the no-aid, and "naïve-to-delay" conditions. The standard errors are comparable. Even allowing for a difference in writing style between the passages, a carry-over effect approaching 27% is clearly not present. We can conclude that a carry-over effect of the type reported by Black is not responsible for the small effects of delay on speech production rate found in our experiment.

Some limitations of our study should be noted. Pilot trials indicated that scripted speech was much easier to produce than extemporized conversations, especially when listening with the longest delay. Possibly, a higher cognitive load may increase the size of the effects of delay. As in our previous study (Stone & Moore, 1999), the conditions of our experiment did not allow the opportunity for participants to get used to any specific delay. Long-term use of hearing aids that introduce a fixed delay may alter the annoyance produced by, and the ability to cope with the delay. We are not aware of studies of acclimatization to delay, but Gatehouse (1992) showed a long-term acclimatization effect to amplification in hearing aids, so some effect may occur for delay also. However, the effect of delay reported here on speech production was very small, and any long-term acclimatization would only weaken the effects.

To investigate short-term acclimatization effects, we analyzed the word production rates for consecutive 15-sec segments of the first four passages read in any session as a function of delay. This is a similar time scale to that used by Black (1950) in his study of the effect of room acoustics on speaking rate. It is during the first four passages of any test that one might expect to see the most acclimatization. The speech recordings were typically around 3 minutes in length. Over the full length of the recording, production rates typically declined by about 5% over the first 15 to 30 sec of any new condition, before becoming very similar across all delay conditions. The largest decline in mean word production rate during this first 15 to 30 sec was for the 43-msec delay. We only started recording data after about 40 to 50 words into the passage: this was reached typically at least 12 sec after commencing to read. We conclude that that most rapid part of acclimatization was achieved partly before, and partly during the very start of the analyzed portions of each passage.

Another limitation of our study was that we only considered conditions under which leakage of sound into the meatus was minimal; we used deeply fitting foam earplugs and thick-walled tubing. This was done partly because we used normally hearing participants, and the effects of leakage would be greater for these than for hearing-impaired people, as illustrated in Figure 4 of Stone & Moore (1999). It is possible that under conditions where the perceptual effect of the leakage was greater, the tolerable delay would be shorter. This might happen, for example, for people with good hearing at low frequencies and poorer hearing at high frequencies.

Finally, we turn to the important issue of the main factors determining the permissible upper limit of delays in hearing aids. The literature on lip-reading (McGrath & Summerfield, 1985) suggests a limit of a 40 msec; beyond that audio-visual integration may be disrupted. The present data on speech production suggest a limit of 30 msec or more. Our previous paper (Stone & Moore, 1999) and the present results indicate that delays up to 20 msec can be subjectively acceptable during speech production and perception, but delays of 30 msec are clearly disturbing. Agnew and Thornton (2000) reported an even lower limit of about 14 msec for subjective disturbance using critical listeners (engineers). It appears that, at present, the most important factor is subjective disturbance, since this becomes adversely affected at shorter delays than lip-reading or speech production. However, one other factor could play a role, namely the interaction of acoustic delays with motor control and proprioceptive feedback systems. For example, tasks such as typing and racquet sports may be adversely affected by a delay between actions (hitting a key or striking a ball) and the sound associated with those actions. We are not aware of any studies on this topic for a low-delay hearing aid application.

Finally, we mention a practical issue. New hearing aid users often need to be counseled that their own voice may sound unnatural due to occlusion, and that it may take some time to become accustomed to the altered sound quality produced by a hearing aid and to learn to make use of the newly audible speech cues provided by a hearing aid. Further counseling will be necessary if hearing aid processing is used that produces delays longer than 15 to 20 msec, to promote perseverance and acceptance of the aid. The use of such processing may be justified if it produces appreciable benefits.

CONCLUSIONS

This experiment made use of two types of hearing aid processing as well as two acoustic environments to investigate the disturbing effects of processing delays in the production and perception of the aid user's voice. The results indicate that the disturbing effects on perception become significant for delays exceeding about 15 msec in an acoustically "dry" environment, and exceeding about 20 msec in an acoustically "live" environment. Objectively determined parameters from speech production did not show significant effects of delay until the delay exceeded 30 msec. Therefore, the principal factor affecting acceptability of processing delays is that of subjective perception. Our results differ from earlier results on effects of Delayed Auditory Feedback which were obtained using very high feedback levels. It appears that, on average, speech production is unaffected by delay for delays up to about 30 msec.

There are two practical implications of this work:

(1) Hearing aid designers who have an algorithm with a long processing delay that provides demonstrable benefit should not be afraid of using delays up to 15 msec: longer delays of up to 30 msec would require that the aid user is forewarned and counseled as to the potential side effects.
(2) For similar levels of subjective disturbance, the processing delay can be about 4 msec longer in a reverberant acoustic environment than in a near anechoic room.

In addition, in both acoustic environments, participants tended to give higher disturbance ratings to the compression processing than the linear processing at low delays, but lower ratings at the longer delays. Although of marginal significance, this may indicate that compression may lead to a reduced tolerance of delay.

ACKNOWLEDGMENTS:

This project was funded by the Medical Research Council (UK). We thank Audiologic of Boulder, Colorado for the use of the Audallions, and the Department of Phonetics and Linguistics, at University College London for loan of the Laryngograph and assistance with SFS. We thank Thomas Baer and David Wigney for comments on an earlier version of this paper. We also thank three reviewers for their helpful comments.

Address for correspondence: Michael A. Stone, M.A., Ph.D., Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England.

Received May 9, 2001; accepted March 15, 2002

REFERENCES

Agnew, J., & Thornton, J. M. (2000). Just noticeable and objectionable group delays in digital hearing aids. *Journal of the American Academy of Audiology*, *11*, 330–336.

Allen, J. B. (1977). Short-term spectral analysis, synthesis and modification by discrete fourier transform. *IEEE Transactions on Acoustical Speech Signal Processing*, *25*, 235–238.

Bachrach, D. L. (1964). Sex differences in reactions to delayed auditory feedback. *Perceptual and Motor Skills*, *19*, 81–82.

Black, J. W. (1950). The effect of room characteristics upon vocal intensity and rate. *Journal of the Acoustical Society of America*, *22*, 174–176.

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders*, *16*, 56–60.

Butler, R. A., & Galloway, T. F. (1957). Factorial analysis of the delayed speech feedback phenomenon. *Journal of the Acoustical Society of America*, *29*, 632–635.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Heto, R., Kei, J., Lui, C., Kiessling, J., Kotby, M., Nasser, N., El Kholy, W., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G., Westermann, S., & Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America*, *96*, 2108–2120.

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *Journal of Speech and Hearing Disorders*, *20*, 333–346.

Gatehouse, S. (1992). The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids. *Journal of the Acoustical Society of America*, *92*, 1258–1268.

Fourcin, A., & Abberton, E. (1971). First applications of a new laryngograph. *Volta Review*, *69*, 507–518.

Howell, P., & Powell, D. J. (1987). Delayed auditory feedback with delayed sounds varying in duration. *Perception and Psychophysics*, *42*, 166–172.

Killion, M. C., Wilber, L. A., & Gudmundsen, G. I. (1988). Zwislocki was right: A potential solution to the "hollow voice" problem (the amplified occlusion effect) with deeply sealed earmolds. *Hearing Instruments*, *39*, 14–18.

Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, *22*, 824–826.

Mackay, D. G. (1968). Age-linked changes in delayed auditory feedback. *Journal of the Acoustical Society of America*, *43*, 811–821.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, *77*, 678–685.

Moore, B. C. J. (2000). Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore "normal" loudness for speech at different levels. *British Journal of Audiology*, *34*, 165–177.

Rankovic, C. M. (1991). An application of the articulation index to hearing aid fitting. *Journal of Speech and Hearing Research*, *34*, 391–402.

Spilka, B. (1954). Some vocal effects of different reading passages and time delays in speech feedback. *Journal of Speech and Hearing Disorders*, *19*, 37–47.

Stone, M.A., & Moore, B. C. J. (1999). Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear and Hearing*, *20*, 182–192.

Tiffany, W. R., & Hanley, C. N. (1952). Delayed speech feedback as a test for auditory malingering. *Science*, *115*, 59–60.

von Békésy, G. (1960). *Experiments in Hearing* [Translated by E. G. Wever]. New York: McGraw-Hill.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, *60*, 213–232.

Zwislocki, J. (1953). Acoustic attenuation between the ears. *Journal of the Acoustical Society of America*, *25*, 752–759.

Reference Note

1 Stromsta, C. A. (1951). A first approximation of distance from vocal chords to cochlea and the transit time of bone-conducted sound from the region of the vocal chords to the region of the cochlea. Unpublished master's thesis, Ohio State University.