Veterans Administration

Journal of Rehabilitation Research and Development Vol. 24 No. 4
Pages 239–260

A microprocessor-based acoustic hearing aid for the profoundly impaired listener*

STUART ROSEN, JOHN R. WALLIKER, ADRIAN FOURCIN, and VIRGINIA BALL**
Department of Phonetics & Linguistics, University College London, and Department of Clinical Physics and Bioengineering, Guy's Hospital, London, England

Abstract—We have found that the larynx-frequency pattern of speech presented as a sinusoid can be of greater communicative value to profoundly hearing-impaired people than the complete acoustic signal. The presence of higher harmonics can give poorer labelling of isolated intonation contrasts and often minimal gain in segmental spectrally-based distinctions. These observations have led to the development of a practical, body-worn, pattern-processing hearing aid that uses a microprocessor to sense the (analogue-processed) speech fundamental frequency, transform it into an appropriate amplitude and frequency region, and generate digitally the required output sinusoid. Our findings have important implications for the design of other signal-processing hearing aids in demonstrating that a simplification of speech can lead to enhanced speech receptive abilities in persons with impaired hearing.

INTRODUCTION

Recent advances in electronic technology have been a boon for many hearing-impaired people, providing them with efficient, small, and effective aids to hearing. Most impaired listeners are reasonably well served by such aids, many of which can be provided in a behind-the-ear, or even in-the-ear form. Unfortunately, there is still a large number of patients, typically those with more profound losses, for whom currently available hearing aids are only marginally effective. Many such patients choose not to wear any aid, and even those who persevere often obtain only limited benefit because of a severe mismatch between their needs and the functioning of the aid. Although some of this group of patients may be too deaf to use any form of acoustic aid, and thus may be suitable only for vibro-tactile or electro-cochlear aids, the majority by far will have some residual hearing that could be exploited by an appropriate aid if one were available.

The number of patients who might benefit from such an aid is surprisingly large. The best estimates may be derived from a recent study by Thornton (37) who bases his findings on the large-scale National Study of Hearing undertaken by the Medical Research Council's Institute of Hearing Research, and on his own more intensive studies in the south of England. Our preliminary results indicate that listeners with four-frequency (0.5, 1, 2, and 4 kHz) average losses of 100 dB or more, yet who still have some measurable residual hearing, would be likely to benefit from a special aid. According to Thornton, approximately 0.05 percent of the population of England, Scotland, and Wales meet this criterion, about 27,500 people. Assuming similar proportions

*This work was supported by the Medical Research Council of the United Kingdom. Parts of this paper have appeared previously in various conference proceedings, progress reports, and review chapters in books (See references 25–27, 38).
**S. Rosen, J.R. Walliker, A. Fourcin, and V. Ball are with the Department of Phonetics & Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, England; J.R. Walliker is also with the Department of Clinical Physics and Bioengineering, Guy's Hospital, London SE1 9RT, England.

in other developed countries, this is a total population of perhaps 136,000 in the EEC alone (excluding Spain and Portugal), and roughly the same again in the USA and Canada.[1]

Our aims in this work are to: 1) determine the residual auditory abilities of profoundly impaired listeners in psychoacoustic tests; 2) apply a speech-pattern approach in order to design a new type of hearing aid that will match the listeners' needs and abilities; 3) test the efficacy of this scheme in laboratory-based studies of speech perception; 4) build a portable version of the new aid, and; 5) evaluate it in field trials. We have made sufficient progress in our first four aims to arrive at the initial stages of a field trial.

Here, we shall present the results obtained so far, in the order of our listed aims, first describing the subjects who participated and then the equipment used in testing.

SUBJECTS

Four profoundly impaired listeners have taken part in our studies, with the degree of testing varying widely from patient to patient.

Our most extensively tested listener, "C" (age 58) had hearing which was, as far as is known, normal until November, 1974 when he fell and sustained a fissure fracture of the left occipital bone. This led to total deafness in the left ear and a severe sensorineural deafness in the right.

"M" (age 47) began to lose her hearing in 1970. A progressive loss was noticed in her right and then her left ear, and despite extensive investigations, no cause was found.

"J" (a native German speaker) had hearing problems from her teens which were diagnosed when she began work at 19. The suspected cause was measles, but there is also a family history of deafness of different aetiologies. At age 54 (she is now 56) her hearing deteriorated further in her left ear.

"R" (age 76) had a high-frequency hearing loss in childhood which has further deteriorated over the last few years. It is thought that forceps delivery at birth may have been the cause of the original loss.

All four subjects ordinarily wear conventional high-power behind-the-ear hearing aids, which assist in lipreading but do not permit them to understand speech by hearing alone. "C" in particular notes receiving much less help with his lipreading when talking to women and children.

Figure 1 shows the pure-tone audiograms of the four subjects, with four-frequency average losses of about 110–115 dB HL. When the threshold could not be measured, the maximum level of the audiometer (120 dB HL at the relevant frequencies) was used in the calculations.

For presentation of auditory stimuli, either Connevans CE8 or Beyer DT-48 circumaural earphones were used for tests not involving sound field listening. Each earphone was fitted with a microphone mounted on the earphone grid that allowed for direct measurement of sound-pressure levels while the listener wore the earphones. Measurements on a KEMAR manikin indicated that the sound pressure measured by this microphone is within 6 dB of that at the tympanic membrane at 1.8 kHz, within 2 dB at 1 kHz, and closer for lower frequencies. For further details, see Rosen (25).

Measurements of the amplitude responses of the two earphones on the listener's head show them to be similar over the frequency range of 40 Hz to 2 kHz, approximating a high-pass filter characteristic with a cutoff frequency of 175 Hz, as described by Stock and Rosen (36).

Sound field testing used a Celestion Ditton 15XR loudspeaker. All tests were done in a quiet, sound-treated room.

Tests with synthetic speech and nonspeech sounds were under the control of a PDP-12 computer. Most signals were hardware generated, and gated on and off by envelopes controlled by 10-bit digital-to-analog (D/A) converters. Signals for the phase experiments were synthesized directly and played out through the same D/As. Synthesized speech sounds

[1] The figure of 0.05 percent was obtained by subtracting 0.0168 (proportion of population with greater than 110 dB four frequency average [FFA] HL, and thus considered by Thornton appropriate for a cochlear implant) from 0.0678 (proportion with greater than 100 DB FFA HL, dropping the requirement for nerve viability as not applicable to this group). This estimate is conservative because it seems likely that there is a significant number of patients with losses greater than 110 dB FFA who would benefit significantly from an appropriate acoustic aid. It does, however, include both pre- and postlingually deafened persons, and the proportion of older prelingual patients who can learn to make effective use of auditory information is uncertain. The statistics for the EEC are from 1983, giving a total population for the European 10 (Belgium, Denmark, Federal Republic of Germany, France, Greece, Ireland, Italy, Luxembourg, Netherlands, and U.K.) of 272,500,000. Statistics for the USA and Canada were obtained at about the same time, giving populations of roughly 230,000,000 and 24,000,000 respectively.

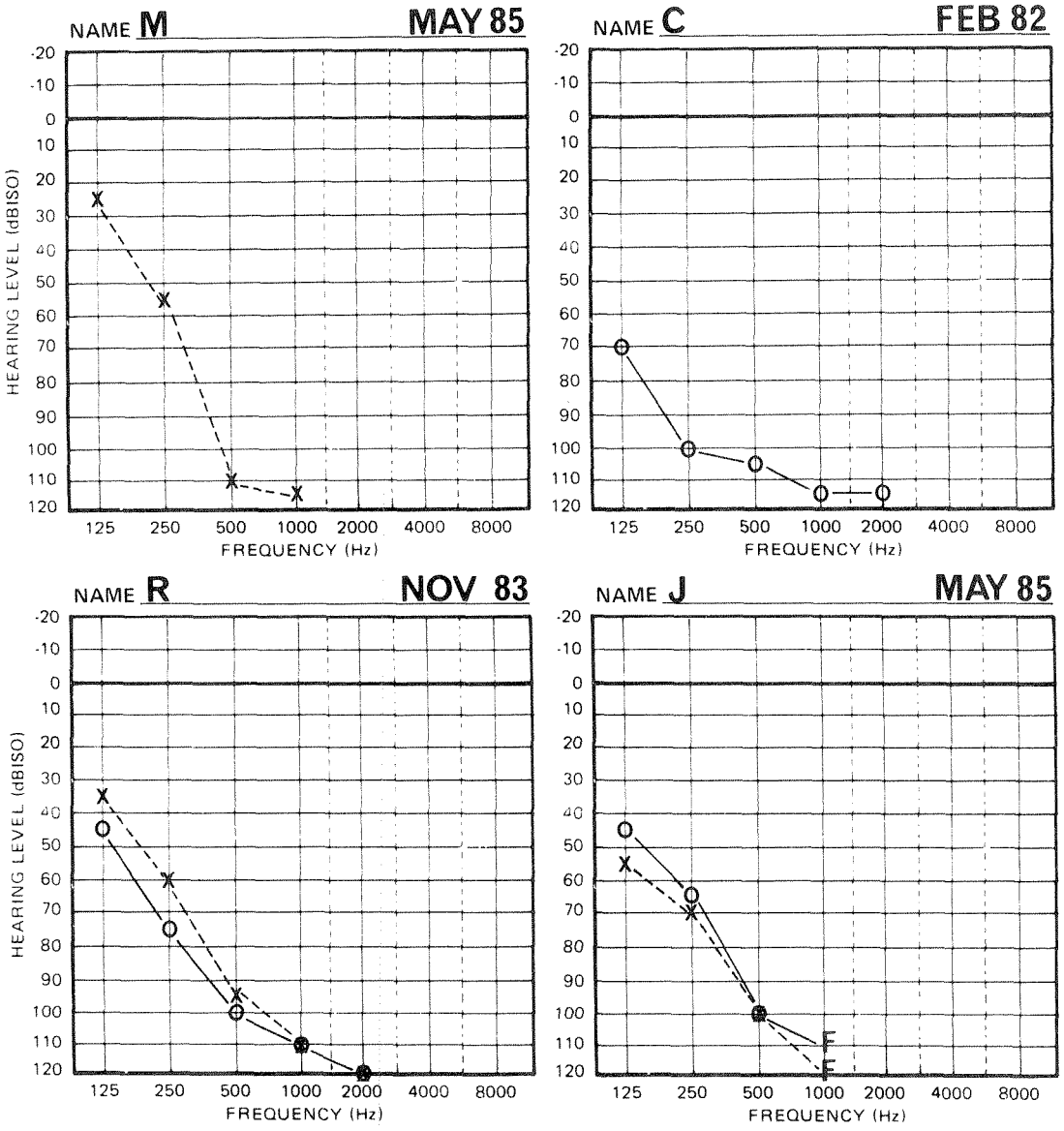

Figure 1.
Pure-tone air-conduction audiograms for the four subjects used in this study. The circles and solid lines refer to the right ear; the crosses and dashed lines refer to the left ear. "F" indicates a sensation of feeling.

were generated by an OVE IIIc serial formant speech synthesizer. Natural speech sounds were recorded and played back using either a Sony U-matic video recorder/reproducer or a Revox B-77 audio recorder/reproducer.

RESIDUAL AUDITORY ABILITIES OF PROFOUNDLY HEARING-IMPAIRED LISTENERS

Many profoundly hearing-impaired patients, including our four subjects, use conventional hearing aids even if they obtain only marginal benefit from them. Their difficulties are directly traceable to the nature of their hearing impairment, and to the design of conventional aids.

Auditory areas

Although traditional pure tone audiograms have their place in portraying hearing loss, we have found it more useful to plot thresholds and comfort and discomfort levels directly in dB SPL, as is shown for the four patients in **Figure 2**. These were determined by the experimenter manipulating the inten-

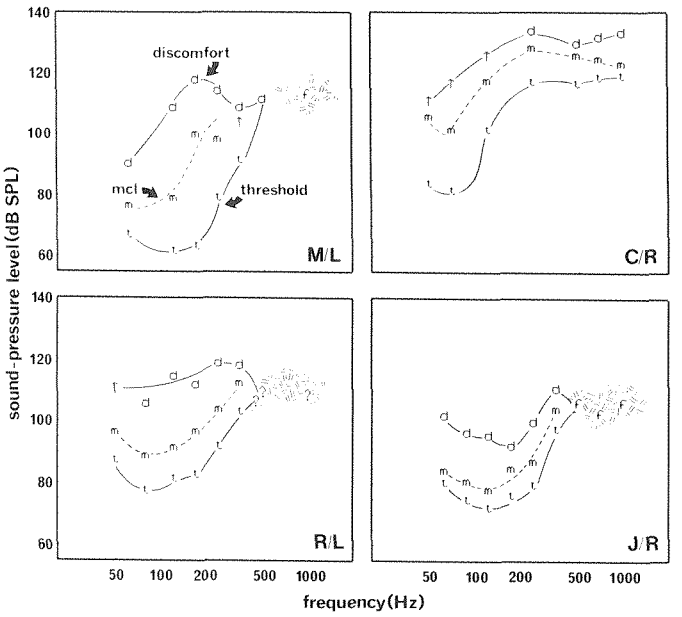

Figure 2.
The auditory areas of the four listeners whose audiograms are shown in **Figure 1.** Only one ear of each was tested, and this is indicated by the "L" or "R" following the subject's initial. Curves were drawn to fit the appropriate points by eye. The absolute thresholds are at bottom and marked with "t." Threshodls for discomfort are marked with a "d," and "most comfortable loudness" (mcl) is indicated by "m" and the dashed lines. Upward arrows generally indicate that the maximum level possible with the equipment had been reached. For "M," however, the upward arrow at about 380 Hz indicates that the sound was "not quite loud enough" to be comfortable, but was starting to be felt. Hatched areas for "M" and "J" indicate frequency regions in which sounds were felt and not heard at the levels used (hence "f"). For "R," question marks are used to indicate that she was not sure whether or not the sounds were heard, but reported they were "more of a sensation than a note" or were "horrid squeaks." ISO standards should not be used to relate these levels to normal hearing thresholds, as the former are based on measurements in 6 cc couplers. Stock and Rosen (36), using the same earphone-mounted microphone technique to monitor sound pressure levels, found normal thresholds to be 30, 20, 12, and 4 dB SPL at 125, 250, 500, and 1,000 Hz, respectively.

sity level of cyclically repeating tone pulses of duration 400 ms (rise and fall times of 50 ms) and interstimulus intervals of about 1.4 s. Intensity levels were measured by the microphone mounted on the earphone grid.

Figure 2 shows that dynamic range is greatest at the lowest frequences (between 50 and 200 Hz) and for three of the four listeners quickly diminishes thereafter. Only "C" seems to have any dynamic range at 1 kHz. The other three listeners find a tone at 1 or 2 kHz "uncomfortable" at sub-threshold levels. (The cause of the discomfort was not clear, but may have been due to tactile stimulation.) The typical rising frequency response of most hearing aids will result in the delivery of the most intense signals into a region in which these patients have no hearing. As a result, when the aid is set to give the highest comfortable output, the all-important low frequencies will be barely, if at all, audible.

Studies like this point up another limitation of conventional audiometry. When sounds are presented only at threshold or near-threshold levels, listeners may be unsure about whether they are hearing or feeling the sound. Thus "R" showed thresholds up to 2 kHz on the pure tone audiogram, when the present experiments suggest that even by 500 Hz, there was no genuine hearing. Both "M" and "J" show a similar effect. Constant questioning of listeners about their sensations during testing (which was an important feature of the measurements of auditory area) may help to resolve some of these discrepancies.

Lack of frequency selectivity

There is a further difficulty. Although the normal ear is known to perform a detailed frequency analysis on incoming sounds, effectively separating different frequency regions into separate "channels" or "critical bands," it seems that such an ability may be completely lost with profound hearing impairment. If one single critical band, extremely restricted in frequency range, is all that remains, the threshold curves shown in **Figure 2** may be nothing more than the amplitude response of the presumed remaining auditory filter. In physiological terms, there may be a small patch of auditory nerve fibres remaining, whose characteristic frequencies are near 100 Hz. The perception of all other sounds would then be mediated through these fibers. One way to test this hypothesis psychoacoustically is to look for instances in which there is more masking from remote than close frequencies. For example, if the single auditory filter were centered at 100 Hz, we would expect, for a probe tone of 250 Hz, that 100 Hz would be a more effective masker than 300 Hz. Unfortunately, such experiments are often inconclusive because of the difficulty of getting any masking at all in such profoundly-impaired listeners.

This presumed loss of selectivity means that

speech information in different frequency bands is mixed together, instead of being kept separate, as in the normal ear. Thus the patients lose the ability to distinguish sounds such as vowels which differ in their spectral shape. Furthermore, background noise can now exert its influence over the entire audible spectrum, so patients are much more affected by it than the normal listener.

Even in the case of profound impairment, however, patients usually have some residual abilities that can be used effectively. For example, they are able to detect changes in intensity, even though the number of discriminable steps between threshold and discomfort is likely to be small, given the small dynamic ranges.

Distinguishing periodic from aperiodic stimuli

Of more interest from the speech-perceptual point of view is that profoundly impaired listeners can distinguish between periodic and aperiodic signals, as long as they fall within their range of relatively good hearing. This ability has been tested using a technique first described in 1968 by Risberg (22), whose primary purpose was to distinguish performances based on residual auditory function from those based on tactile responses. (As our listeners are postlingually deafened with many years of normal or near-normal hearing experience, this is much less of a serious problem than it would be with the prelingually deaf.) We have also found this "tone-noise" test useful in assessing patients undergoing acute electro-cochlear stimulation (7).

The task is simple: the listener is presented with two sounds, one periodic and one aperiodic, and asked to indicate which of the stimuli was the noisy one. In contrast to most psychoacoustic tests, it is the stimulus duration that is varied (with the two stimuli on a given trial of equal duration) in order to estimate the point at which the subject is 71 percent correct using an adaptive tracking procedure (15).

All stimuli had 10-ms rises and falls. The tracking procedure started with a stimulus duration of at least 250 ms. Initial step-size was 40 ms and was decreased by a factor of 2 after each of the first four turnarounds, making the final step size 5 ms. At this point a further 10 turnarounds were run, and the peaks and valleys of these 10 points averaged to estimate the stimulus duration necessary for 71 percent correct performance.

The sounds were presented at a comfortable level, as determined by each listener. In order to eliminate loudness differences as a possible cue, the subject adjusted the two stimuli to be of equal loudness at the start of each session.

Table 1 shows the available results. Listener "C" shows, as expected, a much better performance for the right ("hearing") ear than the left ("feeling") one. Thus "C"'s auditory sensations, as impaired as they are, are considerably more refined than his tactile ones (though better tactile performance might result from different stimulating systems on different parts of the body). Also interesting is the way in which noise bandwidth affects performance, with a general trend for better performances with wider bandwidths, up to a point.

The performance of "M" was considerably better than that indicated by the measured value of 34 ms. The minimum duration of the stimuli allowed by the testing program was 20 ms, and the adaptive procedure often perseverated on this value. In the

Table 1.
Results from the "tone"–"noise" discrimination task. The two entries for subject "C" (C/R and C/L for right and left ears, respectively) are from two different experimental sessions. The second pair of entries for the right ear of subject "C" (C/R) was obtained in a separate session.

	Condition		Duration	
Listener	sine frequency	limits of noise	(ms)	No. tests
M/L	200 Hz	100 Hz–400 Hz	34	2
C/R	141 Hz	20 Hz–20 kHz	41	4
C/L			145	4
C/R	141 Hz	100 Hz–200 Hz	93	4
	200 Hz	100 Hz–400 Hz	40	2

second of the two tests performed, she correctly identified the stimulus pair at 20 ms for 13 of 18 times (i.e., 72 percent of the time), and at 25 ms she correctly identified the stimulus pair 10 times out of 12 (i.e., 83 percent of the time). Therefore, a better estimate of her performance is close to 20 ms. That the performance of "M" is better than that of "C" parallels the differences in their pure-tone thresholds at low frequencies.

Discrimination of changes in fundamental frequency

The ability of the listeners to hear changes in frequency in the low-frequency range (below 400–500 Hz) is of practical importance in the perception of intonation. Unlike normal listeners, however, it appears that the profoundly impaired are best at discriminating frequency changes when the signals are simple sinusoids, rather than multiharmonic pulse trains or speech. Thus, the high-frequency information (that would be transmitted even by a flat-response aid) may interfere with the perception of information in the low frequencies, even though it cannot be used as a source of information itself. Furthermore, changes in the relative phase of the harmonics making up the periodic signal can have important effects on the ability of the impaired listeners to make the discrimination.

That the fundamental frequencies of sinusoids are better discriminated than those of pulse trains is shown by the results of a simple fixed-level, two-interval, two-alternative forced-choice task. During any particular experimental session (40 trials), the two frequencies presented on any given trial were fixed and only the order in which they occurred was randomized. The subject indicated whether the pitch rose or fell from the first to the second. Feedback was provided after every response. Stimuli were presented over one earphone of the Connevans headset at a comfortable level.

Four experimental conditions were obtained by using two different sounds (sinusoids and pulse trains) at each of two different frequency ranges, centred on 100 and 240 Hz. The two frequencies used in a session were either 96.6 and 103.5 Hz, or 231.8 and 248.4 Hz (always a change in frequency of 6.9 percent, referred to the geometric mean of the pair).

Stimuli were 500-ms in duration with 50-ms raised-cosine rises and falls. Sinusoidal stimuli were generated by a digitally-programmable signal source. For the pulse train stimuli, the same generator was used to trigger 200 μs pulses that were lowpass filtered at 2500 Hz (48 dB/octave) before being gated with the appropriate amplitude envelope.

In order to ensure that loudness variations could not aid the subject, each stimulus was jittered in amplitude by a random amount (different for each presentation) within the limits ± 3 dB in 0.25 dB steps. Preliminary loudness matching showed that, on average, the two stimuli of the pair would be of equal loudness when one was adjusted by 0.2 dB.

During any experimental session, only one combination of center frequency and stimulus type was used. Each combination was run three times, in a random order–120 trials per condition.

A summary of the results in terms of percent correct may be found in the left half of **Table 2**. The sample size is large so all statistics use the normal approximation to the binomial distribution to test differences between proportions (35) and one-tailed tests.

There are three main findings: 1) For stimuli that consist of a number of harmonics, sensitivity is greater at low frequencies than at high ($p = .043$); 2) This holds true even for sinusoidal stimuli, although to a lesser degree, which is reflected in the nonsignificant difference between results from the two centre frequencies using the sinusoidal sounds ($p = .087$); and 3) Performance with sinusoids is much better than that with pulse trains, the difference between the two waveforms being most significant at the higher centre frequency ($p = .015$).

These differences in performance may seem rather small when expressed as proportions. It is easier to appreciate their significance when they are converted into a more familiar form, the 75 percent correct Weber fraction. This has been done in the right side of **Table 2**, under the assumption that the psychometric function is a cumulative normal distribution in log frequency symmetrically placed at the center frequency of the stimulus pair to be discriminated. In other words, we assume no time-order error and a linear relationship between log frequency and the normal deviate of the proportion correct. Here we see that the Weber fraction at 240 Hz is nearly twice as big for pulse trains as for sinusoids.

"M" showed a similar result in an adaptive

Table 2.
Results from a frequency discrimination task comparing "C"'s performance with sinusoids and pulse trains at two center frequencies.

	a Proportion correct		b Weber fraction (%)	
Center frequency	Type of sound: Sinusoids	Pulses	Type of sound: Sinusoids	Pulses
100	.858	.767	4.3	6.4
240	.792	.667	5.7	10.8

frequency discrimination task for a center frequency of 100 Hz. In these tests, the intensity of the sinusoid was set to be equal to that of the fundamental in the pulse trains (about 72 dB SPL). Amplitude jitter of ±2 dB was present, and three just-noticeable differences (jnds) determined for each condition. For sinusoids, the jnd averaged 3.2 Hz, and for pulse trains, 4.1 Hz. A much smaller difference (although in the same direction) was found in pilot investigations for a center frequency of 200 Hz. In fact, as an inspection of **Figure 2** shows, a pulse train of 200 Hz whose fundamental is at a comfortable level (about 95 dB SPL in these tests, or 28 dB SL) will have its second harmonic very near threshold, and thus unlikely to have much of an influence on perception. It is as if the listener were perceiving only a sinusoid out of the total complex. For 100 Hz pulse trains, however, the fundamental was about 10 dB SL, and the second harmonic about 5 dB SL. We would therefore expect that whatever differences there were between the perception of pulse trains and sinusoids, they should be bigger at 100 Hz than at 200 Hz.

Stock and Rosen (36) also determined the relative discriminability of frequency changes in sinusoids and pulse trains with an adaptive technique, using listeners having a wide range of hearing losses (normal to profoundly impaired) and center frequencies (125–500 Hz, the major part of the voice pitch range). As had been found in previous studies (11, 12), normal listeners always had smaller jnds for pulse trains than for sinusoids. The profoundly impaired listeners always did better with the sinusoids, although this advantage was much greater at 125 Hz than at 500 Hz.

In our view, these results are best accounted for (in both normals and hearing-impaired listeners) by a model that proposes a primarily temporal analysis of incoming sounds, but preceded by a frequency analysis: e.g., van Noorden (20); Moore and Glasberg (18). For impaired listeners, sinusoids lead to better performance because degraded auditory frequency selectivity allows individual harmonics in complex tones to interact to a much greater degree than is the case in normal-hearing listeners. Thus the waveform presented to the temporal analyzers is more complex, leading to a more ambiguous pitch. In normal-hearing listeners, the lowest harmonics are effectively resolved from one another, presenting the temporal analyzers with simple sinusoids. For a more detailed discussion of these issues, see References 26 and 36.

Phase effects

If the poorer performance of impaired listeners with complex tones is indeed linked to an insufficient resolution of harmonics, we might expect the relative phase of stimulus components to influence the perception of pitch to a much greater degree in the impaired ear. Since relative phase is an important determinant of waveshape, there will be phase relationships that lead to waveforms in which the fundamental periodicity is clearly marked by large peaks, and others in which the periodicity is disguised (as in **Figure 3**). Normal listeners will not be affected by phase relationships because the lower harmonics, which are the most important for the perception of pitch in complex tones (18), are essentially resolved from one another into sinusoids before being temporally analyzed. It is the presumed *interaction* of even the lowest harmonics for the impaired listeners that could give rise to phase sensitivities.

Such effects have been reported by Hoekstra and Ritsma in 1977 (13) and Hoekstra in 1979 (12), albeit for sounds that are only remotely related to speech. Their tests were roughly equivalent to perceiving changes in the fundamental frequency of a speech sound from three harmonics centered at 2 kHz (the so-called "residue"). Hoekstra (12) reported that three of five hearing-impaired listeners were significantly better in discriminating changes in frequency when the components in the complex were "in phase" (with clearly defined peaks), than when the phase of the central component was shifted by 90 degrees (leading to a much flatter wave envelope). Normal listeners (and the other two hearing impaired

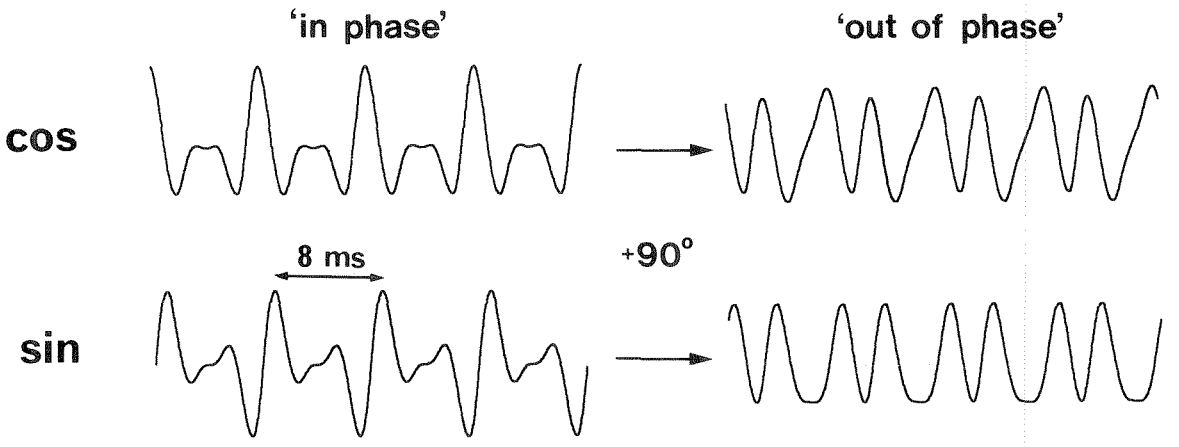

Figure 3.
Examples of waveforms used to test the effect of the phase relationship of harmonic components on the ability of the listeners to discriminate changes in fundamental frequency. All consist of the first three harmonics of fundamental 125 Hz, in which the 250 Hz component is 6 dB more intense than the other two. Both waveforms at left can be considered to be "in phase." At the top, the three sinusoidal components are added up in cosine phase, while for the one at bottom, the waveforms are added together in sine phase. At the right are seen the results of shifting the central component of each complex on the left by 90 degrees. Note how the envelopes of the signals become much flatter, leaving the fundamental period of the signals more difficult to discern from the waveshape.

listeners) showed no difference between the two conditions.

A similar result has been obtained with rather more speechlike sounds with "C," who was tested on his ability to discriminate changes in fundamental frequency for three-harmonic complexes with fundamental frequencies near 240 Hz. The fundamentals of the two stimuli to be compared were fixed at 228.6 and 252 Hz (a change of 10.25 percent) and could be either "in phase" or with the central component shifted 90 degrees "out of phase" (see **Figure 3**). The phase relationship of all the stimuli within a session was always the same. A two-interval two-alternative forced-choice (2I-2AFC) task was used, in which the listener was required to label the direction of the pitch change. To prevent the use of any possible loudness cues, each sound was jittered by a different random amount each presentation, over a range of ±2 dB. **Table 3** shows that the change in fundamental frequency between the two stimuli is more salient for the complexes that are "in phase." Note though that the phase relations are specified at the *input* to the headphones. In fact, when all components were "in (sine) phase," this led to complexes that were in approximate inverted cosine phase (as determined by a microphone mounted on the earphone grid).

Effects of phase on pitch saliency have also been demonstrated with "R," using four-component harmonic complexes in which all components were of equal amplitude. The "in phase" complex had all components in cosine phase, while the "out of phase" complex had alternating cosine and sine terms. In addition, sinusoids at the level of the fundamental in the harmonic complexes were also used. The center frequency of the fundamentals was always 125 Hz, and phase and amplitude corrections were applied to ensure the proper relationships at earphone output as confirmed by the grid-mounted microphone. In a particular experimental session, the frequency difference between any pair of stimuli was fixed, but all three types of stimuli occurred (although comparisons were only made between sounds of the same type). Three frequency differ-

Table 3.
Percent correct in discriminating a fixed change in fundamental frequency in a three-component harmonic complex when the relative phases of the components in the pair of stimuli to be compared are varied.*

Phase relationship "In phase"	"Out of phase"	Statistical significance of the difference
65.6% (of 90)	47.8% (of 90)	$p<.05$

*The central second harmonic was 6 dB more intense than the first and third, which were equal in amplitude. The significance test is a two-tailed test of differences between proportions.

Table 4.
Percent correct in discriminating three fixed changes in fundamental frequency in a sinusoid, and in a four-component harmonic complex when the relative phases of the components in the pair of stimuli to be compared are varied and are in a sinusoid.*

Frequency difference	Phase relationship Cosine In cosine phase	Alternating	Sinusoid
5.9%	80.0	75.0	90.0
7.9%	95.0	72.5	95.0
9.8%	87.5	80.0	90.0
mean	87.5	75.8	91.7

*The only significant differences in the first three rows are for those stimuli involving a frequency difference of 7.9 percent. A two-tailed test of differences between proportions was used. Sinusoids and "in phase" components are discriminated better than alternating phase stimuli at the .01 level.

ences were possible, and two sessions of each were run, making a total of 80 judgements per frequency change per stimulus type. **Table 4** summarizes the results.

Due to the amount of variability, there is not a smooth increase in performance with increasing frequency separations. Even so, it is clear that "in phase" components always lead to stronger pitches than "out of phase" ones, and sinusoids are discriminated best of all, although the differences between these and "in phase" sounds are small.

"In phase" components do not always lead to more salient pitches. **Table 5** shows the results from another frequency discrimination task performed by "C." The sounds were again four-component harmonic complexes with all components at the same amplitude. The fundamental frequencies of the two stimuli to be compared differed by 10.25 percent (238.1 Hz and 262.5 Hz), and the components could be either in sine phase (measured at the earphone *output* as phase corrections were applied), or with alternating sine and cosine terms. No amplitude jitter was present. These results show that the detailed temporal structure of the sound needs to be taken into account, as well as the phase distortions imposed by the listener's auditory system, which may vary from individual to individual (24, 25).

Phase relations may also be of importance to impaired listeners in another way (25, 26). In so far as temporal information is important in the perception of spectral shape, as has been proposed, for example, by Young and Sachs in 1979 (33, 39), impaired listeners will hear changes in phase as changes in vowel quality. In fact, many listeners, both normal and impaired, report phase changes in harmonic complexes as changes in vowel quality (16, 23, 34). Darwin and Gardner (3) have shown changes in vowel labelling performance with changes in phase for normal listeners, and these effects are likely to be stronger in impaired listeners.

Table 5.
Percent correct in discriminating a fixed change in fundamental frequency in a four-component harmonic complex when the relative phases of the components in the pair of stimuli to be compared are varied.

Phase relationship Sine	Alternating	Statistical significance of the difference
60.0% (of 110)	100.0% (of 110)	$p<.001$

In general, then, we expect impaired listeners to be much more influenced by phase changes than normal listeners. This arises directly from degraded frequency selectivity. Phase sensitivity is understood to reflect the *failure* of frequency resolution—only when a sound's constituent sinusoids interact (i.e., lie sufficiently within a single critical band, or auditory filter) will a phase change be detectable. As many impaired listeners have impaired selectivity, it seems likely that phase will play a larger role for them than for normal-hearing listeners (23–26).

Rosen (23, 24, 26) has presented direct evidence for this in a number of moderately impaired listeners, who were shown to be better at detecting certain phase changes than *any* normal listener tested. The profoundly impaired listener will be unlikely to perform better than normals, due to the loss of temporal, as well as frequency resolving power. The relative importance of phase still seems likely to be greater.

Not all phase effects may be negative. Since impaired listeners can hear changes in phase as changes in vowel quality, even when they are poor at distinguishing vowels on the basis of their amplitude spectra, phase manipulations could provide a way of signalling useful information. The problem, though, is that if phase changes also affect the saliency (and indeed even the perceived value) of the voice pitch, how is one to independently manip-

ulate the two perceptual features? If nothing else, a consideration of phase relationships may allow one to manipulate the relative amplitude of harmonic components to obtain a range of vowel qualities, while at the same time maintaining, as far as is possible, the saliency of the pitch changes.

Frequency discrimination for sinusoids across frequency

For the moment, however, it is clear that if one wants only to signal changes in fundamental frequency in the low frequencies, it is best done with a sinusoid. Even so, the profoundly impaired listeners' abilities vary significantly within this range. We have already seen in **Table 2** that for "C" performance is better at 100 Hz than at 240 Hz. Stock and Rosen (36) showed for their profoundly impaired listener that, compared with normal-hearing listeners, performance was much more severely degraded at 500 Hz than at 125 Hz. Therefore, there may be some advantage in "mapping" high pitches into low ones.

A SPEECH-PATTERN APPROACH TO HEARING AID DESIGN

Our general approach has two main components. First, we determine which speech-pattern elements are most useful to the impaired listener. Second, the desired features must be presented in a way matched to the listeners' residual abilities. For the listeners studied here, it is clear that their extremely limited auditory abilities will not allow the understanding of speech by auditory means alone. A simplification of the signal is desirable. Furthermore, given that our subjects are already using visual cues from lipreading to a large extent, we wanted to provide information that is *not* readily visible. For further details of this approach, see References 17 and 26.

Voice pitch information fulfills both these criteria. It is relatively invisible yet provides crucial information about whether the vocal folds are active or not (the segmental feature of voicing) and their vibration rate (the physiological correlate to voice pitch, or intonation). It has been shown in a number of studies to be an effective aid to lipreading (10, 21, 29). Moreover, its frequency and amplitude characteristics match those of the profoundly impaired listeners quite well. Even when all intensity variations within it are suppressed, it still provides a good aid to lipreading. Furthermore, it falls in the frequency range where listeners are relatively good at discriminating frequency changes, especially when presented as a sinusoid. For high-pitched voices, where listeners' abilities to discriminate changes in frequency worsen, the pitch contour may be mapped downwards into a frequency range in which discrimination abilities are better. For this we use an option called "Mapitch," which typically subtracts 50 Hz from all frequencies, not only lowering the contour, but also stretching it in ratio terms.

Our first speech-processing scheme for profoundly impaired listeners is based on these considerations. The fundamental frequency of speech is extracted and presented to the listener as a frequency-modulated sinusoid at his or her most comfortable level. We call this a SiVo aid, for Sinusoidal Voice. Before we discuss a practical implementation of it, we may ask how it serves listeners in their perception of more speech-like sounds.

SPEECH-PERCEPTION WITH THE SiVo AID

The main aim of the following studies is to determine if a simple SiVo signal aids speech reception better than the full complex speech signal. Our approach to assessing a listener's performance is based on the idea that no single test will suffice, and that there must be a balance between tests that investigate analytically the perception of specific speech features, and those that attempt to assess the patient's performance in a more global way. Also, since we view the SiVo primarily as an aid to lipreading, tests are often done audiovisually and concentrate on the assessment of those features that are known to be important aids to lipreading. Rosen, Moore and Fourcin (31) have shown that the voice pitch contours of speech are probably a more important aid to lipreading than just an indication of the presence or absence of voicing, so we will consider those abilities first.

Natural question and statement contours

Our test of the ability to hear changes in natural voice pitch is based on question/statement opposi-

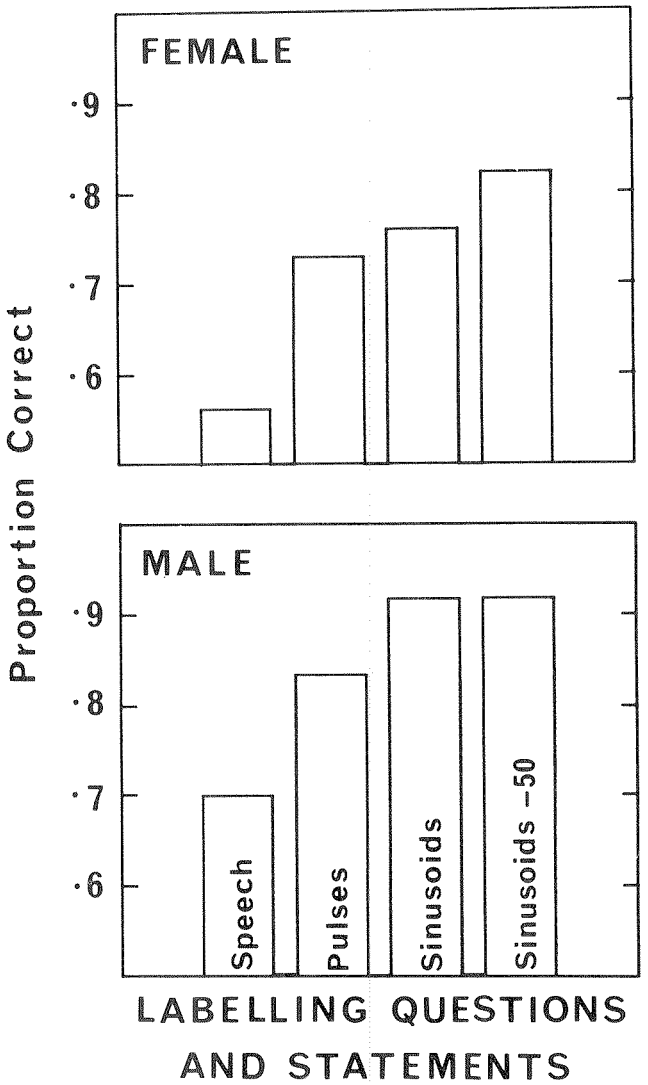

Figure 4.
The results of "C" from the natural speech question/statement task, using a male and female speaker.

tions that can be signalled by a straightforward acoustic contrast. Two-channel recordings were made of a man and a woman uttering eight simple declarative sentences (e.g., "She's reading a newspaper," "It's pouring with rain") with both statement and question intonation; statements were signalled by a falling fundamental frequency contour on the last word and questions by a rise. One audio channel carried the speech signal from a microphone. On the other was the signal from an electro-laryngograph (6), a device that detects vocal fold closures by measuring the electrical conductivity between two electrodes placed on the speaker's neck.

Three experimental sessions were recorded by each of the speakers, both of Southern British origin. A session consisted of two tokens of each sentence spoken with each of the two possible intonation contours. These were randomized without replacement, making a total of 32 trials.

Four different acoustic signals were presented to "C": one was the speech signal itself. The laryngograph signal was utilized in the other three conditions to give information only about the larynx frequency of the speaker. In one, .5 ms rectangular pulses were triggered by vocal fold closures. For the others, fixed-amplitude frequency-modulated sinusoids were presented: either at the larynx frequency of the speaker or at a frequency reduced by a constant 50 Hz relative to the larynx frequency (the Mapitch option).

The patient listened over the Connevans earphone at a comfortable level and responded "question" or "statement." The nature of the contrast was explained using an interactive visual display, the Voiscope® (1). Intensive practice in all conditions preceded testing.

During the experimental sessions proper, "C" wrote his responses on a sheet of paper and received no feedback. The sessions were run in order, the three sessions of the male speaker alternating with the three of the female speaker. The four conditions of stimulus presentation were presented in an order determined by randomly choosing permutations of the 24 possible under the constraint that each combination of session and condition occurred only once. This made a total of 24 sessions.

Mean results as a function of condition are displayed separately for each speaker in **Figure 4**. For both speakers, performance is best with sinusoids and worst with speech, with pulse trains falling between. For the male speaker, it doesn't matter if the sinusoids are lowered in frequency or not. Both sinusoidal presentations lead to identical scores. For the female speaker, however, superior performance comes from sinusoids mapped down by 50 Hz as compared to unmapped sinusoids. Finally, performance with the female speaker is worse than with the male speaker in all conditions. Listening to female speech, for instance, performance is only slightly better than chance at 56.3 percent.

This difference between the perception of male and female voices is especially interesting as it confirms C's observations that he has more trouble understanding women speakers. In our estimation,

this is primarily due to female speakers having higher fundamental frequencies and is another reflection of the fact that "C"'s frequency discrimination abilities worsen with increasing frequency.

The three other listeners have undergone few tests of this sort, and then only with the female speaker. Generally speaking, they perform well (91 percent upwards) as long as they listen over headphones, whether to sinusoids or natural speech. This may simply result from the test being too easy for most listeners. Differences *are* shown when the subjects perform the test free-field using their own hearing aids. In that situation, both "M" and "J" ("R" has not been tested) evidenced better performance for sinusoids at the fundamental (mean of 98 percent) than for natural speech (mean of 84 percent).

Synthetic frequency glides

Since there is always a fair amount of variability in natural speech, even within speakers, we also synthesized a prosodic test that assured us of precise control. In this way, we would be able to test abilities to hear dynamic pitch changes over a range of frequencies, without having to worry about the different frequency ranges that might be used by different speakers.

This test used a continuum of stimuli, all with simple rising or falling fundamental frequency contours (7), for each type of sound and each frequency range desired (illustrated at the bottom left of **Figure 5**). Each stimulus had a constant fundamental for 60 ms both at its beginning and its end, with a 180-ms frequency transition in between which was linear in log frequency. The extreme fall and rise were both changes of an octave. The four intermediate stimuli had their initial and terminal frequencies in equal logarithmic steps between the defining octave interval. Responses were analyzed by computing, for each experimental session, the proportion of times each of the stimuli was labelled a "rise."

These frequency contours could be realized in three different ways. Sinusoidal stimuli were generated by simultaneous amplitude and frequency modulation of a sine-wave generator with the digital-to-analog outputs of the computer. Synthetic speech stimuli were generated by the OVE IIIc speech synthesizer under computer control. The spectral envelope of these stimuli was static in time, with formant values set for an /*a*/. For the lowest frequency range (80–160 Hz), the formant frequencies were appropriate for a man while for the two higher frequency ranges (130–260 and 200–400 Hz), the formant frequencies were appropriate for a woman. Finally, pulse-train stimuli were created by triggering .5 or 1.5 ms positive-going rectangular pulses from the output of either the sine-wave oscillator or the speech synthesizer with all its formants bypassed. All stimuli were presented through the Connevans earphones at a comfortable level chosen by the particular listener.

As in the labelling of natural prosodic contrasts, listeners received extensive training prior to the experimental sessions using live voice and synthetic presentations, labelling the stimuli as "rise" or "fall." Again, the Voiscope proved useful. In fact, for "C," these tests preceded the previous ones and at the beginning of training, he was unable to reliably label even the most extreme stimuli presented on the continuum.

Again, "C" has been tested most extensively. For a particular combination of waveform and frequency range, results were consistent from session to session and so were averaged. The two pulse widths used in the pulse-train stimuli from 200–400 Hz also gave similar results and so were averaged. For further experimental details, see Reference 27.

The main experimental results are found in the seven square boxes of **Figure 5**. Each box contains the results from one experimental condition, with frequency range given by the column and stimulus type given by the row. The results are plotted in the form of a categorization curve: i.e., the proportion of "rise" judgements as a function of stimulus number. The slopes of the categorization curves may be taken as a measure of the accuracy of the subject's responding, with steeper functions indicating better performance. These were quantified by fitting cumulative normal functions to each of the curves, using a maximum-likelihood technique (2), the relevant extracted estimates being used for tests of statistical significance.

Three main findings may be noted, all in agreement with the results obtained with natural prosodic contrasts. First, for the speech stimuli, labeling ability is much superior for low frequency stimuli than for high. "C" is not able to consistently label the higher fundamental frequency stimuli (200–400 Hz) at all, even for the extreme stimuli which change by an octave. The estimated slope of the fitted

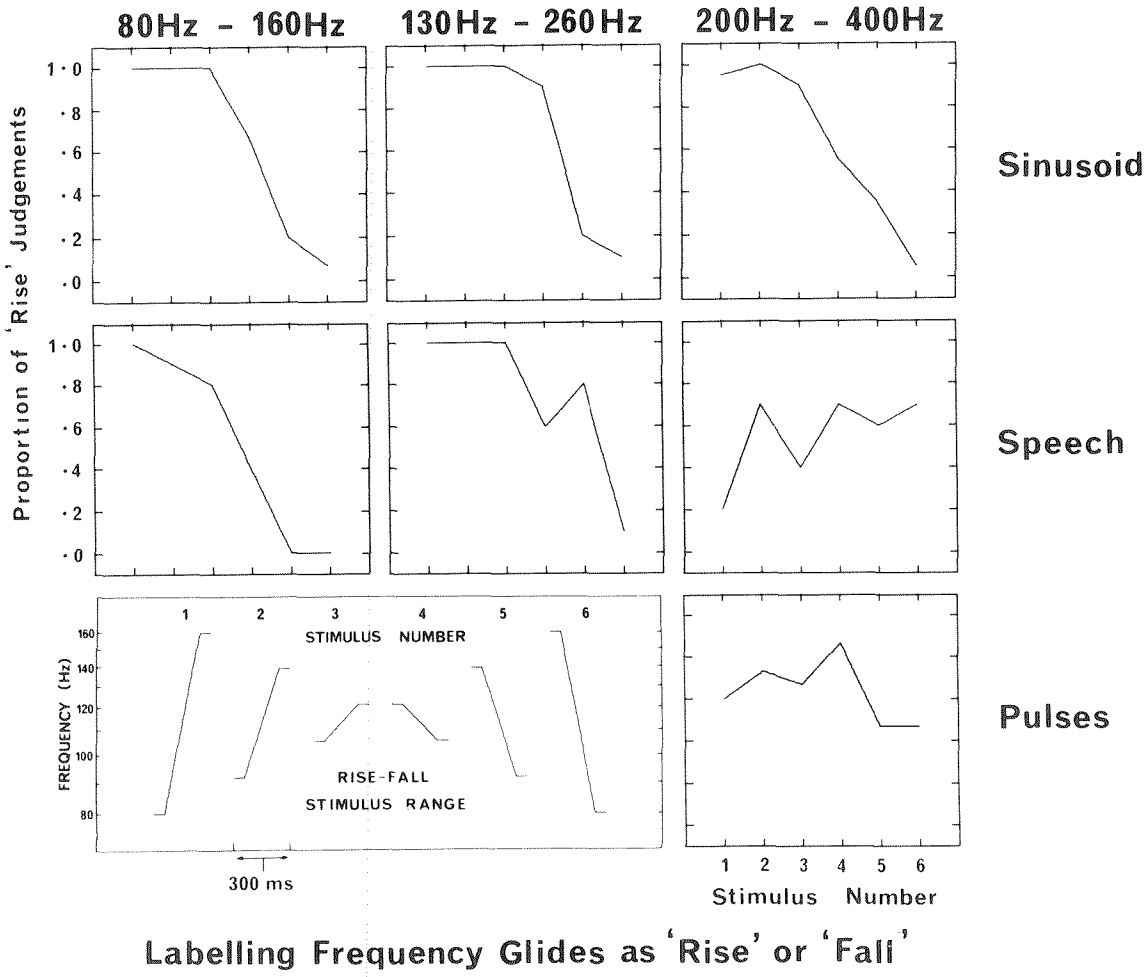

Figure 5.
The results of "C" from the synthetic rise/fall test.

function was not significantly different from zero; and was significantly different from the non-zero slopes of the two lower frequency ranges. That this difference cannot be attributed to the order in which the stimuli were presented is confirmed by the results in the 100–200 Hz range (not shown in **Figure 5**), two sessions of which occurred between the two sessions of the 200–400 Hz set. In the former condition, categorization was good and the slope of the categorization curve was not significantly different from those exhibited by the 80–160 Hz and 130–260 Hz ranges. All three of these conditions gave slopes significantly different from zero. Although the 80–160 Hz range had a steeper estimated slope than the 100–200 Hz range, which was in turn steeper than the 130–260 Hz range, these differences were not significant.

Second, even for the sinusoidal stimuli, labelling performance is superior at low frequencies, even though C has some ability to label the high frequency stimuli correctly. This is reflected in the significantly shallower slope (using a one-tailed test at the .05 level) obtained with the highest fundamental frequency range. There is no statistical difference between the slopes of the two lower frequency ranges.

Finally, for a given frequency range, steeper slopes are always obtained with the sinusoidal stimuli than for the speech stimuli, though only for the 200–400 Hz range is this difference statistically significant (using a one-tailed test at the .05 level). This also holds true for the pulse stimuli which do not have a slope significantly different from zero. **Figure 5** shows clearly that in this frequency range, performance with the sinusoidal stimuli is much superior to that with either speech or pulse stimuli.

There is, however, one way in which these results differ from the previous ones. In the high-frequency

Table 6.
Percent correct in a test involving the identification of four "h-vowel-d" words, done on the basis of the complete speech signal, and a sinusoid at the voice fundamental.

Listener	Earphone presentation SiVo	Earphone presentation Speech	Sound-field speech via hearing aid
M/L	42.5 (of 80)	40.0 (of 80)	25.0 (of 40)
C/R	42.5 (of 80)	39.2 (of 120)	40.0 (of 40)
J/R	20.0 (of 40)	55.0 (of 40)	69.2 (of 39)*

*This condition with bilateral hearing aids.

range, performance with synthetic speech and with pulse trains is not very different. This may be due to the spectral and overall amplitude variations present in natural speech that have been eliminated by using steady-state stimuli. This increased similarity between the stimuli may be responsible for the more similar results. In addition, it is difficult to compare the level of difficulty across the two tasks and it may well be that the synthetic stimuli are rather less physically different in their fundamental frequency contours than the natural stimuli. It is certainly true that the synthetic stimuli are much shorter and therefore may present a more difficult task. In this way, differences between conditions are reduced by the rather poor performance near chance levels.

As for the natural speech question/statement test, the two other listeners tested with these synthetic stimuli ("M" and "J") do considerably better than "C." Both are able to label sinusoids, pulse trains, and "speech" quite well, even in the 200–400 Hz range, as long as the sounds are presented over earphones. "J" does well even when she listens to the sounds free-field through her ordinary aid. "M," however, in this condition is unable to label the 200–400 Hz range appropriately. It thus seems likely that one improvement that could be made to conventional aids for some of these patients would be to flatten the frequency response, thus making them more similar to the response of earphones.

Natural vowels in a consonantal context

It thus seems that these profoundly hearing-impaired listeners hear changes in pitch best when they are carried by a sinusoid. At least, they never do *worse* with sinusoids, and can be significantly better. However, reducing the complex speech signal to a single sinusoid potentially involves a serious loss of information. If it could be shown that no use is made of the spectral information that can interfere with the perception of fundamental frequency, sinusoidal presentations would clearly be to at least some patients' advantage in a practical hearing aid.

The spectral distribution of energy plays a complex role in the cueing of speech contrasts (see Reference 26 for a review). Dynamic differences in spectral shape give important cues to many consonantal oppositions. We chose to test the use of spectral cues in vowel identification because different vowels can be signalled by relatively long-lasting and clearly defined differences in spectral envelope. An inability to use spectral information in such a simple form would almost certainly preclude its use in the more perceptually complex consonantal oppositions.

The four words "heed," "hid," "head," and "had" were chosen as representing a reasonable range of vowel qualities. A male Southern British English speaker whose vowels were appropriate for Received Pronunciation (RP) recorded two lists in which each of these words occurred 10 times. The stimuli were randomized subject to the condition that none occurred more than twice in a row. A two-channel tape-recorder was used: on one channel was the acoustic signal from a microphone while on the other was the laryngograph signal. Listeners were asked to perform the task either on the basis of the complete acoustic speech signal, or from a SiVo-type sinusoid derived from the laryngograph signal. In this way, we can determine if the listener is using information about spectral shape, over and above the cues contained in overall duration ("heed" being the longest and "hid" the shortest). Three normal listeners obtained scores of 32.5 percent, 47.5 percent, and 50 percent correct (of 40 trials), when listening to the voice pitch information alone.

Table 6 presents the results for the three listeners

Table 7.
The formant frequencies used in a test of the ability to identify three synthesized vowels.

Vowel	Formant frequencies (Hz) F1	F2	F3
[i]	280	2620	3380
[*a*]	740	1180	2640
[u]	320	920	2200

tested. Only "J" shows any ability to use the spectral information in speech. Interesting too, is the fact that she performed better with her own hearing aids, which may be due to the frequency response of the aids. On the other hand, it may simply be the result of using two ears. In any case, we feel safe in asserting that neither "C" nor "M" is able to make use of the spectral differences between these stimuli to aid identification. "J" may in fact benefit from added information.

Synthetic vowels

It is, of course, still possible that listeners are able to use some spectrally derived information in speech. The natural set of stimuli uses vowel contrasts that are not as large as they could be. Furthermore, the variability in level typical of natural speech may make it difficult to set levels appropriately. To obviate these problems (as well as eliminate any cues other than spectral shape), we also investigated listening performance with a set of synthesized vowels.

In order to make the test as easy as possible, only the three point vowels, [i] as in "heed," [*a*] as in "hard" (Southern English pronunciation) and [u] as in "who'd," were used. These were chosen because they are maximally different both perceptually and acoustically. Formant frequencies appropriate for British RP were obtained from an unpublished thesis by J.C. Wells, cited in Reference 8. They are listed in **Table 7**.

The vowels were synthesized using the OVE IIIc serial-formant speech synthesizer under computer control. All were 300 ms long and had identical falling fundamental frequency contours. See Rosen and Fourcin [27] for further details.

All the stimuli were presented at a comfortable level chosen by the listener. In order to avoid the use of possible loudness differences between the vowels as a help in identification, the stimuli were first adjusted by each listener to be of equal loudness.

Table 8.
Percent correct in the synthetic vowel test. Chance performance would be 33.3 percent.

Listener	Earphone presentation	Sound-field speech via hearing aid
M/L	86.7 (of 30)	73.3 (of 30)
C/R	40.7 (of 150)	41.3 (of 150)
R/L	61.7 (of 60)	63.3 (of 30)
R/R	66.7 (of 30)	
J/R&L	96.7 (of 30)	100.0 (of 30)

Testing sessions consisted of 30 to 75 trials (10 to 25 judgments per stimulus) in a random order. Both free-field and earphone presentations were used to test the possibility that the frequency response of the listener's hearing aid might lead to better performance than the relatively flat response of the headphones. In addition, it anticipated the potential criticism that poor performances with earphones could be due to lack of experience with them. Each session began with an informal practice.

The results are presented in **Table 8**. As in the previous tests, "C" scores lowest, regardless of the method of presentation. Although significantly different from chance ($p = .023$), it is unlikely that this minor ability will serve to any significant extent in the perception of real speech where the spectral patterns are less distinct from one another and are constantly changing in both frequency and amplitude. "R," too, is unlikely to be able to use many of the spectral cues in natural speech, even though she is scoring rather higher. She performs at a lower level than "M," who even with the ability demonstrated here, was not able to make much use of spectral cues in the natural vowel task. Again, "J" stands out from the other listeners in the extent of her residual abilities.

The identification of intervocalic consonants

In another attempt to discover if any significant information was lost by the presentation of sinusoids alone, we used a video-recorded test requiring the identification of 12 common English consonantal sounds spoken intervocalically (e.g., /*aba*,/ /*ata*/, for /m/, /b/, /p/, /v/, /f/, /d/, /n/, /z/, /s/, /t/, /g/,

Table 9.
Percent correct in the identification of 12 intervocalic consonants, done on the basis of the complete speech signal, and a sinusoid at the voice fundamental.*

Listener	Lips alone	Earphone presentation SiVo	Earphone presentation Speech	Sound-field speech via hearing aid
M		79 (of 48)	88 (of 48)	
C	37 (of 92)	62 (of 143)	56 (of 144)	
C		54 (of 144)		47 (of 144)
R	29 (of 143)	74 (of 142)	74 (of 144)	59 (of 142)

*All SiVo presentations used the Mapitch −50 Hz option, except those of "M," who had no mapping. "C" has two lines entered because the two groups of sessions were run with about 8 months intervening.

or /k/) (32). The sounds were spoken by a female speaker who was simultaneously recorded from a microphone and an electro-laryngograph. Sessions consisted of 48 trials (4 occurrences per consonant) in a random order. In all experimental conditions, subjects viewed the video picture. In one condition, no sound was presented. In another, the complete speech signal was presented. Finally, the laryngograph signal was used to trigger a fixed-amplitude sinusoid that followed the voice pitch contour of the speaker, usually with the Mapitch minus 50 Hz option (except where noted).

The first tests with "C" concentrated on headphone presentation, where speech and SiVo-50 signals were presented at a comfortable level. All sessions done by lipreading alone, where learning effects do not occur (28), were completed first; the two sound conditions were alternated every 48 trials. As **Table 9** shows, performance was best with the sinusoid, followed by speech and lipreading alone. The main effect of adding a sound channel was to reduce voicing errors.

We also tested the possibility that "C" might perform better with his own hearing aid in a further experiment where the speech signal was presented free-field. Three such sessions alternated with three sessions in which "C" heard the voice pitch of the speaker presented as a sinusoid mapped down by 50 Hz. Again, performance was better with the sinusoid than with the full speech signal.

"R" participated in a similar experiment in which all four conditions were presented in a random order three times each. Her results are somewhat better than "C"'s (**Table 9**), but similar in character. Presenting the full speech signal leads to no better performance than a single sinusoid.

Of the three subjects tested, only "M" shows any advantage of the full bandwidth speech signal, and its magnitude is small. "J," who appeared to make good use of the full speech signal in vowel identification, has not yet been tested. What may be a more crucial test is one in which the performance of subjects with their hearing aids is compared to their performance when they use SiVo free-field. Such tests are currently underway.

Investigating the aid's usefulness in a "natural" situation

Since our previous tests looked only at restricted aspects of speech perceptual ability, we felt it important to evaluate the SiVo aid with the more realistic connected discourse tracking (CDT) task of De Filippo and Scott (4). In CDT, a speaker reads and the subject attempts to repeat back verbatim. Nothing may be skipped over. Errors made must be corrected. Performance is quantified by counting the number of words transmitted during a 5-minute testing session and expressing the result as a speed in words per minute.

"C" performed CDT with lipreading in two conditions: using his conventional hearing aid and a microphone-based table-top prototype of the SiVo aid. See Rosen and Fourcin [27] for further details. The speaker was a young woman with a fairly high-pitched voice (modal frequency of 247 Hz) so the SiVo aid mapped the pitch contour down by 50 Hz. Testing took place in a quiet, well-lit room with "C" facing the speaker. The aids alternated between each session. **Figure 6** shows the results in graphical form. The mean tracking rate was 32.5 words per minute with "C"'s own aid (10 sessions) and 44.8

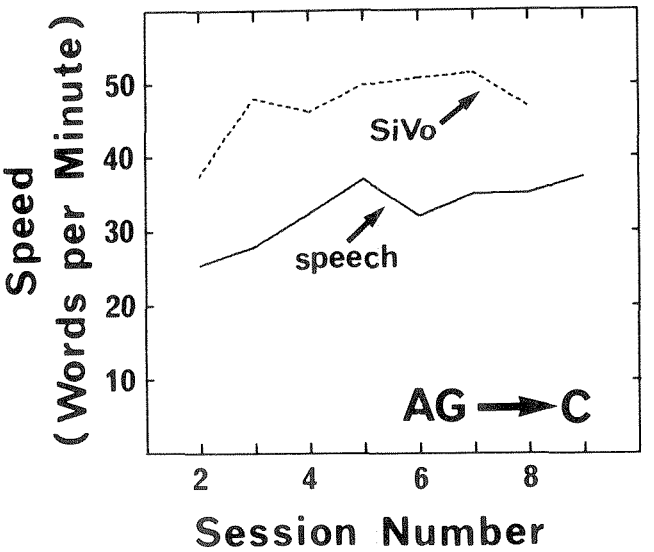

Figure 6.
Three-point smoothed rates obtained by "C" in connected discourse tracking, as a function of session number. The solid line indicates his performance when listening to speech through his own hearing aid, while the dashed line indicates his performance when using a table-top microphone-based version of the SiVo aid using the Mapitch −50 Hz option. In both conditions "C" was also lipreading the speaker.

words per minute with the SiVo aid (9 sessions), a proportional increase of 38 percent. Similar studies are planned for the other listeners.

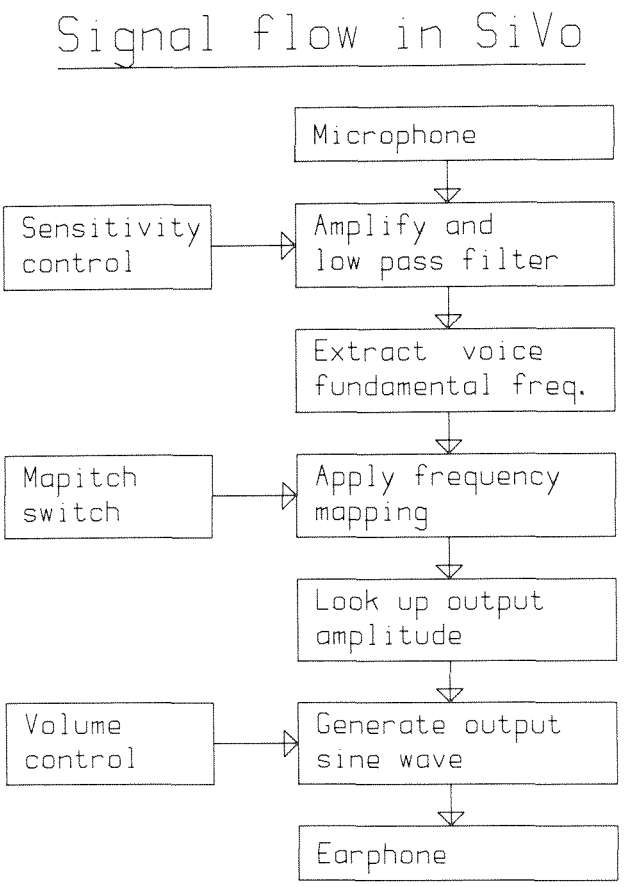

Figure 7.
A signal flow diagram for a portable SiVo.

Summary

The evidence seems clear that, at least for some of these listeners, simplifying the speech signal to a single sinusoid has considerable advantages. Both psychoacoustic and speech-perception tests show that fundamental frequency changes are typically best signalled with sinusoids. Furthermore, the information that is lost in this way does not seem to be of much use to most of the listeners. These results encouraged us to develop a wearable SiVo that could be used daily.

A WEARABLE SiVo AID

The principal requirements for a practical device are that it must receive a speech input and generate an acoustic sine wave output at a sound level the patient can comfortably hear and that follows the pattern of the voice fundamental frequency. In many instances it is desirable to map the output frequency downwards, especially when listening to the higher pitched voices of women and children. The device should respond rapidly to fluctuations in fundamental, preferably on a period-by-period basis. It should, of course, be as small and light as possible with minimum power consumption.

Figure 7 shows the main stages involved. First of all, the speech input is received by a microphone and amplified to suit the following circuitry. The user of the device adjusts a "sensitivity control" to separate the wanted speech signal from the unwanted background noise. The signal then passes to a circuit that extracts the voice fundamental frequency from the incoming speech. The fundamental frequency can be modified by the Mapitch switch, which is controlled by the user. The frequency mapping is usually a simple subtraction of 50 Hz in switch position "b," 80 Hz in position "c" or no change in position "a." In some cases, however, other functions such as frequency range expansion and limiting may be desirable. When the

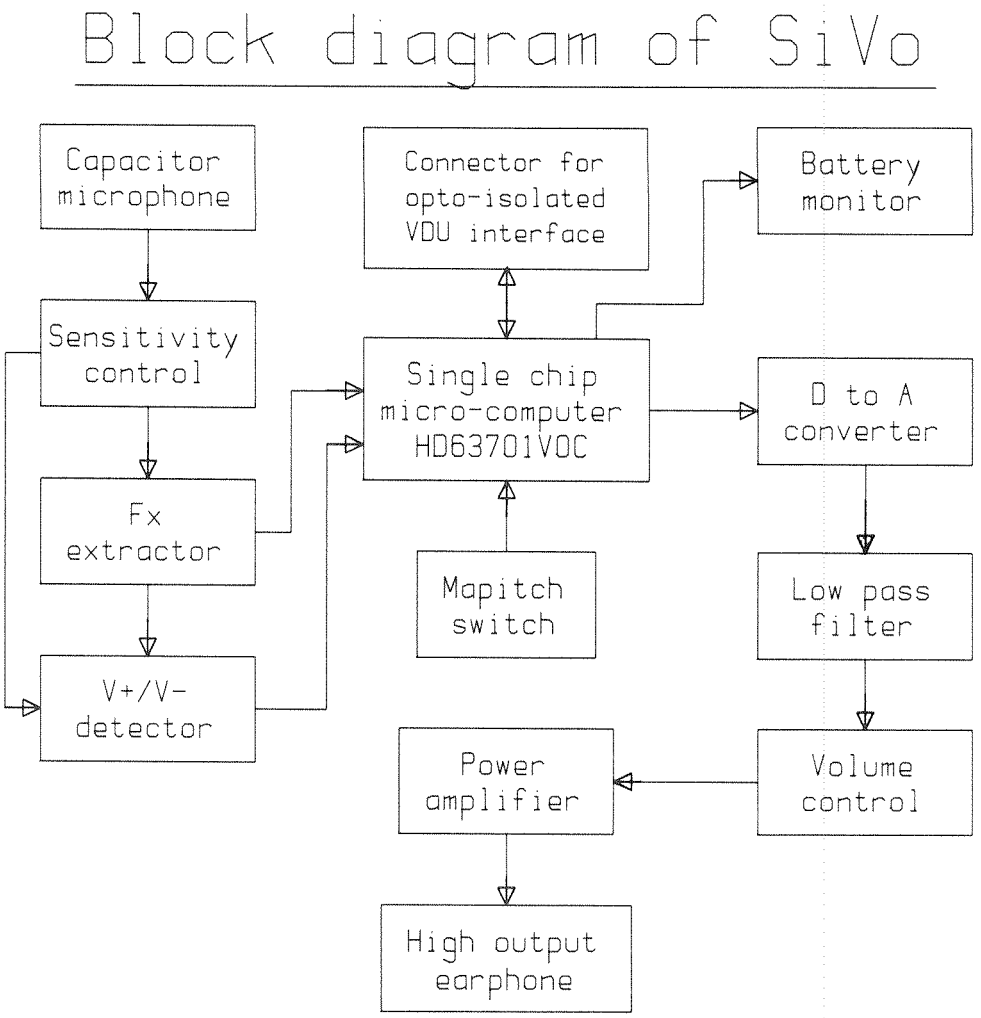

Figure 8.
A block diagram of a portable SiVo.

mapping is complete the system must generate an acoustic sine wave at the correct output amplitude and frequency.

Implementation

Both analog and digital techniques can be used to construct a SiVo device. Our earlier prototypes were based on completely analog circuitry. While these effectively demonstrated the benefits of SiVo, they were relatively bulky because of the large number of individual components used. It also proved difficult to configure analog circuits to match the widely varying requirements of the patients.

The present implementation is based on a hybrid of analog and digital techniques, in which each is used for the functions best suited to it. The overall block diagram is shown in **Figure 8**.

Fundamental frequency extraction is performed by an analog circuit (14). The speech signal is low-pass filtered to remove the bulk of the formant and fricative information, compressed with a logarithmic amplifier, and passed through two peak-picking circuits in order to detect the principal peaks in the speech pressure waveform associated with vocal fold closure. **Figure 9** shows the sequence of progressively simplified waveforms obtained at various stages in the circuit for a naturally uttered "ah" starting with the microphone signal and ending with output pulses in synchrony with the speaker's vocal fold closures.

Quite apart from its simplicity, the peak-picking method has several important advantages over cepstrum (19) and Gold-Rabiner (9) processors, which deal with relatively long blocks of data and destroy the fine temporal structure within them. By pre-

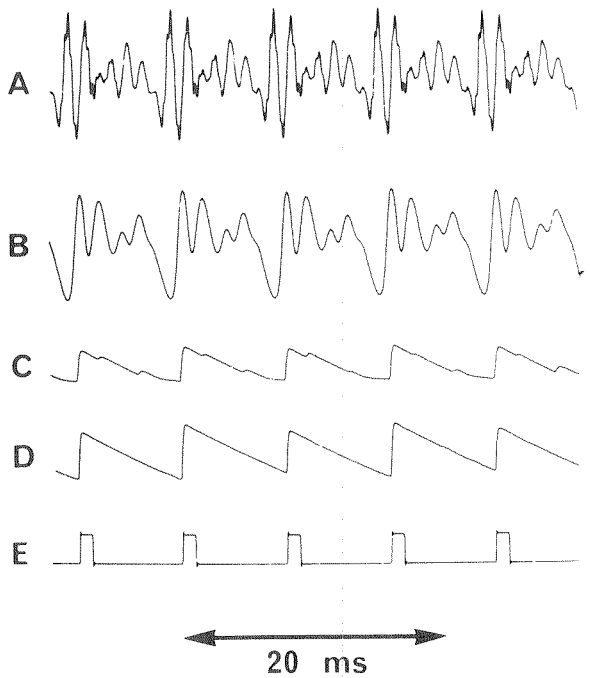

Figure 9.
Internal waveforms from the fundamental frequency extractor. The microphone-sensed speech signal (A) is low-pass filtered and compressed (B), passed through two peak-pickers (C,D), differentiated and clipped (E) before triggering the microcomputer.

serving the detailed time patterning of speech, the characteristic irregularities of creaky voice, for example, may be heard as such by the deaf listener. It has already been shown, in a quiet room with microphone close to the speaker, that the "peak-picked" voice pitch is a better aid to lipreading than one derived from a cepstrum (30).

The pulses from the fundamental frequency extractor are fed into the timer input of an HD63701VOC single-chip CMOS microcomputer. This measures the intervals between pulses to very high precision and tests them to ensure that they are sensible. The computer examines the state of the front panel Mapitch switch to determine whether frequency mapping is required. If so, this is done by table lookup using data stored in erasible-programmable read-only memory (eprom) on the same chip as the microcomputer itself. The use of data tables means that arbitrary frequency mapping functions can be applied simply by modifying the stored data. The computer then looks up another data table containing the relation between output frequency and output sound pressure level. This table is either specially prepared for each individual patient or selected from a library of previously used tables.

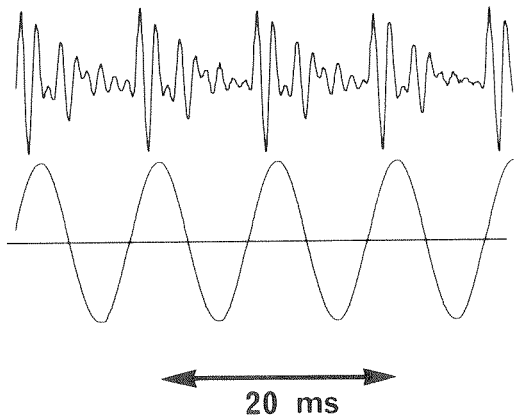

Figure 10.
Speech sound pressure waveform (top) and acoustic output from SiVo (bottom) for the token "ah."

There is now enough information to synthesize a sine wave output using a digital-to-analog (D/A) converter. The output is generated on a cycle-by-cycle basis, always starting and ending at zero crossings. The processing delay in the system is never more than one output cycle, so the device responds well to speech irregularities such as creaky voice.

The stepped sine wave from the D/A converter is lowpass filtered at about 600 Hz, to remove most of the energy at the sampling frequency and its harmonics, before being fed to the user's volume control. The output from this drives an anti-phase pair of constant current output amplifiers connected to the center-tapped high output earphone.

Figure 10 shows the speech sound pressure and earphone acoustic output waveforms for the token "ah." In **Figure 11** the input and output are shown

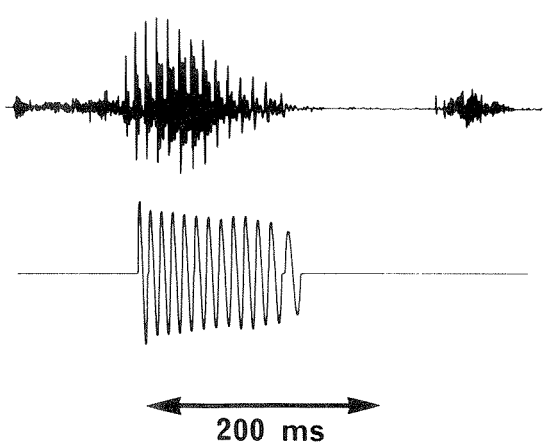

Figure 11.
Speech sound pressure waveform (top) and acoustic output from SiVo (bottom) for the word "coat."

for the word "coat." The amplitude of the sine wave can be seen to vary according to the stored frequency response table (in this case a rising 6 dB/octave emphasis). Note that there is no output corresponding to the unvoiced parts of "coat" (the aperiodic energy associated with the initial [k] and final [t] burst + aspiration). The possibility of signalling this additional feature has already been designed into the present device as indicated by the voiced/voiceless (V+/V−) detector in **Figure 8**. (For further details, see below).

Matching SiVo to the patient

It is very important that the output intensity of the earphone is matched to the listener, who, as seen in **Figure 2**, will often have a very limited dynamic range. For clinical use, conventional audiometers can provide useful guidance in this respect, but there are several problems with their use. The standard frequencies provided by many audiometers are spaced at octave intervals, which gives only a coarse sampling of what may often be a very steeply sloping response. The calibration is subject to considerable uncertainty at the lower frequencies where acoustic leakage around earphones can be a serious problem. The calibration of hearing aid earphones is also subject to error because of the effect of earmold leakage, variable ear canal excise, volume, and the variations between individual transducers. Our laboratory set-up using variable oscillators and earphone-mounted microphones avoids the problems of earphones and audiometers, but not those of hearing aid earphone calibration.

In view of these problems, and particularly for simplicity in the clinic, it is preferable to carry out measurements with closely spaced test tones presented by the patient's own SiVo aid with the earphone and earmold he will ultimately use. In this way, all cross-calibration errors are eliminated. Fortunately, the microcomputer within SiVo is readily able to generate the necessary test signals under external control. A special socket has been provided within the battery compartment for the connection of an external visual display unit (vdu) or desk-top microcomputer via an optical isolator which ensures the patient's electrical safety.

The testing software displays a frequency response curve on the vdu which can be edited with cursor keys. Pressing the space bar generates a test tone at the selected frequency and amplitude. By stepping through the range of frequencies, a good approximation to the "most comfortable level" curve is generated. Better matching can, however, be obtained by comparing each test frequency with a reference level at, for example, 125 Hz.

In the present prototype devices, the operator compares the response obtained with a predefined library of curves stored within the microcomputer and selects the best match using an internal switch. Additional curves can be generated and stored in the memory of the single chip microcomputer if none of the existing ones match the requirements.

Developments in the short-term

The existing system is a prototype for field evaluation of SiVo. "M" and "C" have already been equipped with the device, and use it daily, while continuing to use their behind-the ear aids in certain situations. Although they have experienced the expected problems (difficulties in noisy environments, or with more than one speaker; working out where to wear a relatively bulky device), their overall impressions are favorable.

The development of a smaller and lighter device is well under way. While the present unit makes extensive use of surface mounted components, many are still conventionally packaged. We expect most of these to become available in surface mounted form shortly. The software has been modified to run on a new surface-mounted CMOS microcomputer, (MC 68HC11). This will bring many advantages because not only does it include an analog-to-digital converter and many other components, but it also has an electrically erasible programable read-only memory (EEPROM).

The use of EEPROM to store the operating program and patient-specific data tables will greatly increase the device's versatility and ease of fitting to the needs of the patient. Furthermore, the device can be loaded with diagnostic programs to help locate any problems that might occur either in production or field maintenance.

FINAL REMARKS AND FUTURE DIRECTIONS

Our findings show that the SiVo aid, as it stands, will prove to be of benefit to a significant number of profoundly hearing-impaired listeners. Even so, one of the advantages of our approach is that it

allows a principled path to future improvements. These fall into two main categories.

Firstly, for those listeners who have sufficient residual auditory function, more speech-pattern elements may be included. We have already noted that even profoundly hearing-impaired listeners can distinguish between periodic and aperiodic stimulation, and this provides an avenue for signalling information about voiceless speech sounds. The current devices have been designed to allow the inclusion of a voiceless sound detector that will trigger the generation of pseudo-random noise in a low frequency band clearly audible to the patients when high frequency noise is detected. It may be possible in the appropriate patients to signal broad classes of fricatives by varying the spectral content of the aperiodic stimulation.

The inclusion of amplitude variation would be an additional helpful feature for those with sufficient dynamic range and sensitivity to loudness variations (10, 21). The analog-to-digital converters used for detecting the voiceless information will also allow dynamic envelope control in conjunction with stored data tables.

For the best patients of this group ("J," for example), it may be possible to signal some aspects of spectral shape in voiced sounds, using a combination of harmonic amplitude and phase manipulations. Our initial explorations will focus on the provision of information about F1. An aid which included all these features would no longer be of use only to those with the most profound hearing impairments, so the number of appropriate patients for these more complex aids will be much larger. We thus envisage a range of aids, based on the same ideas, catering for the range of patient performances obtained in the profoundly hearing-impaired group. Alternatively, there could be one aid whose characteristics can be configured to match the patient's abilities.

Secondly, there are many improvements to be made in our methods of speech-pattern element feature extraction, which are neither very robust for distant speakers nor in the presence of background noise. It is likely that these will be substantially improved through the development of algorithms which make greater use of our knowledge about the properties of speech, and through the application of higher performance low-power processors and application-specific integrated circuits. Given the problems with background noise that listeners who are hearing-impaired to any degree typically report, the greatest advantage of a speech pattern approach may be in the theoretical possibility of providing noise-free speech signals.

ACKNOWLEDGMENTS

The authors form part of the External Pattern Input (EPI) Group, and wish to thank their colleagues who have greatly influenced the course of this work. Special gratitude is expressed to Alison Gill for serving as the speaker in the CDT test, Brian Moore and Andrew Faulkner for comments on the manuscript, and David Smith for assistance in analyzing results from the 12 intervocalic consonant test. We are, however, most deeply indebted to our four listeners, without whose patient and informative cooperation this work could not proceed.

REFERENCES

1. **Abberton E and Fourcin AJ:** Electrolaryngography. In: C Code and M Ball (Eds), *Experimental Clinical Phonetics*, London: Croom Helm, 1984.
2. **Bock RD and Jones LV:** *The Measurement and Prediction of Judgment and Choice*. San Francisco: Holden-Day, 1968.
3. **Darwin CJ and Gardner RB:** Mistuning a harmonic of a vowel: grouping and phase effects on vowel quality. *J Acoust Soc Am* 79:838–845, 1986.
4. **De Filippo CL and Scott BL:** A method for training and evaluating the reception of on-going speech. *J Acoust Soc Am* 63:1186–1192, 1978.
5. **Fastl H and Weinberger M:** Frequency discrimination for pure and complex tones. *Acustica* 49:77–78, 1981.
6. **Fourcin AJ and Abberton E:** First applications of a new laryngograph. *Med Bio Illus* 21:172–182, 1971. Also, *Volta Rev* 74:161–176, 1972.
7. **Fourcin AJ, Rosen S, Moore BCJ, Douek EE, Clarke GP, Dodson H, and Bannister LH:** External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings. *Brit J Audiol* 13:85–107, 1979.
8. **Gimson AC:** *An Introduction to the Pronunciation of English*. London: Edward Arnold, 1968.
9. **Gold B and Rabiner L:** Parallel processing techniques for estimating pitch periods in the time domain. *J Acoust Soc Am* 46:442–448, 1969.
10. **Grant KW, Ardell LH, Kuhl PK, and Sparks DW:** The

contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *J Acoust Soc Am* 77:671–677, 1985.

11. **Henning GB and Grosberg SL:** Effect of harmonic components on frequency discrimination. *J Acoust Soc Am* 44:1386–1389, 1968.
12. **Hoekstra A:** Frequency discrimination and frequency analysis in hearing. Ph.D. Thesis, University of Groningen, 1979.
13. **Hoekstra A and Ritsma RJ:** Perceptive hearing loss and frequency selectivity. In: EF Evans and JP Wilson (Eds), *Psychophysics and Physiology of Hearing*, London: Academic Press, 1977.
14. **Howard DM and Fourcin AJ:** Instantaneous voice period measurement for cochlear stimulation. *Elect Lett* 19:776–778, 1983.
15. **Levitt H:** Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477, 1971.
16. **Licklider JCR:** Effects of changes in the phase pattern upon the sound of a 16-harmonic tone. (abstract) *J Acoust Soc Am* 29:780, 1957.
17. **Moore BCJ, Fourcin AJ, Rosen S, Walliker JR, Howard D, Abberton E, Douek EE, and Frampton S:** Extraction and presentation of speech features. In: RA Schindler and MM Merzenich (Eds), *Cochlear Implants*, New York: Raven Press, 1985.
18. **Moore BCJ and Glasberg BR:** The role of frequency selectivity in the perception of loudness, pitch and time. In: BCJ Moore (Ed), *Frequency Selectivity in Hearing*. London: Academic Press, 1986.
19. **Noll AM:** Cepstrum pitch determination. *J Acoust Soc Am* 41:293–309, 1967.
20. **Noorden L van:** Two channel pitch perception. In: M Clynes (Ed), *Music, Mind and Brain*, New York: Plenum Press, 1982.
21. **Risberg A:** The importance of prosodic speech elements for the lipreader. In: HB Nielsen and E Kampp (Eds), *Visual and Audio-Visual Perception of Speech*. Sixth Danavox Symposium, Scandinavian Audiology, Supplementum 4. Stockholm: Almquist and Wiksell, 1974.
22. **Risberg A:** Periodic-nonperiodic test of hearing capacity. *6th International Congress on Acoustics*. Tokyo: B-219 to B-222, 1968.
23. **Rosen S:** Hyperacute monaural phase sensitivity in the hearing-impaired (abstract). *Brit J Audiol* 18:257–258, 1984.
24. **Rosen S:** Monaural phase sensitivity: Frequency selectivity and temporal processes. In: BCJ Moore and RD Patterson (Eds), *Auditory Frequency Selectivity*, New York: Plenum, 1986.
25. **Rosen S:** Phase and the hearing impaired. In: MEH Schouten (Ed), *The Psychophysics of Speech Perception*." Dordrecht, The Netherlands: Martinus Nijhoff, 1987.
26. **Rosen S and Fourcin AJ:** Frequency selectivity and the perception of speech. In: BCJ Moore (Ed), *Frequency Selectivity in Hearing*. London: Academic Press, 1986.
27. **Rosen S and Fourcin, AJ:** When less is more: further work. Speech, Hearing and Language: Work in Progress. *Phonetics & Linguistics*, University College London, 1:3–27, 1983.
28. **Rosen S, Fourcin AJ, Abberton E, Walliker JR, Howard DH, Moore BCJ, Douek EE, and Frampton S:** Assessing Assessment. In: RA Schindler and MM Merzenich (Eds), *Cochlear Implants*, New York: Raven Press, 1985.
29. **Rosen S, Fourcin AJ, and Moore BCJ:** Voice pitch as an aid to lipreading. *Nature* 291:150–152, 1981.
30. **Rosen S, Fourcin AJ, Walliker JR, Douek EE, and Moore BCJ:** External electrical stimulation of the cochlea in the totally deaf. In: J. Raviv (Ed), *Uses of Computers in Aiding the Disabled*. Amsterdam: North-Holland, 1982.
31. **Rosen S, Moore BCJ, and Fourcin AJ:** Lipreading connected discourse with fundamental frequency information. *Brit Soc Audiol Newsletter*. Summer Issue:42–43, 1980.
32. **Rosen S, Moore BCJ, and Fourcin AJ:** Lipreading with fundamental frequency information. *Proceedings of the Institute of Acoustics Autumn Conference 1979*, Windermere. Paper 1A2:5–8, 1979.
33. **Sachs MB and Young ED:** Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. *J Acoust Soc Am* 66:470–479, 1979.
34. **Schroeder MR:** New results concerning monaural phase sensitivity (abstract) *J Acoust Soc Am* 34:1579, 1959.
35. **Snedecor GW, and Cochran WG:** Statistical Methods. Ames, Iowa: Iowa State University Press, 1978.
36. **Stock D and Rosen S:** Frequency discrimination and resolution at low frequencies in normal and hearing-impaired listeners. Speech, Hearing and Language: Work in Progress. *Phonetics & Linguistics*, University College London, 2:193–222, 1986.
37. **Thornton ARD:** Estimation of the number of patients who might be suitable for cochlear implant and similar procedures. *Brit J Audiol* 20:221–229, 1986.
38. **Walliker JR, Rosen S, and Fourcin A:** Speech pattern prostheses for the profoundly and totally deaf. In: *IEE Conference Publication No. 258*, International Conference on Speech Input/Output; Techniques and Applications, London, England, 194–199, 1986. Also in: Speech, Hearing and Language: Work in Progress. *Phonetics & Linguistics*, University College London, 2:223–236, 1986.
39. **Young ED and Sachs MB:** Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J Acoust Soc Am* 66:1381–1403, 1979.