SOUND SEPARATION WITH A COCHLEAR IMPLANT AND A HEARING AID IN OPPOSITE EARS

Peter J. Blamey[1], Christopher J. James[1] & Lois F.A. Martin[2]

[1]Department of Otolaryngology, University of Melbourne
[2]Bionic Ear Institute

ABSTRACT – Two experiments were conducted to investigate the perception of speech and noise presented simultaneously to three subjects with impaired hearing in five monaural and binaural conditions. A broadband noise was found to have no effect on speech perception when the two signals were presented to opposite ears. When speech and noise were presented to the same ear(s), speech perception scores on a closed-set test fell from above 95% at high signal-to-noise ratios (SNR) to 71% at an SNR of about –5 dB. When two speech signals were presented simultaneously at equal intensities (0 dB SNR) speech perception scores fell to 75% or lower, regardless of the ear(s) to which the signals were presented. Thus dichotic presentation helped these listeners to separate speech from a broadband noise, but not to separate two simultaneous speech signals produced by different speakers.

INTRODUCTION

Under normal conditions, listeners hear sounds from more than one source at a time, such as a voice and environmental noise, or several voices speaking at the same time. At least three distinct physical characteristics of the signals are used to separate these sounds perceptually (spatial, spectral, and temporal). Spatially separated sources can be localised and distinguished from one another using interaural timing differences at low frequencies and interaural intensity differences at high frequencies (Rayleigh, 1907). There are also perceptual mechanisms that can be used to distinguish sounds that are heard simultaneously with only one ear. For example, these mechanisms make use of simultaneous onsets and comodulation of different frequency components to form separable streams of auditory information (Bregman, 1990). Another example is the separation of complex sounds that have different fundamental frequencies as in the case of human voices (Darwin & Gardner, 1986).

The focus of the present study is the use of temporal and spectral cues for the separation of sounds presented binaurally or monaurally to listeners who use a cochlear implant in one ear and a hearing aid in the other. With impaired hearing, both of these cues may be degraded by poor temporal and/or spectral resolution in the monaural condition. In the binaural condition there may be additional complications if the hearing loss is asymmetrical and/or there are different temporal delays in the processing in the two ears. The cases to be considered in this report are particularly interesting because the acoustic presentation of sound via a hearing aid to one ear and the electric presentation of the same sound via a cochlear implant to the other ear is almost certain to introduce both temporal and spectral differences between the ears. It is also unlikely that cochlear implant users will have access to fine spectral cues such as those needed to group together harmonically related frequency components.

Without entering into a detailed consideration of proposed models and mechanisms of binaural and monaural separation of sounds, two hypotheses seem plausible: a) that two simultaneous sounds presented diotically (both sounds to both ears) will be more difficult to separate than the same two sounds presented dichotically (one sound to each ear), b) that two simultaneous sounds with generally similar spectral and temporal characteristics (such as two voices uttering a word) will be more difficult to separate than grossly different sounds (such as a voice and a broadband noise).

METHOD

Participants and processors

The three participants in this pilot study were post-linguistically deafened adults who used a multiple-electrode (Cochlear Limited) cochlear implant in one ear, and who had residual hearing in the non-implanted ear. Table 1 summarises the relevant audiological information for the non-implanted ear of each participant, together with some of the factors known to have an effect on speech perception scores for individual cochlear implant users (Blamey et al, 1992, 1996). Two of the participants normally used a hearing aid and cochlear implant together, while the other wore only the cochlear implant. For the purposes of this experiment, all participants were fitted with a benchtop hearing aid based on a Motorola DSP 56303 evaluation module with additional microphone and amplifier circuits to drive an Oticon AN270 button receiver in the non-implanted ear. If the participant normally wore a hearing aid, the fixed linear gain of the benchtop hearing aid was set to equal the gain of the participant's own aid (within 2 dB) across the frequency range from 125 Hz to 4 kHz as measured in a hearing aid test box. If the participant did not normally wear a hearing aid, the fixed linear gain of the benchtop hearing aid was set according to the NAL prescription with correction for severe and profound hearing losses (Byrne & Dillon, 1986). The implant signals for this study were presented via the participants' own Sprint cochlear implant speech processors with their usual speech processing strategies as listed in Table 1.

Participant	S1	S2	S3
Hearing Loss at 250 Hz	65 dB HL	75 dB HL	30 dB HL
500 Hz	105 dB HL	75 dB HL	60 dB HL
1 kHz	110 dB HL	85 dB HL	120 dB HL
2 kHz	115 dB HL	120	No response
4 kHz	120 dB HL	No response	No response
Hearing Aid Use	Yes	No	Yes
Hearing Aid Ear	Left	Left	Right
Age	66 years	48 years	65 years
Aetiology	Progressive	Genetic	Progressive
Duration of hearing loss	22 years	25 years	16 years
Implant experience	7y 5m	5y 3m	1y 1m
No of electrodes in use	16	19	6
Mean dynamic range	40.6 SL	38.4 SL	31.9 SL
Speech processor strategy	Mpeak	ACE	CIS
Implant Model	CI-22	CI-24M	CI-24M

Table 1. Summary of audiological details for the participants.

Separation of two voices

In the first experiment, speech recognition scores for the two voices were used as a comparative measure of sound separation in the five experimental conditions, on the assumption that better sound separation would result in higher word recognition scores. Stimuli were spondees (bisyllabic words with equal stress on each syllable) spoken by two different speakers: a male adult and a female adult who are also the second and third authors of this report. These words were chosen from a set of 40 spondees that had already been recorded for each of the two speakers, and were selected to be easy to distinguish from one another in both of the monaural conditions to be tested ie hearing aid alone and cochlear implant alone. This was necessary because two of the participants had very poor speech recognition in the hearing aid ear. The RMS levels of the digitally recorded speech tokens were equalised. Each word was stored in a single channel of a wave file with the onset of the word within 10 ms of the start of the file.

Computer software and hardware were set up so that the stimuli could be mixed and presented via a line input directly to the benchtop hearing aid and the cochlear implant speech processor to produce the monaural, diotic, and dichotic signals required for this study. Five conditions were tested: HA in which the two voices were mixed and presented to the hearing aid only; CI in which the two voices were mixed and presented to the cochlear implant only; DIOTIC in which the two voices were mixed and presented to both the cochlear implant and the hearing aid simultaneously; MCIFHA in which the male voice was presented to the cochlear implant and the female voice to the hearing aid dichotically; and FCIMHA in which the female voice was presented to the cochlear implant and the male to the hearing aid dichotically.

Prior to starting the experiment, the individual spondees spoken by each speaker were presented in the HA, CI and DIOTIC conditions in a practice procedure. The input levels to the hearing aid and the cochlear implant speech processor were adjusted individually so that the loudness of the stimuli was equal in each ear, with the overall loudness at a comfortable listening level. It was also checked that each participant could recognise which speaker presented each of the spondees, and that they could recognise the individual spondee presented. This was done by presenting a block of stimuli (4 of each spondee) in a random order and asking the participant to select the spondee spoken from a list. If the participant scored over 95% correct in all three of the HA, CI and DIOTIC conditions, the participant was judged to be ready for the more difficult experiment involving simultaneous presentations of the two voices. Participant S2 was able to perform this task with a set of 10 spondees (5 for each speaker). S1 and S3 performed the task with 6 spondees (3 for each speaker).

The combined stimuli were presented to S1 and S3 in blocks of 18 (2 x 9 combinations) and to S2 in blocks of 50 (2 x 25 combinations) in a random order. Two seconds before each presentation trial, either the male words or the female words were displayed on a computer screen, together with a heading "MAN'S VOICE" or "WOMAN'S VOICE," respectively. After the trial, the participant was asked to respond with the word from the list on the screen that had been spoken by the indicated speaker. For each of the combinations within a block, the participant was asked once for the male speaker's word and once for the female speaker's word. Two to five blocks of trials were presented to each participant in each of the five conditions listed above.

Results for the first participant indicated that the diotic score was slightly higher than the two dichotic scores, contrary to hypothesis a). It seemed possible that the participant may have had difficulty in switching his attention rapidly from one ear to the other or from one voice to the other, so the task was repeated with a different blocking structure. Blocks of 27 stimuli were presented in each condition (3 x 9 combinations). Within each block, the participant was always asked to respond with the same speaker's word ie always the male speaker, or in another block always the female speaker. Thus the participant did not need to switch his attention between ears or between speakers within a block of stimuli.

Separation of a voice and a noise

In this experiment, the relative levels of speech and noise were varied adaptively to find the signal-to-noise ratio (SNR) where 71% of the words were recognised correctly. It is assumed that separation of the speech (and recognition of the speech) becomes more difficult as the SNR decreases. The stimuli were three spondees spoken by the female speaker with the carrier phrase "The next word is ...". The same three spondees were used as in the first experiment ie "teapot", "drawbridge" and "football", but these were different tokens recorded with the carrier phrase. Each recorded stimulus was set to the same RMS amplitude. The stimuli were presented in a persistent background of speech-shaped broadband noise under 7 different conditions. As in the first experiment HA, CI and DIOTIC conditions were used, where the speech and noise were mixed and presented to one or both ears. In the dichotic NCIFHA condition, noise was presented to the implanted ear and the female voice to the hearing aid. In the dichotic FCINHA condition, the noise was presented to the hearing aid and the woman's voice to the implanted ear. The remaining conditions (HA0 and CI0) were carried out with no noise and the voice presented to the hearing aid or cochlear implant, respectively.

The SNR was varied in each condition using an adaptive procedure in which the SNR was increased by 2 dB every time the listener responded incorrectly, and decreased by 2 dB after two correct responses in a row (Levitt, 1971). This up-down procedure oscillates about the SNR where the listener scores 71% correct. The chance score in this 3-alternative-forced-choice task is 33%. The numerical values of SNR refer to the ratios of RMS amplitude for the speech and the speech-shaped noise. The adaptive procedure was terminated after 6 turning points had been found, and the average of the last four turning points was taken as the asymptotic SNR. Because it was expected that the noise might have little effect in the dichotic conditions and we did not wish to present uncomfortably loud sounds, the SNR was reduced by reducing the speech level when the SNR was negative, and by increasing the noise level when the SNR was positive. Thus the procedure would typically start with the speech at the comfortable level. The adaptive procedure would increase the noise until it was at the same RMS level as the speech (0 dB SNR) and then the level of the speech would start to decrease. In the HA0 and CI0 conditions, the level of the speech was decreased to find a speech reception threshold with no noise.

RESULTS

Separation of two voices

The two different blocking methods produced no significant differences in the results obtained in any condition or for any subject, so the results were combined. Figure 1 shows the percentage of correct responses for each subject in each of the 5 conditions. It is clear that the listeners were unable to separate the two voices completely as the scores drop significantly below 95% in all conditions. On the other hand, the scores are all significantly above the chance score of 33%. Analysis of Variance (ANOVA) indicated that there were no significant differences between any of the five conditions shown in Figure 1. The ANOVA used subject, condition, and blocking method as independent variables.

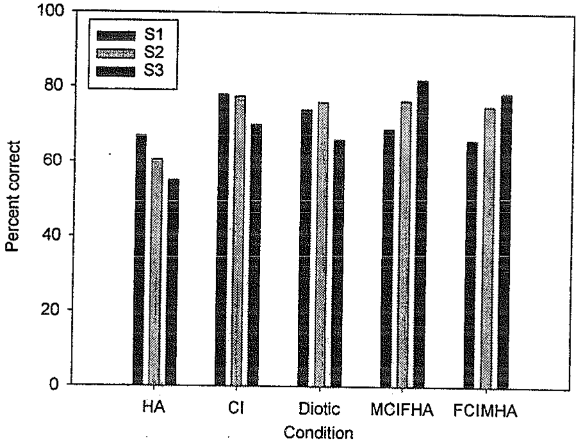

Figure 1. Percentages of correct responses by subject and condition in the separation of voices experiment.

Separation of a voice and a noise

Figure 2 shows the asymptotic SNR values for each subject in each of the 7 conditions tested. Each value shown is the mean of two or more adaptive procedures. For positive SNRs, the speech is at a comfortable loudness and the RMS level of the noise is lower than the RMS level of the speech by the indicated number of dB. For negative SNRs, the noise is at a comfortable loudness, and the RMS level of the speech is lower than the RMS level of the noise by the indicated number of dB. High SNR indicates poor separation of speech and noise. Low SNR indicates good separation of speech from noise.

ANOVA with subject and condition as independent variables was followed by post-hoc t-tests using the Bonferroni method to compare the SNRs in the different conditions. The mean SNRs for the CI, Diotic, and HA conditions were not significantly different from one another ($p>0.05$). The mean SNRs for the FCINHA, NCIFHA, CI0, and HA0 conditions were not significantly different from one another ($p>0.05$). However all the SNRs for the first 3 conditions were significantly different from all the SNRs in the second group ($p<0.001$).

DISCUSSION

Hypothesis a) that it is more difficult to separate sounds diotically than dichotically.

Hypothesis a) predicts that the speech perception scores in the voice separation experiment should be higher for the dichotic conditions than the diotic condition, and the SNRs for the dichotic conditions should be lower than for the diotic condition in the speech and noise experiment. The hypothesis was supported by the results when one of the sounds was a broadband noise and the other was speech. The hypothesis was rejected when the sounds were both speech signals. The result for speech and noise is consistent with masking experiments in which a masker has a much greater effect on a probe in the same ear than on a probe in the opposite ear (Zwislocki, 1972). When two speech signals are presented together, interference takes place regardless of the ear(s) of presentation. This result is not consistent with known monaural and binaural masking effects. It is more consistent with the binaural experiments of Studdert-Kennedy and Shankweiler (1970), which suggest that speech features from the two ears are processed independently and then recombined at one location in the left hemisphere (of almost all right-handed listeners and most left-handed listeners). The interference between the two signals probably occurs at the recombination stage where there is usually a small right-ear advantage, but otherwise the speech features derived from the two ears are treated equivalently. In the dichotic conditions of the two-voice experiment, S1 and S3 had higher scores for the words presented to the hearing aid, and S2 scored higher for words presented to the implant. This was a right ear advantage for S2 and S3, and a left ear advantage for S1.

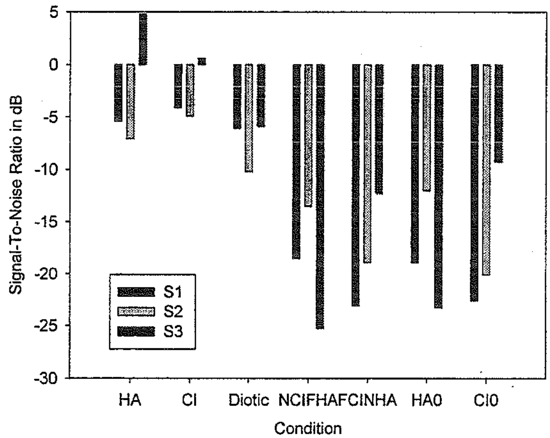

Figure 2. Mean asymptotic signal-to-noise ratio in each condition of the separation of speech and noise experiment for each subject.

Hypothesis b) that it is it is more difficult to separate two voices than to separate a voice and a broadband noise.

This hypothesis was supported by the results. In fact, the equality of the SNRs in the NCIFHA and FCINHA conditions with the HA0 and CI0 conditions, respectively, demonstrates that the broadband noise had no measurable effect on the perception of the speech signal when the voice and the noise were presented to opposite ears. In the HA, CI, and Diotic conditions, the mean SNRs were negative for a level of 71% correct. Thus the scores at zero dB would have been greater than 71% at zero dB SNR. The percentages correct were 59%, 75%, and 72%, respectively in the two-voice experiment (at zero dB signal to noise ratio).

It is interesting to note that Armstrong et al (1997) found a significant advantage for diotic listening with an implant and hearing aid together over monaural listening with a cochlear implant for open-set sentence perception in quiet and in 8-talker babble at an SNR of 10 dB. In the present study, there were no statistically significant differences between the diotic and monaural conditions in either experiment. The discrepancy between the studies may be due to the different materials and noise used, or the low number of subjects in the present study.

CONCLUSIONS

The separation of sounds into streams by the auditory system involves different mechanisms for dynamic speech signals and stationary broadband noises. These effects which are observed in normally hearing listeners are also present in subjects who have a hearing aid in one ear and a cochlear implant in the other. The separation and fusion of sounds presented to different ears, and the potential advantage of one ear or one device over the other may have consequences for the development of binaural processors for people with impaired hearing.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the National Health and Medical Research Council of Australia (project grant 970257), the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and the Bionic Ear Institute. The research was conducted under the auspices of the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (project 96/289H). The authors also gratefully acknowledge the valuable contribution of the participants who volunteered their time for these experiments.

REFERENCES

Armstrong, M., Pegg, P., James, C. & Blamey, P.J. (1997) *Speech perception in noise with implant and hearing aid*, American Journal of Otology 18, S140-S141.

Blamey, P.J., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindstrom, B., Nedzelski, J., Peterson, A., Shipp, D., Staller, S. & Whitford, L. (1996) *Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants*, Audiology & Neuro-Otology 1, 293-306.

Blamey, P.J., Pyman, B.C., Gordon, M., Clark, G.M., Dowell, R.C. & Hollow, R.D. (1992) *Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients*, Annals of Otology, Rhinology, & Laryngology 101, 342-348.

Bregman, A.S. (1990) *Auditory scene analysis*, (MIT Press: Cambridge, MA).

Byrne, D. & Dillon, H. (1986) *The National Acoustics Laboratory (NAL) new procedure for selecting the gain and frequency response of a hearing aid*, Ear & Hearing 7, 257-265.

Darwin, C.J. & Gardner, R.B. (1986) *Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality*, Journal of the Acoustical Society of America 79, 838-845.

Levitt, H. (1971) *Transformed up-down methods in psychoacoustics*, Journal of the Acoustical Society of America 49, 467-477.

Lord Rayleigh (1907) *Our perception of sound duration*, Philosophical Magazine 13, 214-232.

Studdert-Kennedy, M. & Shankweiler, D. (1970) *Hemispheric specialization for speech perception*, Journal of the Acoustical Society of America 48, 579-594.

Zwislocki, J.J. (1972) *A theory of central auditory masking and its partial validation*, Journal of the Acoustical Society of America 52, 644-659.