RELATIONSHIP BETWEEN SPEECH RECOGNITION AND SELF-REPORT MEASURES

Andrea Simpson [a)],
The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation,
The University of Melbourne, and The Bionic Ear Institute
384- 388 Albert Street East Melbourne. 3002 Australia

Hugh J. McDermott
Department of Otolaryngology, The University of Melbourne,
and The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation
384- 388 Albert Street, East Melbourne. 3002 Australia

Adam A. Hersbach
The Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation,
and The Bionic Ear Institute
384- 388 Albert Street, East Melbourne. 3002 Australia

[a)] Electronic Mail: asimpson@bionicear.org

ABSTRACT The performance of a prototype digital high-power hearing instrument was evaluated using tests of speech understanding in quiet and a questionnaire. The subjects were 26 adults with moderate-to-profound sensorineural hearing loss, most of whom were experienced hearing aid users. There was no significant difference between the group mean scores on monosyllabic word tests conducted in quiet for the prototype device versus the subjects' own hearing aids. However, responses to the questionnaire showed that 85% of the subjects preferred the prototype device to their own hearing aid(s). Although there was a positive correlation between the questionnaire results and the speech recognition score differences, overall there was a stronger tendency for subjects to prefer the prototype device than could be explained by their speech test results alone.

I. INTRODUCTION

Within the last decade, a range of fully digital hearing aids has become available on the commercial marketplace for clinical use. Digital technology has the advantage over its analog counterpart by being able to perform complex operations while consuming little power [4]. These instruments are capable of implementing a range of signal processing algorithms designed to improve speech intelligibility, listening comfort, and sound quality for people with a hearing impairment.

There are a wide variety of digital hearing aid products for the clinician to choose from. There is some evidence to suggest that despite implementing different processing techniques, digital hearing aids provide similar performance outcomes. For example, a study carried out by Harnack Knebel and Bentler [6] compared real and perceived benefit for two commercial digital hearing-aids. In that study, no significant differences were found between the hearing aids with objective testing of speech recognition.

In addition, clinicians should be aware of a general tendency for new devices to be preferred over existing technology for reasons other than objectively measurable performance improvements. Bentler et al. [1] compared users' preferences for identical hearing-aids after they had been labelled 'analog' or 'digital' at random, and described accordingly to the subjects. Strong preferences were observed for the devices labelled 'digital,' even when they were, in fact, analog hearing-aids.

The aim of the experiments reported below was to examine the relationship between objective measures of speech intelligibility in quiet listening conditions, and subjective measures obtained by means of a questionnaire when evaluating a prototype digital hearing-aid that was designed specifically for use by people with a moderately severe-to-profound hearing loss.

II. METHODS

A. The Prototype BTE Device

The test instrument evaluated in the trial was a prototype behind-the-ear (BTE) digital power instrument claimed to be suitable for people with hearing threshold levels that exceed 50 dB HL at all frequencies. It was omnidirectional and specified to have a maximum output and a maximum gain of approximately 140 dB SPL and 80 dB, respectively (measured in an ear simulator). The gain could be adjusted separately in five partially-overlapping frequency bands and covered a frequency range of 100-4800 Hz. Three main signal processing schemes could be selected during programming to suit the hearing characteristics of the user. These included a programmable amplitude compression scheme and two alternative schemes that provided essentially linear amplification but had slightly different output limiting techniques. The prototype aid did not have any unique signal processing features when compared to other digital aids. In appearance the test hearing aid resembled the patients' own aids. Subjects were provided with some information about the prototype aid as well as being told it was a test instrument.

It was possible to program several different sets of signal processing parameters into the test instrument when it was fitted to each user. These programs could be selected manually by the user to suit the ambient listening conditions. In the experiments described below, the impact on the users' perceptual performance of using only one of these programs was evaluated. Program 1 was selected as it was intended to provide appropriate amplification for most listening situations based on measurements of the hearing characteristics of the aid user.

B. Subjects

Twenty-six adults, comprising 11 women and 15 men, volunteered to participate in the trial. Relevant information about them is provided in Table 1. Their hearing threshold levels, measured conventionally under headphones, are listed in Table 2. The majority of subjects had moderate to severe hearing losses suitable for aiding by the test instrument. However, some subjects who had relatively good low-frequency hearing thresholds were included in the study because their losses were severe or profound from 1 kHz upwards.

For all subjects, hearing losses were assumed to have primarily a sensorineural origin, based on the results of hearing thresholds measured by bone conduction. In one subject (S9), an earlier assessment had indicated a retrocochlear lesion. Seven of the subjects had one unaidable ear where hearing

Table 1. Relevant information about the subjects who participated in the study, and their hearing-aids

Subject	Age	Sex	Probable etiology of hearing loss	Type of own hearing-aids	Features of own hearing-aids	Ears fitted with own hearing-aids	Processing strategy of test instrument	Ears fitted with test instruments
S1	26	F	Congenital	Widex L8	Digitally programmable	Binaural	Linear	Binaural
S2	72	F	Unknown	Bernafon RB15	Digitally programmable	Left	Linear	Left
S3	48	F	Congenital	Phonak PPCLC	Analog	Binaural	Linear	Binaural
S5	27	M	Congenital	Phonak Novaforte E4	Digitally programmable	Binaural	Linear	Binaural
S6	82	M	Industrial noise exposure	Widex Diva	Fully automatic digital aid with adaptive beamformer and compression	Binaural	Compression	Binaural
S7	42	M	Otosclerosis	Phonak PPSC	Analog	Left	Linear	Left
S8	79	M	Industrial noise exposure	Bernafon RB15	Digitally programmable	Binaural	Linear	Binaural
S9	68	F	Bilateral acoustic neuromas	Oticon 380P	Analog	Left	Linear	Left
S10	47	F	Unknown	Canal Aid Dynamic Equalizer II	Fully automatic digital aid	Binaural	Linear	Binaural
S11	54	F	Premature presbyacusis	Phonak PICS	Digitally programmable	Binaural	Linear	Binaural
S12	59	F	Otosclerosis	Starkey A-13	Analog	Right	Linear	Right
S13	69	M	Industrial noise exposure	Phonak PICS	Digitally programmable	Binaural	Linear	Binaural
S14	71	M	Industrial noise exposure, Cholesteatoma	Bernafon RB15	Digitally programmable	Left	Compression	Left
S15	69	M	Unknown	Widex Senso CX+	Fully automatic digital aid with adaptive directional microphone and compression	Binaural	Compression	Binaural
S16	54	F	Ototoxic drugs, family History	Resound Canta 7	Fully automatic digital aid with adaptive directional microphone and compression	Binaural	Linear	Binaural
S17	63	M	Industrial noise exposure	Widex Senso CX+	Fully automatic digital aid with adaptive directional microphone and compression	Left	Compression	Binaural
S18	74	M	Industrial noise exposure	Bernafon AA310	Digitally programmable	Binaural	Compression	Binaural
S19	61	F	Viral infection	Phonak Claro 21 daz	Fully automatic digital aid with adaptive directional microphone and compression	Right	Linear	Right
S20	60	F	Unknown	Oticon E39P	Analog	Binaural	Linear	Binaural
S21	76	M	Industrial noise exposure	Bernafon SB13	Digitally programmable, fixed directional microphone	Right	Compression	Right
S22	78	F	Miniere's disease	Bernafon SB13	Digitally programmable, fixed directional microphone	Binaural	Linear	Binaural
S23	80	M	Industrial noise exposure	Starkey Sequel	Analog	Binaural	Compression	Binaural
S24	69	M	Industrial noise exposure	Bernafon AA310	Digitally programmable	Binaural	Compression	Binaural
S25	74	M	Unknown	Bernafon LS16D	Digital hearing-aid with adaptive directional microphone and compression	Binaural	Linear	Binaural
S26	64	M	Industrial noise exposure	Phonak PPCLC	Analog	Binaural	Linear	Binaural
S27	46	M	Unknown	Phonak PPSC	Analog	Binaural	Linear	Binaural

thresholds at all frequencies where measured at 90 dB HL or greater, or were wearing only one hearing aid at the time of assessment. In these cases the fitting and evaluation of the hearing aids was carried out on only the single aided ear. One subject (S17) had aidable thresholds in both ears, but had been wearing a hearing aid in the left ear only. This subject was fitted binaurally for this trial. All subjects were experienced hearing aid users. Subjects were not paid for their participation in the experiments, although expenses such as travel costs were reimbursed.

C. Speech test materials

Consonant-vowel Nucleus-Consonant (CNC) word lists were presented from audio recordings [7]. There were 50 words per list, spoken by a female with an average Australian accent. Each word was a monosyllable such as "church". Each word consisted of three phonemes, making a total of 150 phonemes per list. No lists of words (other than practice lists) were repeated for any subject during the trial. The order in which lists were presented to subjects across sessions was randomized. The average level of the words, when measured at the subject's listening position (about 1 m from the loudspeaker), was 55 – 60 dBA. These levels, which are similar to the levels of speech in normal conversation, were generally perceived to be comfortably loud when heard by the subjects through their hearing-aids.

D. Procedure

Aid fitting

The hearing aid usage and medical history of each subject was documented during the first test session. A pure-tone audiogram, including both air and bone conduction, was obtained, and the electro-acoustic characteristics of each subject's own hearing aid(s) were measured and recorded. Most of the subjects' own hearing aid(s) had been fitted using NAL-RP fitting guidelines [2]. The NAL-RP formula aims to maximize speech intelligibility for the listener in both quiet and noise using linear amplification. Table 1 includes relevant details of each subject's own aids. Gain and output measurements with signal levels of 60 and 90 dB SPL were carried out using a standard 2-cm^3 coupler (Madsen Aurical) with both hearing instruments.

The test instruments were fitted to each subject using appropriate fitting software, with which user-selectable normal and noise-reduction programs were created. The software programmed the test instruments to provide target gains at each frequency as well as other signal processing parameters. In general, linear amplification was selected when the average hearing loss at 0.5, 1, 2, and 3 kHz was equal to or greater than 70 dB HL, whereas amplitude compression was selected in cases where the average hearing thresholds were lower (better). Table 1 provides relevant details of the final programs selected for each subject. The subjects' pure-tone thresholds were entered into the fitting software to derive an initial fitting suggestion. These settings were altered at the first follow-up session based on subject feedback. No changes were made to the fitting if the subject was happy with the sound quality of the device. If required, the programming of the test instruments was adjusted to approximate the amplification characteristics of the subject's own hearing aids, based on 2-cm^3 coupler measurements. Such an adjustment was performed for 14 of the subjects (S1, S2, S3, S5, S6, S7, S8, S9, S11, S14, S16, S17, S20, and S22), and resulted in only small differences between the gain these subjects received with the test instruments and with their own aids.

Word recognition in quiet

For all evaluations of speech intelligibility, each subject was tested individually in a medium sized sound-attenuating booth. Initially, the volume controls on each subject's own hearing aids were set for comfortable listening of speech at a conversational level in quiet conditions. For most subjects, this was the default volume control setting. This setting was noted and fixed for all following test sessions involving those aids. A practice CNC word list was then presented to familiarize subjects with the testing procedure and materials. Subjects were instructed to repeat each word immediately after hearing it, and to guess if unsure. After the practice list, two lists were used to test subjects in each of two conditions: (1) using their own hearing aids, and (2) using the test instruments with Program 1 enabled. Subjects' responses were analyzed to determine the number of phonemes correctly recognized out of a total of 150 phonemes per list. Responses from the practice list were excluded from the data analysis.

A counterbalanced sequence of testing was applied in an attempt to minimize the confounding effects of acclimatization over time (Gatehouse, 1992). Initially, subjects were tested with one list using their own hearing-aids. They were then asked to take the test instruments home, and use them in place of their own hearing-aids as much as possible. Each subject wore the test instruments for a total period of 10 – 14 weeks. The CNC word tests were carried out during the final two sessions of this period with the test instrument on Program 1. At the end of the trial period, subjects reverted to wearing their own hearing-aids. After a further two weeks, a final test was carried out to obtain a score for a second CNC word list using the subjects' own aids.

Self-assessment

At the conclusion of the trial, each subject was asked to complete a questionnaire which was designed to elicit responses comparing the test instruments with their own hearing-aids. The questionnaire, which was adapted from the Shortened Hearing Aid Performance Inventory for the Elderly, or SHAPIE (Dillon, 1994), comprised 23 questions. Subjects completed the questionnaire in the laboratory. Responses were indicated by marking a horizontal line printed immediately after each question. Half of the line was marked "Own hearing-aid," and the other half "Experimental hearing-aid." The position of the label "Own hearing-aid" on either the right or left half of each line varied randomly. Each half of the line carried marks labelled with the words "slightly better," "better," and "much better," spaced regularly and symmetrically about the midpoint. Thus, the midpoint of the line corresponded to a response indicating that the subject judged the two types of hearing-aid to be indistinguishable. Subjects were able to respond "Not applicable" to any question. In the analysis of

Table 2. Hearing threshold levels (dB HL) for the subjects who participated in the study

Note: Asterisks indicate levels that were limited by the maximum possible output of the audiometer.

		Frequency (kHz)								
Subject	Ear	0.25	0.5	0.75	1	1.5	2	3	4	8
S1	L	55	70		80	85	85	90	110	110*
	R	55	65		75	85	95	105	105	110*
S2	L	35	50	70	85	85	85	95	120	110*
S3	L	90	95	115	120*	120*	120*	110	120*	110*
	R	80	95	105	115		120*	120*	120*	110*
S5	L	60	65	85	85		80	80	85	90
	R	40	65	65	85		85	70	70	95
S6	L	45	55		70		75	65	75	95
	R	35	45		65		65	75	80	85
S7	L	55	70	80	90		85	75	90	110*
S8	L	50	65	80	85	120*	120*	120*	120*	110*
	R	45	65	80	90	115	120	115	115	110*
S9	L	65	65	70	90	95	100	120	105	110*
S10	L	45	65		80	90	90	100	120	110*
	R	40	60		80	95	90	95	100	110*
S11	L	50	60	80	90	90	90	85	90	75
	R	50	55		70	75	80	105	95	110*
S12	R	70	65		55	70	85	95	95	110*
S13	L	50	35		60	80	90	90	100	90
	R	30	35		60	65	70	85	90	75
S14	L	40	40		45	35	50	105	100	110*
S15	L	50	75	75	85	95	85	90	110	110*
	R	40	50	60	75	75	70	65	75	75
S16	L	60	75	90	105	115	120	120*	120*	110*
	R	70	95	95	105	115	120	120*	120*	110*
S17	L	55	50		55	65	70	80	85	110*
	R	55	50		60	70	75	85	110	110*
S18	L	40	55		65	80	80	70	80	70
	R	30	45		55	65	90	85	80	70
S19	R	55	60		65	70	80	80	85	110*
S20	L	60	60	70	80	90	100	95	110	110*
S21	R	50	50		50	60	70	80	90	110*
S22	L	90	85		85	85	80	80	85	100
	R	35	45	60	70	100	110	95	95	110*
S23	L	40	50		55	60	70	85	105	110*
	R	35	45		55	55	65	80	100	110*
S24	L	20	35	45	60	60	65	80	75	100
	R	30	45	50	60	75	80	105	85	100
S25	L	55	60		70	75	90	120	120	110*
	R	55	60		65	75	70	80	95	110*
S26	L	55	70	65	65	75	75	100	120	110*
	R	55	70	70	70	70	75	90	110	110*
S27	L	10	50	60	70	115	115	120	115	110*
	R	10	35	60	70	115	115	115	115	110*

each subject's data, questions answered with such a response were omitted. Otherwise, each subject's response to each question was assigned an integer value ranging from –5 (for the response "Own hearing-aid much better") to +5 (for the response "Experimental hearing-aid much better").

IV. RESULTS

A. Word recognition in quiet

For the CNC word test in quiet, mean phoneme scores for each subject with their own hearing-aids and with the test instruments on Program 1 are shown in Fig. 1. Although there was some variability among subjects, the group mean scores (rightmost columns) showed almost no difference in phoneme scores between these two conditions. A paired t-test on these data confirmed that the scores were not significantly different

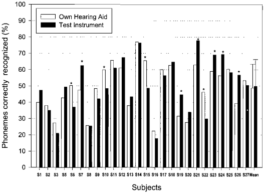

Figure 1. Mean phoneme scores recorded for the 26 hearing-impaired subjects when listening to monosyllabic words in quiet. Filled columns show scores obtained using the test instruments (with Program 1), and unfilled columns show scores obtained using the subjects' own hearing-aids. Scores averaged across subjects are shown in the pair of rightmost columns, with error bars indicating one standard deviation. Asterisk symbols indicate statistical significance ($p < 0.05$).

($t = -0.506$, $df = 25$, $p = 0.62$). Further analysis was carried out on subjects' individual scores using a Chi-squared test. As shown, 6 of the subjects (S7, S19, S21, S23, S24, S26) obtained significantly higher scores ($p < 0.05$), and 4 subjects (S6, S10, S15, S22) obtained significantly lower scores ($p < 0.05$) with the test instruments than with their own aids on this test. The remaining subjects' scores were not significantly different between the two test conditions.

B. Self-assessment

To analyse the results from the comparative questionnaire, the numbers assigned by each subject as responses were averaged across the 23 questions. The mean response values are shown for each subject in Fig. 2. Positive values, plotted on the right of the graph, indicate preference for the test instruments, whereas negative values, plotted on the left, indicate preference for the subject's own aids. Although preference ratings varied considerably, all but 4 of the subjects indicated that they preferred the test instrument to their own hearing-aids. The exceptions, subjects S6, S11, S15, and S16, indicated that they had only a relatively small preference for their own hearing-aids. It is noteworthy that three of these subjects (S6, S15, and S16) owned hearing-aids that employed relatively sophisticated signal processing schemes (see Table 1).

V. DISCUSSION

For the subjects who participated in the study, the test instrument provided perceptual performance approximately equal, on average, to the performance of the subjects' own hearing aids when listening to words presented at a moderate level in quiet conditions. This outcome was not unexpected, particularly because the test instruments were specifically adjusted for

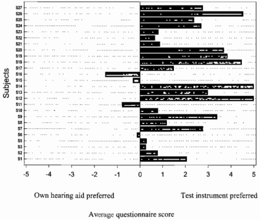

Figure 2. Mean scores from the questionnaire provided to the subjects. Each horizontal bar shows, for each subject, the average across all questions of numerical values assigned to that subject's responses. As shown on the horizontal axis, the possible values range from –5 (own hearing-aid preferred) to +5 (test instrument hearing-aid preferred).

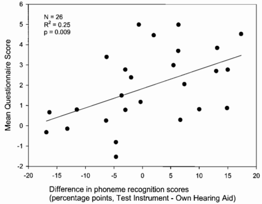

Figure 3. The relationship between the mean questionnaire score for each subject (vertical axis) and the difference in phoneme recognition scores obtained by the same subjects for the two types of hearing-aid evaluated in the study (horizontal axis). The score difference was calculated by subtracting the phoneme recognition score for the monosyllabic words test in quiet using the subject's own aids from the corresponding score obtained with the test instruments (on Program 1). The straight line shown fitted to the data indicates that about 25% of the variance in the questionnaire scores is accounted for by the variance in the phoneme score differences.

about half the subjects at the first follow-up session after initial fitting to provide a gain and frequency response similar to that of their own hearing aids. Interestingly, the large majority of subjects who showed no significant differences in scores between the two devices had this adjustment made to the test device. For many of the remaining subjects, it is probable that the initial programming of the test instruments also provided electro-acoustic parameters similar to those of their own aids. These speech perception results are consistent with the findings reported by Harnack Knebel and Bentler [6].

However, the results of the questionnaire administered in the present study showed that 22 of the subjects preferred using the test instrument rather than their own aids in many everyday situations. During the trial, each subject was aware of which aid they were using, and therefore it is possible that the positive results from the questionnaires reflect a general tendency for the new devices to be preferred over their existing hearing aids. Could this bias have affected the results (shown in Fig. 2) from the questionnaire used in the present study?

To investigate this issue, the questionnaire results were plotted as a function of the difference in phoneme recognition scores (in quiet) for each subject when using the test instruments compared with their own hearing-aids. These data, and a fitted straight line, are shown in Fig. 3. A statistical analysis revealed that the questionnaire scores were moderately correlated with the difference in phoneme scores ($r = 0.5$). The fitted line has a positive slope that was confirmed to be significantly different from zero ($p = 0.009$). The straight line shown fitted to the data indicates that about 25% of the variance in the questionnaire scores is accounted for by the variance in the phoneme score differences. The remaining 75% of the variance may be accounted for by a variety of factors, including test-retest variance.

As shown in Fig. 3, subjects who obtained a larger improvement in speech understanding when using the test instruments compared with their own aids were more likely to have provided positive responses to the questionnaire. Consequently, it seems likely that the preferences for the test instruments were related to the subjects' personal judgments of its perceptual performance (relative to that of their own hearing-aids), rather than reflecting only a bias associated with their involvement in the trial. However, the observation that the phoneme score differences are approximately evenly distributed around zero, whereas the mean questionnaire scores are mostly positive, suggests that, on average, subjects had a stronger tendency to prefer the test instruments overall than can be explained by differences in their objectively-measured ability to understand speech. In general, this outcome is consistent with that reported by Bentler et al. [1] where subjects showed a preference for new technology.

Speech perception in quiet is only one aspect in which a hearing aid can provide benefit for the listener. There are many additional listening environments which would affect how a listener would judge sound quality. These other environments include listening in noise, music, and environmental sounds. It is possible that the test instrument may have provided perceptual benefits for the subjects in ways that were not measured in the current study. This may also account for the majority of subjects' preference for the test instrument rather than their own hearing aids.

VI. CONCLUSIONS

The results of these evaluations of a prototype digital high-powered hearing instrument can be summarized as follows.

1. Recognition of words presented in quiet did not differ

significantly, on average, between the test instruments and the subjects' own hearing-aids.

2. Based on responses to the questionnaire, 22 of the 26 subjects preferred the test instruments to their own hearing-aids overall.
3. Across subjects, a moderate positive correlation was found between the questionnaire responses and the difference in objectively measured speech intelligibility for the test instruments in comparison with the subjects' own hearing-aids.
4. It is important to examine speech intelligibility as well as subjective measures when assessing the performance of hearing instruments.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support of the Commonwealth of Australia through the Cooperative Research Centres program. The second author received financial support from the Garnett Passe and Rodney Williams Memorial Foundation. Thanks go to two anonymous reviewers who commented on a previous version of this manuscript. The contributions of many colleagues to this work are greatly appreciated. In particular, we thank Dr. Robert Cowan, Catherine Sucher, Rodney Millard, Justin Zakis, Katherine Henshall and Dr WaiKong Lai. We are also especially grateful to all of the hearing aid users who participated as subjects in the experiments.

REFERENCES

1. Bentler, R. A., Niebuhr, D.P., Johnson, T.A., Flamme, G. A. "Impact of digital labeling on outcome measures." Ear. Hear. 24, 215-24 (2003)
2. Byrne, D., Dillon, H. "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid." Ear. Hear. 7, 257-65 (1986)
3. Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R.M., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, H.A., El Kholy, W., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G., Westerman, S., Ludvigsen, C. "An international comparison of long-term average speech spectra." J Acoust Soc Am. 96, 2108-2120. (1994)
4. Dillon, H. "Shortened Hearing Aid Performance Inventory for the Elderly (SHAPIE): a statistical approach." Aus. J. Audiol. 16, 37-48. (1994)
5. Dillon, H. Hearing Aids, Boomerang Press: Sydney. (2001)
6. Gatehouse, S. "The time course and magnitude of perceptual acclimatization to frequency responses: evidence from monaural fitting of hearing aids." J Acoust Soc Am. 92, 1258-68. (1992)
7. Harnack Knebel, S.B., Bentler, R.A. "Comparison of two digital hearing aids." Ear. Hear. 19, 280-9. (1998).
8. Peterson, G., Lehiste, I. "Revised CNC lists for auditory tests." Journal of Speech and Hearing Disorders. 27,62-70. (1962).