The TIDE project OSCARS

A. Faulkner[1], N. van Son[2], CM. Bcijk[2]

[1] Dept. Phonetics and Linguistics, University College London, London,UK;
[2] Instituut voor Doven, Sint-Michielsgestel, The Netherlands

Introduction

The aim of the OSCAR project is to develop and evaluate a novel DSP-based acoustic and vibrotactile aid that is designed to be adaptable to the needs and sensory capacities of a wide range of the profoundly and totally hearing disabled population. At present this heterogeneous population is served by a diversity of specialised aids. The OSCAR aid is a body-worn speech processing unit that anticipates a versatile product that can meet the needs of many, and so allow economies of scale in manufacture.

The acoustic processing approach has been defined jointly by UCL and Oticon A/S. The tactile stimulation approach has been determined by KTH (Spens et alM 1996). The speech processing unit, known as SiVo-3, incorporates both conventional whole-speech presentation and noise-resistant speech selective analysis that extracts information chosen for its effectiveness in supplementing lipreading. Speech-analytic or whole-speech processing are selected by a switch so that the user has instant access to the processing that is more appropriate to their immediate needs.

The main activities of the project are: to determine the added value of speech analytic processing in unimodal and bimodal stimulation configurations, to explore the potential of bimodal stimulation configurations, and to characterise the users likely to benefit from these processing and stimulation methods. The evaluations include both laboratory-based assessments of speech perception and production, and subjective user assessments of the effectiveness of the demonstrators in everyday use.

Speech analytic processing

The speech-analytic mode of the OSCAR aid is aimed at enhancing the benefit of a hearing aid when lipreading in everyday conditions of background noise. It employs a noise-resistant artificial neural network (ANN) algorithm (Wei et al., 1993) that can extract the temporal voicing pattern and voice fundamental frequency from noisy and reverberant speech. The analytic mode also detects the presence of high-frequency speech energy so that voiceless speech sounds can be encoded. For both voiced and voiceless speech, the amplitude envelope is also extracted.

In auditory presentation, the speech information extracted and matched by the OSCAR SiVo-3 processing unit has been shown in laboratory studies to have utility for profoundly impaired listeners (Faulkner et al., 1993). The voicing pattern and voice fundamental frequency are encoded as a frequency-controlled sinusoid, while voiceless speech is encoded as a random noise with a spectrum and dynamic range matched to the user's comfortable hearing area. Both sinusoid and noise signals are amplitude modulated according to the extracted amplitude envelope. The frequency and intensity range of these acoustic signals is controlled to match the user's residual hearing.

For the tactile sense, voiced and voiceless excitation and the associated amplitude envelopes are represented by two vibrators. The voiceless pattern is presented by a newly developed tangential vibrator that stimulates the non-Pacinian skin receptors (Huss & Spens, 1996). Fundamental frequency is not coded in the tactile signal.

Conventional "whole-speech" processing

For acoustic use, the OSCAR aid provides a reference mode using conventional amplification, but with more flexible control of the frequency response that is typical of high-gain hearing aids. The specification of the conventional hearing aid reference has been defined by Oticon. It represents a state-of-the-art hearing aid regarded as appropriate to our user group:

a. acoustic output to a maximum of 140 dB SPL from 100 Hz to 2 kHz;
b. insertion gain determined by the POGO rule (McCandless & Lyregaard, 1983) extended to specify the 125 Hz gain;
c. maximum output limited to average uncomfortable listening level over 125 Hz to 2 kHz;
d. switchable low-frequency cut (18 dB/oct, 1 kHz).

Speech-analysis methods

The speech analytic processing of the OSCAR aid is a development of that used in the SiVo-2 hearing aid that was evaluated in the preceding STRIDE project (Faulkner, 1995). The ANN algorithm has been trained using both quiet and noisy speech to produce an output that corresponds as closely as possible to the cycle-by-cycle period of larynx vibration as indicated by a laryngograph signal. Compared to appropriate reference methods, such algorithms are superior in the voiced/voiceless classification of speech in moderate levels of noise, and show accuracy in fundamental frequency estimation that is acceptable for hearing aid applications (Bosnian & Smoorenburg, 1997a; Faulkner & Wei, 1997).

Field trials

The main objectives of the field trials are threefold:

1. For an acoustic aid: to identify the characteristics of those hearing impaired users for whom switchable speech analytic processing mode in an acoustic aid provides worthwhile benefits, and the listening conditions in which these benefits occur.
2. For a tactile aid: to estimate the additional benefit of noise-resistant speech analytic processing and of the OSCAR two-vibrator configuration over standard tactile aids.
3. For a bimodal aid: to investigate the additional benefit of combined auditory and tactile stimulation for appropriate users compared to unimodal stimulation.

This paper presents selected results from the unimodal acoustic aid trials. A full report of the trial results, including data from the unimodal tactile aid. The TIDE project USCAJi >v and the bimodal system, and speech production data and subjective user assessments relating to the acoustic, tactile and bimodal aids, will be reported elsewhere (van Son et al., 1997).

Acoustic aid field trials

Some individual subjects have been found previously to benefit in speech perception (Faulkner et al.,1992) and in voice control (Ball et al., 1990) from the simple acoustic signal provided by the speech analytic SiVo mode. However, simplification of speech in itself does not appear to be of benefit for the majority of the profoundly hearing impaired population (Bosman & Smoorenburg, 1997b). The main benefit expected here is in the perception of speech in noise, arising from the noise-resistant analysis of the SiVo-3 fundamental frequency extraction algorithm.

Trial subjects

The results overviewed here are based on post-training speech perceptual data from a group of 12 profoundly hearing impaired users. Data were collected for three languages, English (at UCL: n=7), Dutch (at Instituut voor Doven: n=3) and French (at Fondation Rothschild, Paris: n=2). Subjects for the acoustic aid trials were selected on criteria designed to include those most likely to benefit from speech-analytic processing:

1. Post-lingual profound hearing loss of primarily sensori-neural origin.
2. Audiometric criteria for the better ear: HL at 4 kHz>120 dB ISO, HL at 2 kHz>100 dB ISO; measurable hearing at 250 and 500 Hz.
3. Age between 18 and 75 years.
4. Limited benefit from conventional hearing aids as measured in the lipreading of consonants.
5. Access to a family member or friend able to act as training partner. Clinical training involved the balanced use of both the conventional and speech-analytic modes of the SiVo-3 aid. Training included the perception of prosody and rhythm, consonantal contrasts, vowel contrasts (in the conventional mode only) and environmental sounds.

Table 1. Unimodal field trial phases

Speech perceptual assessments

/vCv/ identification test
Audio-visual materials were prepared using 14 consonants that are common to Dutch, English, French and Swedish (m, b, p, v, f, n, 1, r, y, d, t, s, sh, k) in an intervocalic /a/ vowel context. Materials were produced by female native speakers of Dutch, English, French and Swedish. All were digitally acquired using M-JPEG compression (using Videologic Mediaspace/ DVA 4000 hardware) and edited to provide digital audio-visual data files for a minimum of ten tokens of each consonant. The digital files were additionally processed to provide materials with added background noise having speech-spectrum frequency shaping. The noise rms. level was set 5 dB or 10 dB below the rms. level of the speech. Each individual vCv test session used 4 randomly selected tokens of each consonant. Two vCv test lists were presented in each pre-training and post-training test condition.

Sentence recognition test
For the Dutch subjects, sentence materials consisted of everyday Dutch sentences of eight to nine syllables (Plomp & Mimpen, 1979), read by a female speaker. The audio visual recordings were provided by the Laboratory of Experimental Audiology of the University Hospital Utrecht. For the English subjects, standard video recordings made by UCL of the BKB sentences were used (Foster et al., 1993). The BKB sentence test comprises 21 lists of 16 sentences, with a total of 50 scored key words for each list. In each test condition, two sentence lists were presented to each subject. The French subjects were tested using sentences based on the English BKB sentences.

Validity of conventional SiVo-3 mode as a reference condition
A comparison of scores in the post-training speech tests between the user's previous aid and the conventional mode of the SiVo-3 unit confirmed that this mode is acceptable as a reference aid. In both vCv and sentence tests, scores in quiet were higher on average than those with users' own aids, but in neither case was this difference significant.

Comparison of analytic and conventional modes as a function of signal-to-noise ratio
The analysis here focuses on speech perception scores in quiet and noise using the conventional and analytic processing modes of the SiVo-3. Nine subjects of the OSCAR user group completed tests in both quiet and noise. Reference is also made

to tests carried out previously in a comparable group of 14 profoundly hearing impaired subjects, using the SiVo-2 aid and the subject's own conventional aid.
The SiVo-2 aid, like the analytic mode of SiVo-3, provided voice fundamental frequency and speech amplitude information, but SiVo-2 did not provide voiceless frication information. It differed also from the SiVo-3 in using a different configuration of the MLP fundamental period extraction algorithm. While the differences in detail between aids used by the OSCAR subjects and the second subject group cannot be disregarded, it is nevertheless reasonable to group the data together to clarify the effects of the analytic The TIDE project OSCAR 41 processing and to gain a larger subject group to identify audiometric correlates of performance.

Fig. 1. Group vCv identification: mean % correct and 95% confidence limits. Data are shown for both aid conditions at three signal-to-noise ratios. Unaided visual performance is shown on the right. C: SiVo-3 conventional mode, A: SiVo-3 analytic mode.

Audio-visual vCv identification
The group means displayed in Fig. 1 show considerable inter-subject variability. In quiet, scores are somewhat higher using the conventional mode, but as background noise is introduced, this pattern is reversed. An ANOVA showed a main effect for the noise condition ($F=6.1$, $p<.01$), but no effect of processing mode. There was, however, a significant interaction of S/N ratio and aid mode: $F=8.4$ ($p<.01$), showing that performance using the analytic mode deteriorates significantly less in background noise than is the case with the conventional mode. At the 5 dB S/N ratio the range of scores with the conventional mode is essentially the same as in the unaided condition, while the bulk of the scores in the analytic condition are above the mean unaided score.

An examination of the within-subject differences between the two aid modes (see Fig. 2) makes it clear that at the 5 dB S/N, virtually all subjects show higher scores with the analytic mode. In quiet, most subjects are performing better with the conventional mode.

Sentence identification
For the sentence test, only one signal-to-noise ratio was tested for each subject. This was 5 dB for the Dutch subjects and 10 dB for English subjects. The Dutch and English sentence in noise data did not differ significantly, and so the results were pooled into a general NOISE condition.

Fig. 2. Mean and 95% confidence limits of within-subject differences between analytic and conventional mode scores in vCv identification.

Globally, the trends mirror the vCv results. Group data are shown in Fig. 3. Performance in quiet is better with the conventional mode than with the analytic mode, however the analytic mode scores are less reduced in the presence of background noise. An ANOVA showed a main effect of the noise condition ($F=5.2$, $p<.05$) but not of the aid mode. Here the interaction effect of noise aid mode marginally missed statistical significance ($F=3.9$, p=.O57). Within subject differences between the two modes are shown in Fig. 4. As for the vCv data, most subjects showed higher scores in quiet using the conventional mode, while in noise, the mean difference favoured the analytic mode.

Fig. 3. Group sentence identification scores: mean % correct and 95% confidence limits. Data are shown for both aid conditions in quiet and noise. Unaided visual performance is shown on the right

Consistency of aid effects over subjects

Individual subject scores show internal consistency between the vCv and sentence data. Fig. 5 shows differences between the analytic and conventional mode scores for individual subjects. For sentences in noise, six of the nine OSCAR users showed higher scores using the analytic mode compared to conventional amplification, while three showed scores that were lower. The same six subjects, together with two others, showed higher scores for vCvs at the 5 dB S/N ratio using the analytic mode. The pattern here is also consistent with that seen in data from fourteen English subjects who have been tested with the SiVo-2 aid and conventional aids. Difference scores for this subject group are shown in Fig. 6.

Fig. 4. Mean and 95% confidence limits of within-subject differences between analytic and conventional mode scores in sentence identification.

Relations with audiometric measures

Speech scores in the different aid conditions show patterns of correlation with hearing loss and dynamic range measures that are complicated by strong negative correlations between hearing loss at higher frequencies and unaided lipreading scores. To seek characteristics that may define the population expected to benefit from the availability of a noise-resisting speech analytic mode in a hearing aid, we have examined relations between audiometric measures and the within-subject differences between analytic and conventional mode scores for speech in noise. Across the OSCAR group and the SiVo-II test groups these differences, averaged over the sentence and vCv tests, show a significant (p=0.05) positive correlation of 0.45 with hearing loss at 1 kHz. This relationship is shown in Fig. 7. On the basis of the lower 95% confidence limit of the regression fit, subjects with HL at 1000 Hz of greater than 115 dB are likely to show better performance in noise with the SiVo-3 analytic processing. From the intercept of the regression line, subjects with HL at 1000 Hz of less than 105 dB are unlikely to show better performance for speech in noise with analytic processing. Subjects with intermediate HLs at 1000 Hz show a wide range of difference scores, including the largest of the differences favouring analytic processing in noise.

Fig. 5. Individual subject differences between analytic and conventional modes: OSCAR subject group. Subjects are ordered from left to right according to the average difference score over the speech tests shown.

Figure 6. Individual subject differences between analytic and conventional modes: SiVo-II subject group. Subjects are ordered from left to right according to the average difference score over the speech tests shown.

The data showed no significant relationships between lower frequency losses and the enhancement of lipreading in noise using the analytic mode, but measurable low frequency hearing is clearly a precondition for benefit and was indeed a selection criterion for the trial group. Since the relations found between 1000 Hz HL performance account for only 20% of the variance, it seems likely that speech-based testing would be required in addition to audiometry in order to identify those likely to benefit if this analytic processing were clinically available.

Fig. 7. Average of differences in speech in noise scores vs. HL at 1 kHz. Points numbered between 1 and 10 are for the OSCAR trial group,*and points numbered from 21 upwards for the SiVo-II subject group. The solid lines show the linear regression fit and 95% confidence limits.

Conclusions from acoustic aid trials

The use of noise-resistant speech analytic speech processing does appear to represent a significant advantage for speech in noise for most of the OSCAR test subjects, especially those with more extreme hearing losses at 1000 Hz. An analysis of subjective user assessments of the OSCAR prototype and a survey of the profoundly impaired population is being carried out to estimate the feasibility of a commercially produced aid that includes analytic processing. This will be reported in the OSCAR project final report (OSCAR, 1997).

Other components of the OSCAR project

The project has also included other activities related to the general value for the profoundly hearing impaired population of the clinical use of aids and of assessment methods based on an understanding of speech perception. These include:

1. The development and use of a PC-based system for audio-visual speech perceptual assessment known as ASTEC (Speech Assessment Test Editor and Controller: Pavlovic et al., 1995). The MS-Windows based ASTEC software allows the implementation of a wide range of speech perceptual tests and includes a database for the storage and analysis of clinical data.
2. A survey by Oticon A/S of the prevalence of profound and total hearing loss and the rehabilitation approaches in use in Europe for this population.
3. Investigations of multi-band compression in acoustic aids for the profoundly hearing impaired (Drullman and Smoorenburg, 1996).
4. Studies at UCL of the potential of profoundly impaired hearing for the perception of speech-related spectral information (e.g. Deeks & Faulkner, 1994; Vickers & Faulkner, 1996).

References

Ball V, Faulkner A, Fourcin AJ. The effects of two different speech coding strategies on voice fundamental frequency control in deafened adults. Br. J. Audiol. 1990; 24: 393-409.

Bosman AJ. Smoorenburg GF. Evaluation of three pitch tracking algorithms at several signal-to-noise ratios. Acta Acustica, in press, 1997a.

Bosman AJ, Smoorenburg GF. Speech reading supplemented with auditorily presented speech elements in the profoundly hearing impaired. Audiology 1997b; 36: 29-45.

Deeks J, Faulkner A. Residual spectral and temporal processing in signalling timbre contrasts in synthetic vowels. Speech,' Hearing and Language, Work in progress, Dept. Phonetics and Linguistics, University College London, 1994; 8: 141-162.

Drullman R, Smoorenburg GF. Multichannel amplitude compression for the profoundly hearing impaired. In: van Son N, Coninx F. eds. Proc. International Sensory Aid Conference 1996, Sint-Michielsgestel, The Netherlands, 1996; 42-45.

Faulkner A, Walliker JR, Howard IH, Ball V, Fourcin AJ. New developments in speech pattern element hearing aids for the profoundly deaf. Scand. Audiol. Suppl. 1993; 38: 124-135.

Faulkner A. Final Report of TIDE Project 133/206: Ref. STRIDE/ 1995/2, Department of Phonetics and Linguistics, University College London. 1995.

Faulkner A, Wei J. Final report of Workpackage A: Speech analytic methods, in preparation. Ref. OSCAR-UCL-1997-WPA, 1997.

Foster JR, Summerfield AQ. Marshall DH, Palmer L, Ball V, Rosen S. Lipreading the BKB sentence lists: corrections for list and practice effects. Br. J. Audiol. 1993; 27: 233-246. !

Huss C, Spens K-E. Radial and tangential tactile stimulation; some masking aspects. In: van Son N, Coninx F, eds. Proc. International Sensory Aid Conference 1996, Sint-Michielsgestel, The Netherlands, 1996; 6-8.

McCandless G, Lyregaard PE. Prescription of gain/output (POGO) for hearing aids. Hear. Instruments 1983; 34(1): 16-21. OSCAR. Final report of TP1217 OSCAR, in preparation. Ref. OSCAR-UCL-1997.

Pavlovic C, Brousseau M, Howells D, Miller D, Hazan V, Faulkner A, Fourcin A. Analytic assessment and training in speech and hearing using a poly-lingual workstation. EURAUD. In: Placencia Porrero I, Puig de la Bellacasa R, eds. The European Context for Assistive Technology, IOS Press, Amsterdam, 1995; 332-335.

Plomp R. Mimpen AM. Improving the reliability of testing the speech reception threshold for sentences. Audiology 1979; 18: 43-52.

Son N van, Beijk C, Faulkner A. Final report of Workpackage E: Field trials, in preparation. Ref. OSCAR-UCL-1997-WPE (1997).

Spens K-E, Huss C. Dahlquist M, Agelfors E. Characteristics of, and preliminary results from a hand held two-channel vibro-tactile speech communication aid for the deaf. In: van Son N, Coninx F, eds. Proc. International Sensory Aid Conference 1996, Sint-Michielsgestel, The Netherlands, 1996; 30-33.

Vickers DA, Faulkner A. Noise spectrum discrimination by severe-to-profoundly hearing-impaired listeners. In: Kollmeier B, ed. Psychoacoustics, Speech and Hearing Aids, World Scientific, Singapore, 1996; 25-28.

Wei J, Howells D, Fourcin AJ, Faulkner A. Larynx period and frication detection methods in speech pattern hearing aids. Speech, Hearing and Language. Work in progress, Dept. Phonetics and Linguistics, University College London, 1993; 7: x 269-276.