

A three-microphone system for real-time directional analysis: Toward a device for environmental monitoring in deaf-blind

Erik Borg, MD, PhD; Lennart Neovius, MSc, Techn. Lic.; Magnus Kjellander, MSc

Ahlsén Research Institute, Örebro Medical Centre Hospital, S-701 85 Örebro, Sweden; Department of Physiology and Pharmacology, Karolinska Institute, S-171 77 Stockholm, Sweden; Hitech Development AB, Åldermansvägen 19-21, 171 48 Solna, Sweden

Abstract—Sound localization is a problem for the hearing impaired, deaf, and deaf-blind. A laboratory prototype of a three-microphone digital system for real-time directional analysis of sound sources that is mounted on eyeglasses has been developed. A cross-correlation algorithm is used for determinations of time and phase differences between the three microphone signals. The equipment was evaluated in a sound-treated test room with noise and speech stimuli from 12 different directions. It has an adequate precision (within approximately ± 10 degrees) for speech and noise bursts in the frequency range 0.5–4 kHz, and for testing with words in a sound field the error percentage is very close to 0 at 0-degree azimuth. The robustness of the algorithm in terms of signal-to-noise ratio is documented and reaches about +8 dB in different directions for speech and up to 0 dB for noise bursts. It is concluded that the construction of a robust system for directional analysis with sufficient efficiency for further experiments has been achieved.

Key words: *directional analysis, sensory disabled, sound localization.*

This material is based on work supported by the Swedish Council for Social Research (Grant #94-0053:3C) and grants from Örebro Medical Centre Hospital and the Silent School Association.

Address all correspondence and requests for reprints to: Erik Borg, MD, PhD, Ahlsén Research Institute, Örebro Medical Centre Hospital, S-701 85, Örebro, Sweden; email: anita.dandenell@orebroll.se.

INTRODUCTION

Sound localization in man and animals is based on the detection of binaural differences in intensity and time of arriving sound waves as well as phase differences and spectral cues (for a recent review see reference 1). Information on the direction to a sound source also contributes to selective attention, listening in noise, and speech reading (2,3). Localization of events in the surroundings even contributes to the feeling of control and security (3,4). In subjects with hearing impairment, unilateral deafness, or bilateral deafness, sound localization ability is deteriorated or absent.

Determination of direction of a sound source is a classic problem in military and marine constructions for target localization. In addition to the task of making a rapid or real-time analysis, there are usually high requirements on robustness against noise and in reverberant acoustic conditions. The recent development of computer programs for virtual reality has given analyses and simulation of acoustic environments a new application that is rapidly expanding (5). The method for directional analysis proposed in this study has been selected for low complexity and straightforward parameter settings.

For the application of real-time directional analysis in human environmental monitoring and communication for individuals with sensory impairment, there is an

additional requirement: small size and light weight. There are at least three groups that would benefit from such devices. First, persons with unilateral hearing loss or deteriorated sound orientation ability would greatly benefit from being able to quickly localize the speaker in a group of communicating persons in order to supplement the acoustic cues with speech reading (3,4). Second, totally deaf people would benefit from an improved ability to monitor the acoustic environment, for instance in traffic, and to rapidly identify the presence of a signing person in the vicinity (provided the person produces some attention-getting acoustic signals).

The third group is made up of deaf-blind individuals who have no acoustic orientation or identification ability and very poor or no vision. They may rely on a small visual field for sign communication, for instance the Usher I subjects, or they may be totally blind and need direct manual contact for communication, e.g. Tadoma (6). It is often reported that these persons startle and may be frightened by unexpected approaches. There is, at present, little knowledge of how deaf-blind people get information about and monitor the environment and how to improve on their situation in this respect (for a review see reference 7).

When neither deaf-blind individuals' audition nor vision is sufficient to give environmental information, the cutaneous senses are the most promising substitute. Several attempts have been made to design instruments for sound localization and to test the orientation ability when using vibrators on the skin. No portable instrument has been tested so far, but the laboratory results are promising. A recent review has been presented (8). The ear-worn two-channel instrument designed by Weisenberger and colleagues (9) is the closest attempt, presented so far, to design a portable instrument for sound localization. However, more extensive documentation and application has not yet followed the initial presentation.

In an ongoing project, we have approached these problems from several aspects. In one study, we analyzed attitudes of the deaf-blind towards environmental interaction and the strategies they use for environmental control (10). In a second study, we designed a laboratory instrument that performs real-time directional analysis using the input from three microphones with real-time analysis based on the principle described by Kaneda (11). With this system, directional data are presented to the user by means of vibrators. In a third study, deaf and deaf-blind subjects evaluated the entire system.

The purpose of this present study is to describe the analysis program and the test results with the system

placed on mannequin and human heads in our sound environmental chamber (12). Subject performance is examined in a companion report (19).

METHODS

Sound Localization Instrument

Figure 1 shows a schematic drawing of the three microphones and the four vibrotactile transducers mounted on a pair of eyeglasses. The microphones (mini electret, type AGK 417/C with a flat frequency response from 20 Hz to 20 kHz) were mounted in front of the ears and on the front piece between the glasses. The four vibrators (KASEI A4B-03-W) were mounted on soft springs to keep skin contact pressure reasonably constant. Two vibrators were placed on the mastoid bone behind the ears, two in contact with the lateral part of the orbita. The present article only describes the results of the analysis program and the participating humans were not asked to indicate perceived directions.

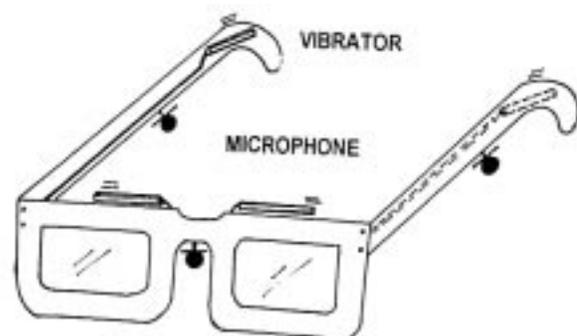


Figure 1.

Mounting of three microphones and four vibrators on eyeglasses used for analyses and coding direction to sound sources.

Directional Analysis

A cross-correlation method is used for directional analysis. **Figure 2** shows a block diagram for the analysis algorithm. The algorithm is based on Kaneda (11). With three microphones and limited computing power, it gives an adequate, easily controlled, robust 360° directional analysis. With this result, comparison with other methods was found less important for the study.

The algorithm uses a time delay, peak hold estimation technique for reducing the sensitivity for echoes. In the proposed system, combinations of three correlations

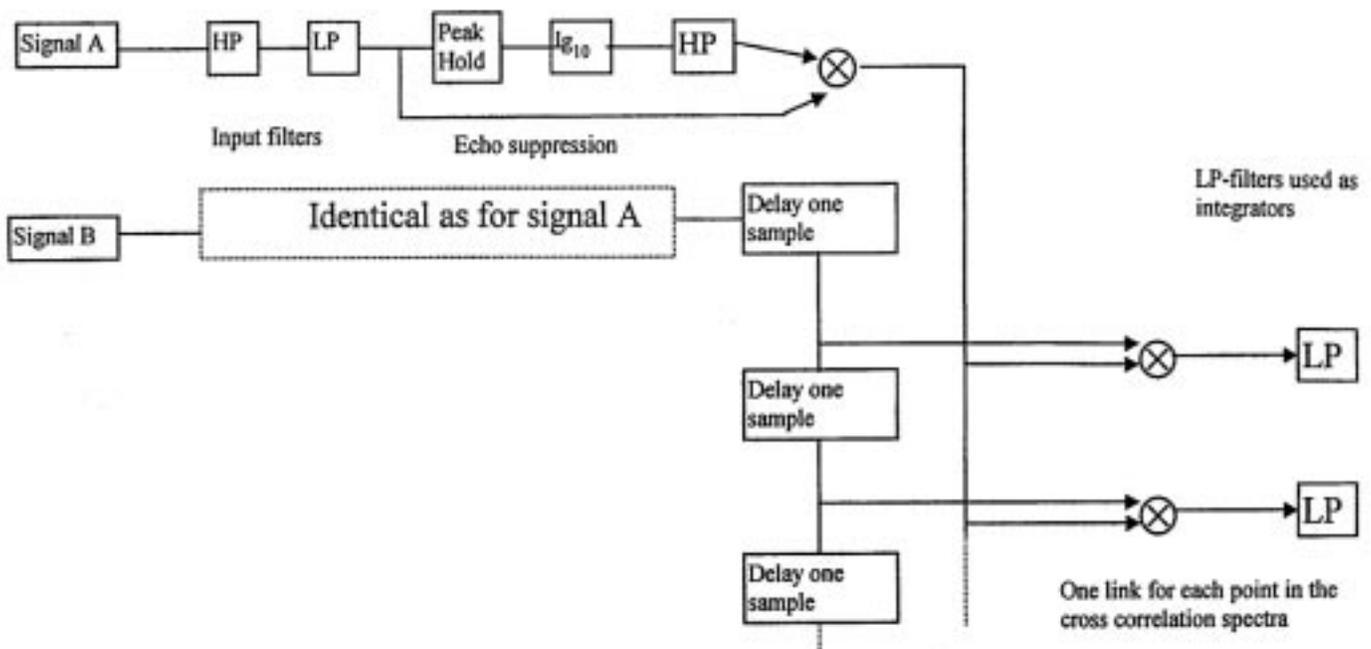


Figure 2.

The DSP implementation as a time discrete model. To obtain a complete analysis, the three input signals are cross-correlated in three pairs. In the correlation between right and left channels, 17 spectral points are used (= the number of links in the picture above), in the two others, 11 points are used.

are used for unambiguous directional analysis. We have assumed that the precision needed for deaf-blind subjects is in the order of 45° , enough to decide the direction for turning of the body or reaching out the arm.

The three microphones are mounted in a triangle. The signals from the microphones were filtered, amplified, and then fed to a DSP board in a personal computer. The microphone signals were then A/D converted at 12 kHz. The input signals from the three microphones were band-pass filtered with selectable cutoff frequencies. In the present studies, the cutoff frequencies have been chosen to match human speech long-time average spectra.

The filtered signals are then fed into echo-suppression circuitry, using the Haas effect. It consists of a peak-hold block with the logarithm calculated on the output, and the signal is differentiated to give more distinct pulses to the correlation function. The decay time of the peak-hold block is adjusted to have the same time as the reverberation decay of the room. In the present study, the parameter values for cutoff frequencies and decay times were chosen by the test leader on the basis of previous pilot testing and were kept constant for all tests and subjects. In the final equipment we plan to have the subject able to adjust the parameters. Kaneda's study (11), using

narrow-band signal, showed that the phase component was strongly affected by reflecting sounds. In the current implementation, also using wide-band signals, an additional branch to include phase information improved the results.

A time-domain cross-correlation is then performed on the preprocessed signals for each microphone pair. The front direction microphone has 17 correlation points, whereas the side pairs have 11 points. The reason for using different numbers of points is that the microphones are not symmetrically mounted. After the cross-correlation computation an exponential decay average is calculated on the correlation data.

The directional analysis algorithm is implemented using the Aladdin Interactive DSP Workbench software (13) and then integrated into a Microsoft Visual Basic application for control. From the application, the test leader can adjust filter cutoff frequencies, threshold level, and decay time constants.

The frequency dependence of the inter-aural time delay was measured and found to be comparable to data in the literature (14). Time delay is not unequivocal for the frequencies above >1.5 kHz. Therefore, for optimal result, the equipment needs to be calibrated for each user.

Testing Procedure

All of the tests were performed in the environmental sound chamber (12), which consists of a sound-treated room designed according to the LEDE principle (15). The test subject was seated in the center of a circular arrangement of twelve loudspeakers (30 degrees apart). **Figure 3** shows the experimental setup with the eyeglasses on the test subject who is seated in the center of the loudspeaker array.

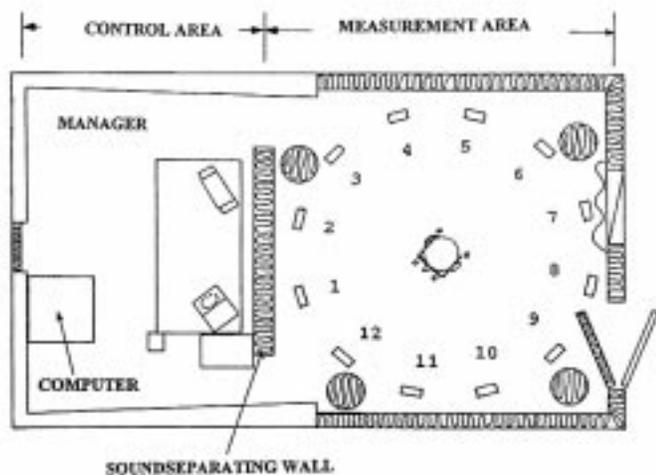


Figure 3
Sound environmental chamber for testing of directional response and presentation of simulated or recorded sound environments.

Since the purpose of this experiment was to evaluate the algorithm, the subjects were only passive carriers of the eyeglasses, and did not indicate sound direction. The reverberation time of the room was below 0.2 seconds, which is shorter than the value for a conventional office or living room. Two tests were used. In the first, the Word Localisation (WL 360) test, monosyllabic words (approximately 0.8-second duration) were presented randomly from the 12 loudspeakers, with a total of 60 presentations. A speech spectrum noise (-6 dB/octave above 1.0 kHz) could be added to the remaining 11 loudspeakers and the signal-to-noise ratio varied. In the second test, the stimuli consisted of 1-second bursts of 1/3 octave noise band (48 dB per octave). White noise masking was used for determination of signal-to-noise ratio for the narrow band noise bursts. The test signals (words, noise bursts) were presented at approximately 5-second intervals from pseudo-randomized directions, a total of 5 in each direction. For the present purpose of determining the capability of the analysis program, the vibrators were disconnected

and the angle calculated by the program was the dependent variable. The eyeglasses were tested both on a polystyrene mannequin head and on human test subjects.

The calibrations were performed with the 1-inch microphone (Briel & Kjaer) placed on the center of the head. The speech level and the level of the speech spectrum noise, the narrow band noise, and the white noise levels were determined as the rms level dBA (described in more detail in reference 12).

RESULTS

Frequency Characteristics

Figure 4 shows the “error,” i.e., the median difference between the calculated angle and the presentation angle of the selected loudspeaker as a function of azimuth for four human subjects. The stimuli were 1/3-octave band noise bursts at the frequencies 500, 1,000, 2,000, and 4,000 Hz at 63 dB SPL presented in a quiet background. The bandwidth of the analysis system was 250–6,000 Hz and the integration time 250 ms. Except for 1,000 Hz at 90-degree azimuth (outside the scale) the errors were small. At this direction there was a 180-degree reversal in more than half of the presentations, the reason for which is probably irregularities in the acoustics of the test room.

Speech Stimuli

Figure 5 shows the median and inter-quartile range (shaded) for speech stimuli from the WL 360 test presented at 73 dB SPL in a quiet background. The value of the median is presented inside the circle and the source angle outside the circle. The results were obtained through placing the eyeglasses on 10 human heads. The frequency limits of the analysis were 250 Hz and 2,500 Hz. As can be seen, there is practically no error at 0 degrees and 180 degrees, and the error is largest around ± 90 degrees, i.e., up to about 15 degrees. The variability is reasonably small. The median values of the calculated directions show a small (0 to 17°) but significant difference from the presentation directions ($p < 0.05$ for all angles except 30°; Wilcoxon-signed ranks test).

Susceptibility to Interfering Noise

The deterioration of the program predictions in noise environments was determined with the WL 360, with speech signals or band-limited noise in the speech frequency range at different interfering noise levels (speech frequency noise for speech signals and white

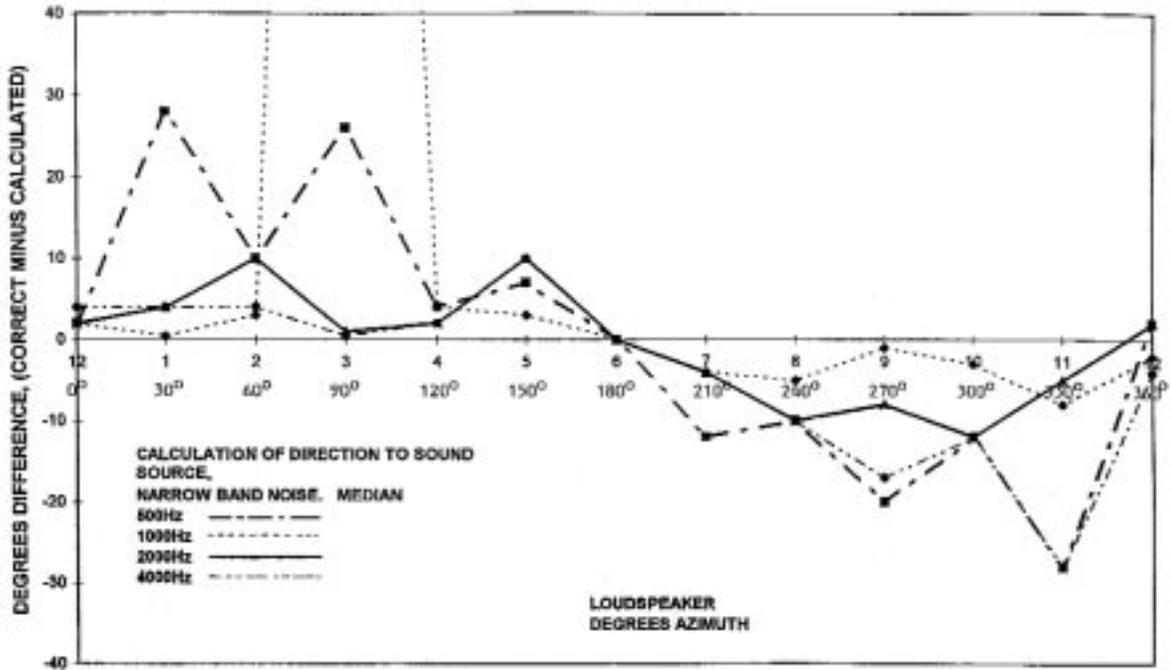


Figure 4. Errors (median) in calculated sound direction, as a function of azimuth for the cross-correlation algorithm of the three-microphone instrument. Tests on four different human heads. Stimuli bursts of 1/3-octave noise bands, 63 dB SPL.

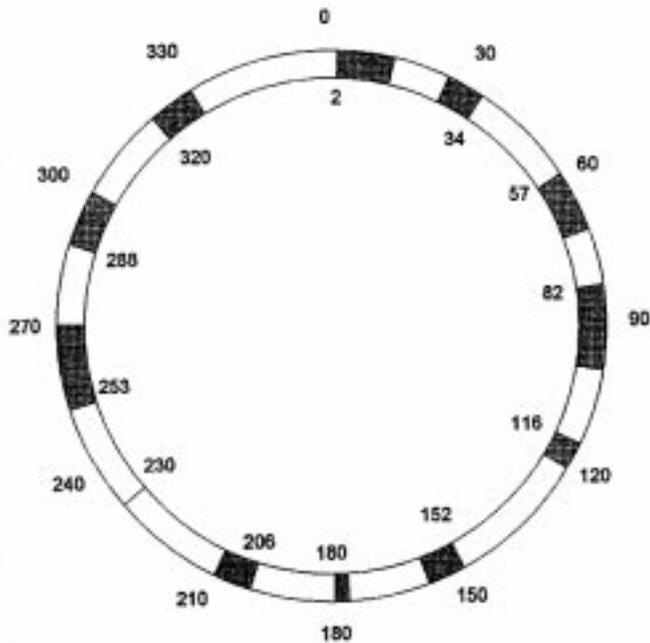


Figure 5. Error (median, inter-quartile range) for calculated direction for the three-microphone system mounted on human heads as a function of azimuth. 73 dB SPL. 10 individuals. Source direction: figures outside the circle. Response: figures inside the circle (median). Inter-quartile range: Shaded area.

noise for band-limited noise stimuli). **Figure 6** shows the median error for four directions of speech presentation at 73 dB SPL as a function of the noise level for five separate measurements on a mannequin head. It is seen that the errors are small for noise levels up to 63–65 dB SPL, i.e., a signal-to-noise ratio of +10 to +8 dB.

A separate test was performed to determine the signal-to-noise ratio for 1/3-octave noise bursts (signal) in a white-noise background. The narrow-band noise was centered at 500, 1,000, 2,000, and 4,000 Hz. The measurements were performed on three human heads. The sound-pressure level at which the analysis program lost track (errors >20 degrees) of the sound source was determined. The sensitivity for different frequencies of the narrow-band noise is shown in **Table 1**, and the role of the presentation angle is shown in **Table 2**. It is evident that the equipment is more robust for low-frequency signals and for signals coming from 0° azimuth.

DISCUSSION

Choice of Algorithm

In previous attempts to use the vibratory sense for sound localization, direct, amplified microphone signals

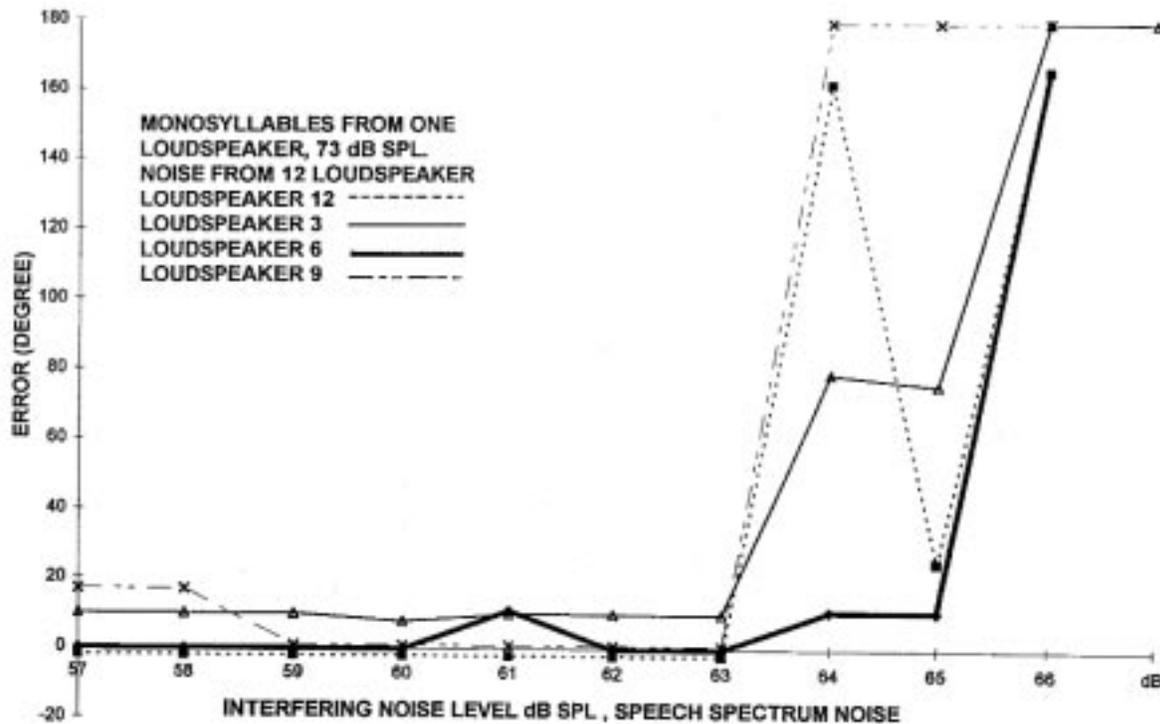


Figure 6.

Noise susceptibility of the three-microphone directional instrument. Monosyllables, 73 dB SPL, were presented from 4 different directions (0, 90, 180, 270 degrees azimuth). Median error as a function of interfering speech frequency noise level. Signal-to-noise ratio was between +8 and +10 dB. (Five separate measurements on a mannequin head.)

Table 1.

Frequency-dependent signal-to-noise ratio for narrow-band noise (63 dB SPL) in white background noise. Average for 3 human heads and 12 signal directions

Frequency	500 Hz	1000Hz	2000Hz	4000Hz	Average
S/N Average	+1 dB	+3 dB	+6 dB	+13 dB	+5 dB

Table 2.

Direction-dependent signal-to-noise ratio for narrow-band noise (63 dB SPL) in white background noise. Average for three human heads and four signal frequencies

Azimuth	0°	90°	180°	270°	Average
S/N Average	0 dB	+6dB	+8 dB	+7 dB	+5 dB

have been utilized (for a review see reference 8). Only rarely has signal processing beyond filtering been used. Gescheider (16) has introduced an artificial increase in delay between the signals to the two vibrators in order to better the low-frequency features of the vibratory sense. He found that an increase of up to 4 msec between stimulus clicks to two vibrators increased the perceived displacement of the sound source. However, a 6-msec

inter-stimulus delay gave rise to two separate sensations. Niioka and colleagues (17) used a time-intensity conversion in their instrument. The vibrators were applied to the fingers, i.e., one of the most sensitive regions for the vibratory sensations.

The results of these studies were surprisingly good, particularly if free head movements and some training were allowed. There has been no analysis on the

robustness in noise environments. We have chosen the Aladdin DSP Workbench because it allows real-time calculations, great flexibility, and it will allow the subjects to control important parameters such as threshold, echo-suppression, frequency limits, *et cetera*. The system can be miniaturized and developed to include certain sound-identification features.

The algorithm is designed to mimic the principles of direction analysis used by the auditory system. At this point, however, we have only used the temporal aspects of the signal. The intensity difference seen at high frequencies can also be included and probably can improve the performance in noisy environments.

There is a lack of data on the performance of previously described equipment for sound localization by cutaneous stimulation in interfering noise. Using the present equipment, the signal-to-noise ratio with narrow-band noise bursts as stimuli is as good as 0 dB at 0-degrees azimuth. For speech stimuli in speech frequency noise, i.e., a difficult listening condition, it is up to +8 dB. This is obviously considerably below normal listening ability where speech recognition is preserved in signal-to-noise ratios around -5 dB. However, it is comparable to the corresponding values for subjects with hearing impairment (18).

Vibratory Coding

The present system is digital and controls four vibrators to indicate eight directions separated by 45 degrees (19). The code is simple: For 0, 90, 180, and 270 degrees azimuth the two front, side, and back vibrators were active, respectively. For the 45-degree directions only one vibrator was activated.

In earlier attempts, only slightly transformed signals have been presented to the vibrators. Weisenberger et al. (9) used two vibrators with low- and high-frequency analogue signals, thus providing both intensity and time information. The results were poor at 500 Hz, but reached, on average, 60-percent correct at 2,000 Hz. Gescheider (16) increased the delay between the two vibrators. With the previous coding system the intensity difference rather than time difference has been important in creating the perceived direction. Niioka and colleagues (17) utilized this property of the skin and converted time to intensity in order to produce clear phantom images. Weisenberger (20) came to the same conclusion as Niioka and colleagues, and explained the poor result at 500 Hz in terms of the poor time resolution of the skin, which they suggested would be improved by a time-intensity conversion.

Our assumption is that the directional information does not require a higher resolution than 45 degrees. Sighted persons can detect objects quickly enough with 45° resolution, and individuals with Usher I and a small visual field will have a considerably improved situation. Deaf-blind persons would then know in which direction to focus their attention and to search for contact with their hands. There is intuitively an advantage to have the vibrators on the head. An analogue signal coding with vibrators on the head is difficult to make reliable, since the cutaneous senses have low accuracy in this region and the underlying bone will cause interaction due to bone conduction between the vibrators (for further information see the review by Borg, reference 8). A digital system offers a wide range of possibilities and, as shown in a companion manuscript (19), the accuracy of digital vibrator coding is high and no training is needed.

Future Developments

The present results indicate that the DSP algorithm has the capacity to be beneficial in listening conditions where hearing aids are particularly useful, i.e., when the signal-to-noise ratio is better than about +10 dB. This value is probably an underestimation of its capacity for everyday listening. The directional device provides information about a wide variety of sounds and is not so focused on human voices as a hearing aid. Rather, results with noise bands are likely to be more valid in describing the real robustness of the device in relevant listening situations. A signal-to-noise ratio around 0 dB is a very promising result with respect to possible practical applications. Since the device is provided with threshold control, the subject can also decide at what level of attention he will monitor the environment. The details of the issue of control over analysis parameters will be studied in future work. The algorithm used will be further improved by using intensity differences. It can be supplemented with facilities for identification of certain environmental sounds such as footsteps, voiced sounds, automobile noises, *et cetera*.

The importance of perception and control of the sound environment has been emphasized only recently by Söderlund (21). Little attention to this issue has been given in the past. Plant (22) touches on the problems of environment and Wallin (23) points out that a drawback of cochlear implants for deaf people is the lack of directional information. The present result shows that using the three-microphone system, directional information can also be extracted in fairly noisy environments.

The potential application of directional information is valuable not only for deaf-blind individuals, but for all those with various degrees of hearing loss, unilateral or bilateral, with deteriorated directional hearing. The benefit is expected to be twofold: benefit in terms of localization *per se*, but perhaps primarily in terms of facilitating speech reading or sign communication.

CONCLUSION

It is concluded a) that it is possible to design a head-mounted three-microphone four-vibrator device for real-time detection, localization, and coding of direction of sound sources; and b) that the precision of the analysis program and the robustness in noise environments is sufficient to warrant further development aiming at wearable equipment useful for deaf and deaf-blind individuals.

ACKNOWLEDGMENTS

We are grateful to Kasper Marklund, MSc, Anders Hjälm, David Josefsson, Mats Wilson, Yvonne Behrenth, and Urban Wiklander for valuable contributions to programming, construction of equipment, and testing, and to Professor Emeritus Arne Risberg, Assistant Professor Karl-Erik Spens, Assistant Professor Gunnar Jansson, and Professor Birger Roos for constituting a reference group for the project and providing valuable comments to the work and the manuscript. This study has been approved by the Ethical Committee, Örebro Medical Centre Hospital, Sweden.

REFERENCES

1. Blauert J. Spatial hearing: the psychophysics of human sound localization. Rev. ed. Cambridge, MA: MIT Press; 1997.
2. Summers IR, editor. Tactile aids for the hearing impaired. London: Whurr Publishers; 1992.
3. Johansson K. A "happy" approach to speechreading: the effects of facial expression, emotional content, and script information on speechreading performance. [dissertation]. Linköping: Linköping University; 1998.
4. Hansson H. Monoauralt döva: audiologiska, socialpsykologiska och existentiella aspekter. (Monaural deaf persons: audiological, social psychological and existential aspects). [dissertation]. Stockholm: Stockholm University; 1993.
5. Rosenthal DF, Okuno HG, editors. Computational auditory scene analysis. Mahwah, NJ: Lawrence Erlbaum Associates; 1998.
6. Reed CM. Tadoma: an overview of research. In: Plant G, Spens KE, editors. Profound deafness and speech communication. London: Whurr Publishers; 1995. p. 40–55.
7. Rönnberg J, Borg E. A review and evaluation of research on the deaf-blind from perceptual, linguistic-communicative, and social-rehabilitative aspects. Scand Audiol. In press 2001.
8. Borg E. Cutaneous senses for detection and localization of environmental sound sources: a review and tutorial. Scand Audiol 1997;26:195–206.
9. Weisenberger JM, Heidbreder AF, Miller JD. Development and preliminary evaluation of an earmold sound-to-tactile aid for the hearing impaired. J Rehabil Res Dev 1987;24:51–66.
10. Rönnberg J, Borg E, Samuelsson E. Strategies and experiences for monitoring of environmental events. Scand Audiol. In press 2001.
11. Kaneda Y. Sound source localization for wide-band signals under a reverberant condition. J Acoust Soc Jap (E) 1993;14:47–8.
12. Borg E, Wilson M, Samuelsson E. Towards an ecological audiology: stereophonic listening chamber and acoustic environmental tests. Scand Audiol 1998;27:195–206.
13. Neovius L. Parameter-controlled synthesizers in speech research and applied text-to-speech systems. [dissertation]. Stockholm: Royal Institute of Technology; 1995.
14. Kuhn GF. Physical acoustics and measurements pertaining to directional hearing. In: Yost WA, Gourevitch G, editors. Directional hearing. Berlin: Springer Verlag; 1987. p. 3–25.
15. Davis D. The LEDE-concept. Audio Magazine 1987;15:48–58.
16. Gescheider GA. Cutaneous sound localisation. J Exp Psychol 1965;70:617–25.
17. Niioka T, Ifukube T, Yoshimoto CF. Basic studies of a tactual sound localizer for the deaf. J Acoust Soc Jap 1977;33:250–8.
18. Hagerman B. Clinical measurements of speech reception threshold in noise. Scand Audiol 1984;13:57–63.
19. Borg E, Rönnberg J, Neovius L. Vibratory coded directional analysis: evaluation of a three-microphone/four-vibrator DSP system. J Rehabil Res Dev 2001;38:257–263.
20. Weisenberger JM. Communication of the acoustic environment via tactile stimuli. In: Summers IR, editor. Tactile aids for the hearing impaired. London: Whurr Publishers; 1992. p. 83–189.
21. Söderlund G. Tactiling and tactile aids: a user's viewpoint. In: Plant G, Spens KE, editors. Profound deafness and speech communication. London: Whurr Publishers; 1995. p. 25–39.
22. Plant GL. The use of tactile supplements in rehabilitation of the deafened: a case study. Aust J Audiol 1979;1:76–82.
23. Wallin A. The cochlea implant: a weapon to destroy deafness or a supplement for lipreading. A personal view. In: Plant G, Spens KE, editors. Profound deafness and speech communication. London: Whurr Publishers; 1995. p. 219–30.

Submitted for publication July 11, 2000. Accepted in revised form December 7, 2000.