Department of Veterans Affairs

Journal of Rehabilitation Research and Development Vol. 30 No. 1, 1993
Pages 95–109

Digital signal processing (DSP) applications for multiband loudness correction digital hearing aids and cochlear implants

Norbert Dillier, PhD; Thomas Frölich, DiplIngETH; Martin Kompis, DiplIngETH; Hans Bögli, DiplIngETH; Waikong K. Lai, PhD
Department of Otorhinolaryngology, University Hospital, CH-8091 Zürich, Switzerland

Abstract—Single-chip digital signal processors (DSPs) allow the flexible implementation of a large variety of speech analysis, synthesis, and processing algorithms for the hearing impaired. A series of experiments was carried out to optimize parameters of the adaptive beamformer noise reduction algorithm and to evaluate its performance in realistic environments with normal-hearing and hearing-impaired subjects. An experimental DSP system has been used to implement a multiband loudness correction (MLC) algorithm for a digital hearing aid. Speech tests in quiet and noise with 13 users of conventional hearing aids demonstrated significant improvements in discrimination scores with the MLC algorithm. Various speech coding strategies for cochlear implants were implemented in real time on a DSP laboratory speech processor. Improved speech discrimination performance was achieved with high-rate stimulation. Hybrid strategies incorporating speech feature detectors and complex decision algorithms are currently being investigated.

Key words: *digital hearing aids, experimental digital signal processor, multiband loudness correction, noise reduction algorithm, speech coding strategies for cochlear implants.*

INTRODUCTION

New generations of fast and flexible single-chip digital signal processors (DSPs) allow the implementation of sophisticated and complex algorithms in real time which required large and expensive hardware only a few years ago. Signal processing algorithms such as nonlinear multiband loudness correction, speech feature contrast enhancement, adaptive noise reduction, speech encoding for cochlear implants, and many more offer new opportunities for the hard-of-hearing and the profoundly deaf. A recent review report on the status of speech-perception aids for hearing-impaired people came to the final conclusion that major improvements are within our grasp and that the next decade may yield aids that are substantially more effective than those now available (1).

This paper concentrates on three areas of DSP applications for the hearing impaired and investigates algorithms and procedures which have the potential of being effectively utilized and integrated into existing concepts of rehabilitation of auditory deficits in the near future. The first area concerns the problem of interfering noise and its reduction through an adaptive filter algorithm. The second area deals with the reduced dynamic range and distorted loudness perception of sensorineurally deaf persons and its compensation through a multiband loudness correction algorithm. The third area covers new processing strategies for profoundly deaf persons with cochlear implants. While the first DSP application does not rely on the individual hearing loss data of a subject, and thus could serve as a general preprocessing stage for any kind of auditory

Address all correspondence and requests for reprints to: N. Dillier, Dept. of Otorhinolaryngology, University Hospital, CH-8091 Zürich, Switzerland.

prosthesis, the second and third applications are intimately related to psychophysical measurements and individual fitting procedures. In the case of cochlear implant processing strategies, a number of technical and physiological constraints have to be considered. Although these three areas could be considered as three different projects that would need to be treated separately in more detail, there are many common aspects of signal processing and experimental evaluation. The use of similar tools and methods has been beneficial and fruitful throughout the theoretical and practical development of the respective projects. The complexity and interdisciplinary nature of the current auditory prostheses research problems requires collaboration and teamwork which, it is hoped, will lead to further practical solutions.

EVALUATION OF THE ADAPTIVE BEAMFORMER NOISE REDUCTION SCHEME FOR HEARING-IMPAIRED SUBJECTS

A major problem for hearing aid users is the reduced intelligibility of speech in noise. While the hearing aid is a great relief for conversations in quiet, its usefulness is often drastically reduced when background noise, especially from other speakers, is present. Several single-microphone noise reduction schemes have been proposed in the past which can improve sound quality but often failed to improve intelligibility, when tested with a normal-hearing or hearing-impaired subject in real life situations such as competing speakers as noise sources (2,3,4,5,6).

Multimicrophone noise reduction schemes, such as the adaptive beamformer (7,8,9), make use of directional information and are potentially very efficient in separating a desired target signal from intervening noise. The adaptive beamformer enhances acoustical signals emitted by sources coming from one direction (e.g., in front of a listener) while suppressing noise coming from other directions. Thus, the effect is similar to the use of directional microphones or microphone arrays. Because of the adaptive postprocessing of the microphone signals, the adaptive beamformer is able to combine the high directivity of microphone arrays with the convenience of using only two microphones placed in or just above the ears.

The aim of our investigation was to optimize the algorithm for realistic environments, implement a real-time version, and estimate its usefulness for future hearing aid applications.

Method

Parameter Optimization. Based on the work of Peterson, et al. (8), a Griffiths-Jim beamformer was chosen with two microphones placed at each ear. In the first stage, the sum and difference of the two microphone signals were calculated. The former contained mainly the desired signal, the latter mainly noise. A self-adaptive filter then tried to cancel out as much of the remaining noise in the desired signal path as possible.

While the system works quite well in anechoic rooms, its performance is not very satisfactory in reverberant conditions, because a part of the desired signal will also come from other than the front direction. In that case, the adaptive beamformer will mistake the desired signal for noise and try to cancel it as well. Therefore, it is necessary to detect the presence of a target signal and to stop the adaptation of the filter during these intervals (10,11). The method of target signal detection implemented in our system is based on the comparison of the variance of the sum (Sigma, Σ) and difference (Delta, Δ) signals which are easily obtained by squaring and lowpass-filtering the frontend signals of the Griffiths-Jim beamformer. An optimal threshold value of 0.6 ($\Sigma/[\Sigma + \Delta]$) was determined empirically through variations of the noise signal, filter parameters, and room acoustics. Other procedures have been considered as well but were either less efficient or computationally too expensive. We found that the performance of the adaptive beamformer can be notably increased by optimizing the adaptation inhibition.

Figure 1 shows measurements of the directivity characteristic with and without adaptation inhibition in a moderately reverberant room (RT = 0.4 sec). Because the microphones at the ears of a dummy head were omnidirectional, the beam not only pointed to the front (0°), but also to the back (180°) of the listener. It can be seen that the adaptation inhibition increased the difference between two sources located at an angle of 0° and 90° from around 6 dB to approximately 10 dB. The sharp profile of the beam with the adaptation inhibition present, when compared with the rather smeared

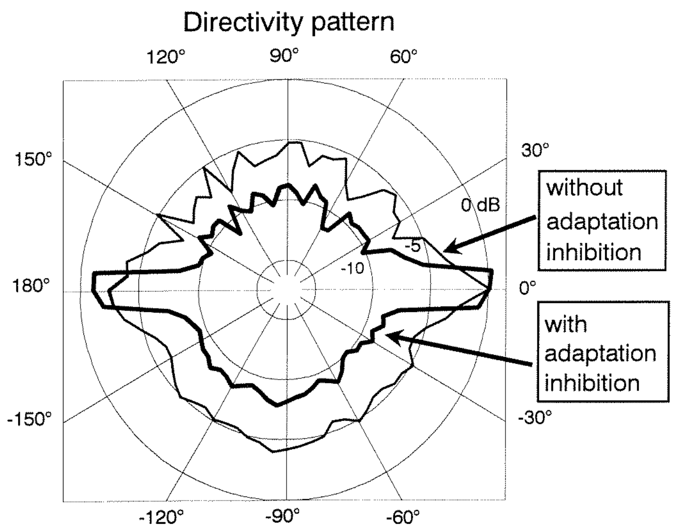

Figure 1.
Effect of speech detection and adaptation inhibition on the directivity pattern of the adaptive beamformer. Front direction (nose): 0°, left ear: 90°.

pattern without adaptation inhibition, is a result of the target signal detection scheme which not only enhances the directivity of the adaptive beamformer, but is also responsible for the shape and opening angle of the beam.

The adaptive beamformer depends on a great variety of acoustic and design parameters. To analyze the sensitivity of these parameters, theoretical as well as experimental investigations were carried out based on adaptive filter and room acoustics theories. A computer program was written which prompts the user for room acoustic and design parameters, as well as directivity and position of a target and a noise source. The program then predicts the improvement in signal-to-noise ratio (SNR) which can be reached with a perfectly adapted beamformer. The predictions were experimentally verified. It is beyond the scope of this paper to discuss the theoretical analysis in detail or to describe the influence of all parameters investigated.

Three parameters were found to influence the performance of the adaptive beamformer most: the reverberation time of the room, the length of the adaptive filter, and the amount of delay in the desired signal path of the beamformer (12).

Figure 2 shows measurement data from an experimental setup, where white noise was generated by two loudspeakers located at a distance of 1 m from a dummy head. The desired signal source was placed at 0°, the noise source at 45° to the right of the head. Thus, the two microphone signals picked up SNRs as indicated in **Figure 2** by squares (comparing the output of the beamformer with the signal of the microphone which is directly irradiated by the noise source) and triangles (comparing the output with the contralateral microphone signal).

The variation of reverberation time (**Figure 2**, left) was achieved by performing the measurements in a number of different rooms whose acoustic characteristics had been determined. It can be seen that a moderate amount of reverberation reduces the effectiveness of the beamformer to a few decibels.

The reverberation time of rooms where a noise reduction would be most advantageous can usually not be influenced. The design parameters of the adaptive beamformer however can be optimized. The most important of them is the length of the adaptive filter. When varying this parameter, it can be seen (**Figure 2**, right) that very short filters are able to switch to the microphone signal with the better SNR, while filters of at least about 500 taps in length are required to reach an additional 4 dB.

A computationally inexpensive way to optimize the adaptive beamformer is the selection of an optimal delay in the desired signal path. We found that for longer filters such as a 512 tap filter, the choice of a reasonable delay is important. It was also found that the optimum depends on the actual acoustical setup and will lie between 25 percent and 50 percent of the filter length.

Implementation. A real-time version of the adaptive beamformer was implemented on a PC-based, floating point digital signal processor (TMS320C30, Loughbourough Sound Images Ltd.). For the update of the adaptive filter, a least mean squares (LMS) adaptation algorithm (13) was employed. After evaluating several different adaptation algorithms, we found that this widely used and well-understood algorithm is still the best choice for our hearing aid application. An adaptation inhibition, as discussed above, was implemented.

With this system, filters of up to 500 taps in length at a sampling rate of 10,000 per second can be realized. For our implementation, a high-performance DSP had to be used whose battery power requirements are still excessive for a commercial hearing aid. The system is, however, flexible enough to allow an evaluation of the main aspects of the

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

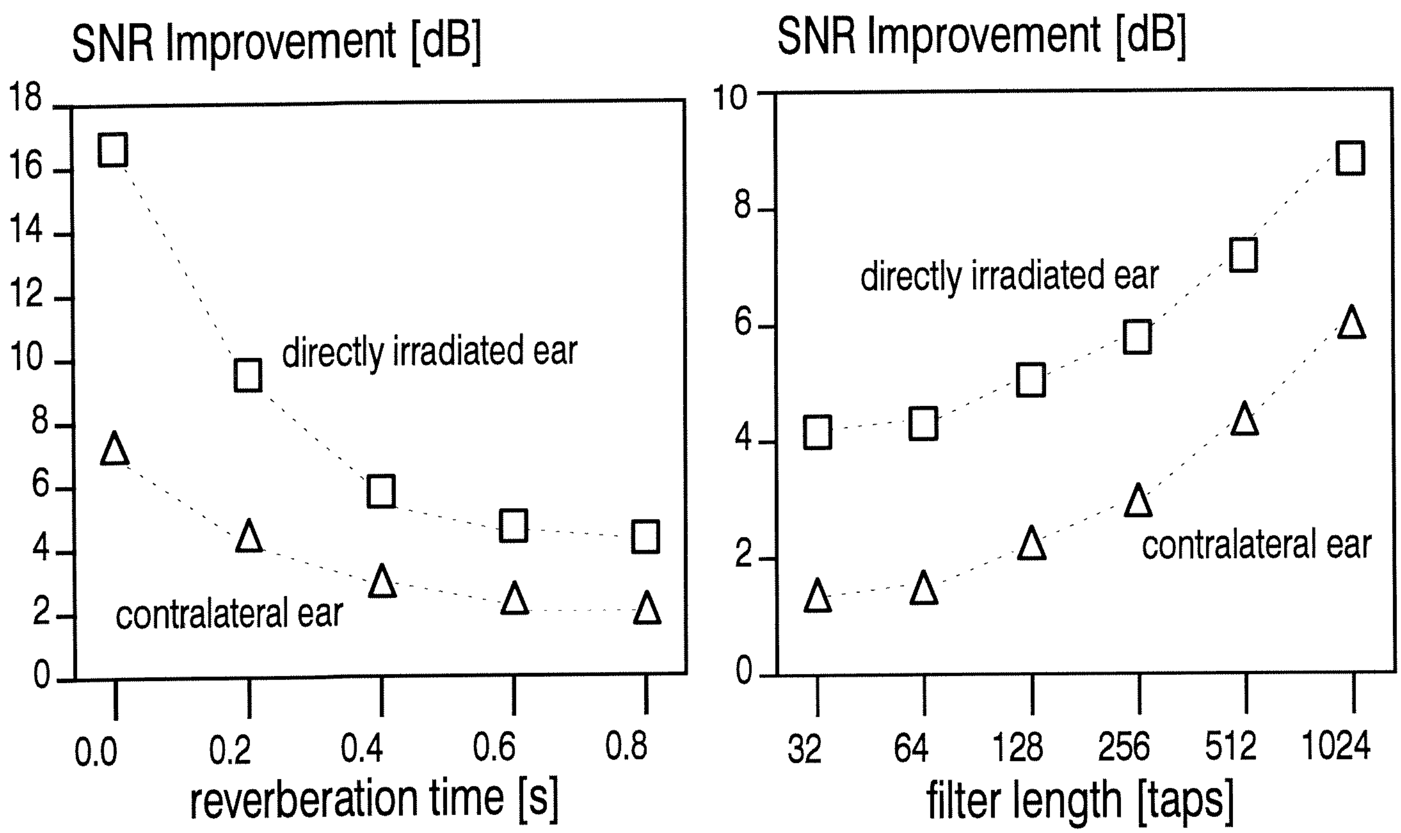

Figure 2.
SNR improvement as function of reverberation time (left) and filter length (right).

algorithm with the potential of future miniaturization.

Evaluation Experiments. To ensure that the adaptive beamformer really is a useful algorithm for hearing aid users, intelligibility tests were carried out with normal-hearing and hearing-impaired subjects. Speech test items were presented by a computer and consisted of two-syllable words with either the medial vowels or consonants forming phonological minimal pairs. Four response alternatives were displayed on a touch-sensitive computer display from which the subjects had to select a response by pointing to it with a finger. No feedback and no repetitions of test items were provided. Fifty or one hundred words per condition were presented and the number of correctly identified items were corrected for the chance level. Phoneme confusion matrices could be generated for further analysis of transmitted information (14).

Most of the test material was recorded in a test room with an average reverberation time of 0.4 sec. Speech was presented via a frontal loudspeaker at 70 dB sound pressure level (SPL). Speech-spectrum-shaped noise which was amplitude-modulated at a random rate of about 4 Hz was presented 45° to the right at different SNRs. A dummy head with two microphones located in the left and right conchas was used to pick up sounds. The distance between the dummy head and either of the two loudspeakers was 1 m. In the experiments with normal-hearing subjects, the sounds were presented directly via earphones monaurally or binaurally (unprocessed conditions) or binaurally after modification by the adaptive beamformer (processed condition). In the experiments with the hearing aid users, the unprocessed or processed sounds were presented via a loudspeaker in a sound-treated room.

This setup was thought to represent some important aspects of everyday listening situations: 1) the size and reverberation time of the room is typical for offices and living rooms in our surrounding; 2) a fair amount of head shadow is provided by the dummy head; and/or, 3) because of the integrated adaptation inhibition, the adaptation of the

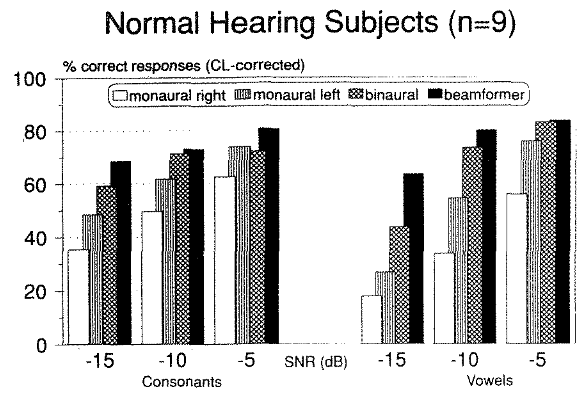

Figure 3.
Evaluation of the adaptive beamformer with normal hearing subjects (n = 9).

beamformer did not have to be performed before the experiments were started. To include even more realistic situations such as moving and multiple sound sources, a part of the investigation was performed with stereophonic recordings of cafeteria noise.

Intelligibillity tests were first performed with nine normal-hearing subjects. As can be seen in **Figure 3**, intelligibility with the adaptive beamformer is highest compared with either one of the microphone signals or even with the binaural listening situation. This is true for all SNRs tested and consonant as well as vowel tests, the largest increase in intelligibility being achieved at low SNRs. Most of these differences were statistically significant.

The noise reduction scheme is primarily meant to be useful to hearing-impaired persons, and so it was also tested with six hearing-impaired volunteers, all of whom were regular users of conventional hearing aids **(Figure 4a)**. As for the normal-hearing subjects, the adaptive beamformer increased intelligibility, when compared with either one of the two unprocessed microphone signals.

The ability of the beamformer to cope with acoustically complex situations and several competing speakers as noise sources is documented by the increased intelligibility in the cafeteria setup **(Figure 4b)**. Note, however, that the differences between the two monaural conditions (directly irradiated and contralateral ear) could not be seen anymore, due to multiple sound sources from all directions. At 0 dB SNR, the beamformer did not provide any benefit for consonant recognition, which is probably due to erroneous behavior of the speech-detector algorithm

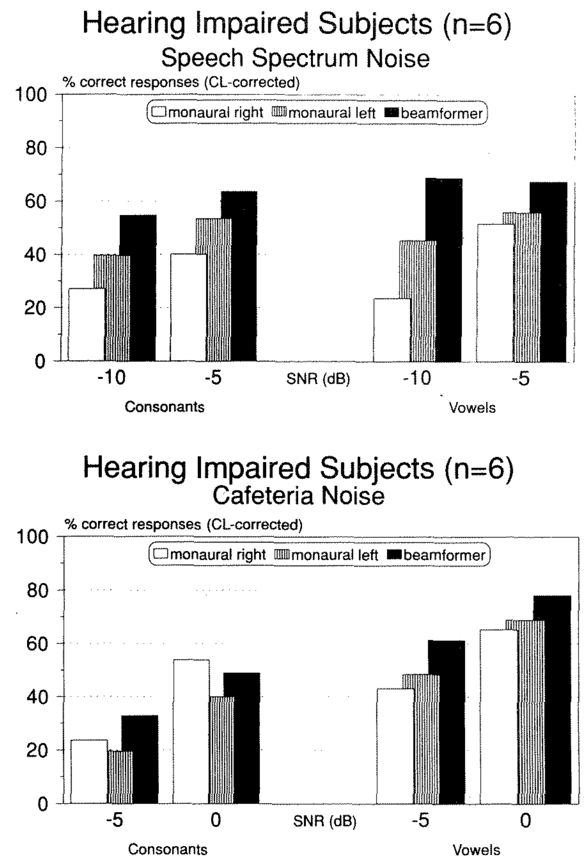

Figure 4.
Evaluation results of hearing-impaired subjects listening with their own hearing aid either to the unprocessed right or left ear signal or to the output of the adaptive beamformer. a) Speech spectrum shaped noise. b) With cafeteria noise.

in some conditions. There was, however, still an improvement for vowel recognition which was statistically significant.

Conclusions

From our investigation, we conclude that three conditions are important for the design of an adaptive beamformer noise reduction for hearing aids: 1) an adaptation inhibition has to be provided; 2) filters of no less than approximately 500 taps should be used; and, 3) the delay should be set to a reasonable value, preferably somewhere between 25 and 50 percent of the filter length. When these requirements are met, the adaptive beamformer is able to improve intelligibility for normal-hearing as well as for hearing-impaired subjects, even in rooms with a realistic amount of reverberation and in complex acoustic situations, such as a cafeteria.

Journal of Rehabilitation Research and Development Vol. 30 No. 1 1993

DIGITAL MULTIBAND LOUDNESS CORRECTION HEARING AID

In this section, an implementation of a new algorithm for a digital hearing aid is described which attempts to correct the loudness perception of hearing-impaired persons. A variety of multiband signal processing algorithms have been proposed and tested in the past (15,16,17,18,19,20,21,22). Some have failed, others showed large improvements in speech recognition scores. The aim of our study was the real-time implementation of an algorithm which restores normal loudness perception similar to the concept suggested by Villchur (23) and others.

A new aspect of our implementation concerns the close relation between psychoacoustic measurements and the ongoing analysis of the incoming signal to determine the required gains in the different frequency regions.

Methods

The digital master hearing aid consists of a DSP board for a 386-PC (Burr Brown PCI 20202C-1 with TMSC320C25) with analog-to-digital (A/D) and digital-to-analog (D/A) converters. Programmable filters and attenuators (custom-made PC-board) are used to prevent aliasing and to control the levels of the signals. In addition to the laboratory version, a wearable device was built which has been tested in preliminary field trials. The processing principle is similar to the frequency domain approach described by Levitt (24) and is illustrated in **Figure 5**.

Magnitude estimation procedures are used to determine loudness growth functions in eight frequency bands that are subsequently interpolated to obtain an estimate of the auditory field of a subject. Sinewave bursts of 8–10 different intensities within the hearing range are presented in random order. After each presentation, the subject judges the loudness of the sound on a continuous scale with annotated labels ranging from very soft to very loud. The responses are entered via a touch screen terminal. The whole procedure is repeated at eight different frequencies.

Figure 5 displays an example of such a magnitude estimation. It can be seen that the function of the hearing-impaired subject starts at a much higher intensity and is much steeper than the curve of the normal-hearing group. The difference between the two curves provides the intensity dependent hearing loss function at this frequency.

The sound signal is sampled at a rate of 10 kHz, segmented into windowed blocks of 12.8 ms and transformed via fast Fourier transform (FFT) into the frequency domain. The amplitude spectrum is modified according to preprogrammed gain tables. The modified spectrum is then transformed back into the time domain via an inverse FFT and reconstructed via an overlap-add procedure. The modification of the spectrum is done by multiplying each amplitude value with a gain factor from one of the gain tables. The selection of the tables is controlled by the amplitude values in the different frequency regions. Thus, the signal is nonlinearly amplified as a function of frequency and intensity.

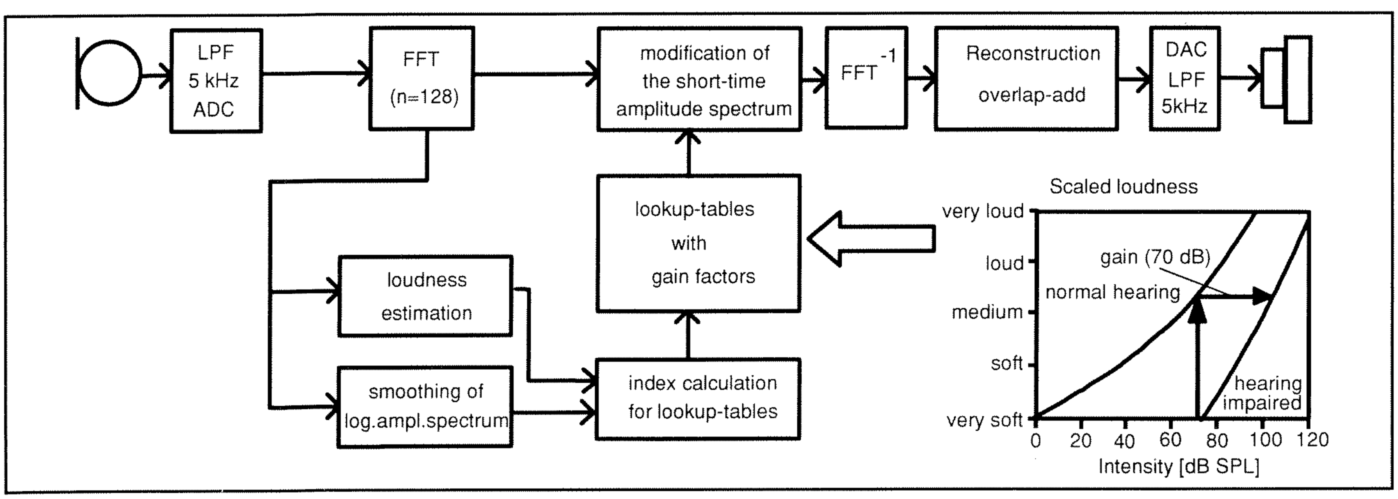

Figure 5.
Processing steps for digital multiband loudness correction algorithm.

Gain tables are generated by interpolation of the measured hearing loss measurements over the whole frequency range. These functions describe how much gain is needed for every frequency to restore normal loudness perception. To prevent a spectrum flattening through independent modification of adjacent amplitudes, a smoothed spectrum is used to determine the gain factors. Thus, the fine structure of the spectrum will be preserved.

However, by determining the gain factors directly from the amplitude values, the reconstructed signal would generally become too loud, because the hearing loss functions were measured using narrow band signals and the incoming sounds consist mainly of broadband signals. The procedure was therefore modified to account for the loudness perception of complex sounds as follows (see **Figure 6** for an example of the processing steps): the frequency (f_{sin}) and amplitude (L_{inp}) of a sinewave signal are estimated which would produce a perceived loudness equal to that of the incoming complex signal. In this simple loudness estimation model, the intensity of this equivalent sinewave signal is determined as the total energy of the complex signal and its frequency as the frequency of the spectral gravity of the amplitude spectrum.

The spectral gravity was chosen because it always lies near the maximal energy concentration which is mainly responsible for loudness perception. The correct gain factors which will restore normal loudness perception (SL_N) are thus obtained by using the raised smoothed spectrum through the point of the equivalent sinewave signal.

Evaluation Results

The algorithm was tested with 13 hearing-impaired subjects in comparison with their own hearing aids under four different conditions in order to find out whether the scaling and fitting procedures were useful approaches for hard-of-hearing subjects and whether or how much the speech intelligibility could be improved with the multiband loudness correction (MLC) algorithm in quiet and noisy environments. The subjects had only minimal experience with the laboratory hearing aid, whereas they had used their own aids for more than a year. The function and fitting of the hearing aids were checked prior to the speech tests. Even though some hearing aids were equipped with an automatic gain control, the threshold of compression was set higher than the levels used in the experiments. Subjects with moderate to severe flat sensorineural hearing loss were selected, which allowed simultaneous processing of the whole frequency range with sufficient numerical resolution. The same test procedure was used as in the previous section.

Figure 7 displays speech intelligibility scores for a consonant and vowel multiple choice rhyme test. At the higher intensity level (70 dB, open squares), which would correspond to an average speech communication level in quiet rooms, the scores with the subject's own hearing aids were rather high, especially in the vowel test. The difference between the new algorithm and the subject's own hearing aid was about 10 percent.

At reduced presentation levels (60 dB, open circles) which would be characteristic for a more difficult communication situation, with a speaker talking rather softly or from a more remote position, the scores with the conventional hearing aids became significantly lower for most subjects. The advantage of the loudness correction becomes more apparent. Almost all subjects achieved scores of from about 80 to 90 percent correct-item-identification, which was about as high as in the 70 dB condition. Subjects with very poor results with their own hearing aids profited most from the digital hearing aid.

In noisy conditions where the hearing aid users experience most communication problems at an SNR of 0 and -5 dB (filled triangles and diamonds, respectively) the results were different from subject to subject. Most of them achieved higher scores with the digital hearing aid. Two subjects scored slightly worse with the DSP algorithm. One subject could not discriminate the words under this condition with her own hearing aid; with the MLC algorithm she obtained a score close to 50 percent.

Conclusions

The automated audiometric procedures to assess the frequency and intensity specific amount of hearing loss proved to be adequate for the calculation of processing parameters. The 13 subjects did not experience major problems in carrying out the measurements. All subjects who had not reached 100 percent intelligibility with their own hearing aids showed improved intelligibility of isolated words with the digital processing, especially at low levels. Some subjects also showed substantial discrimina-

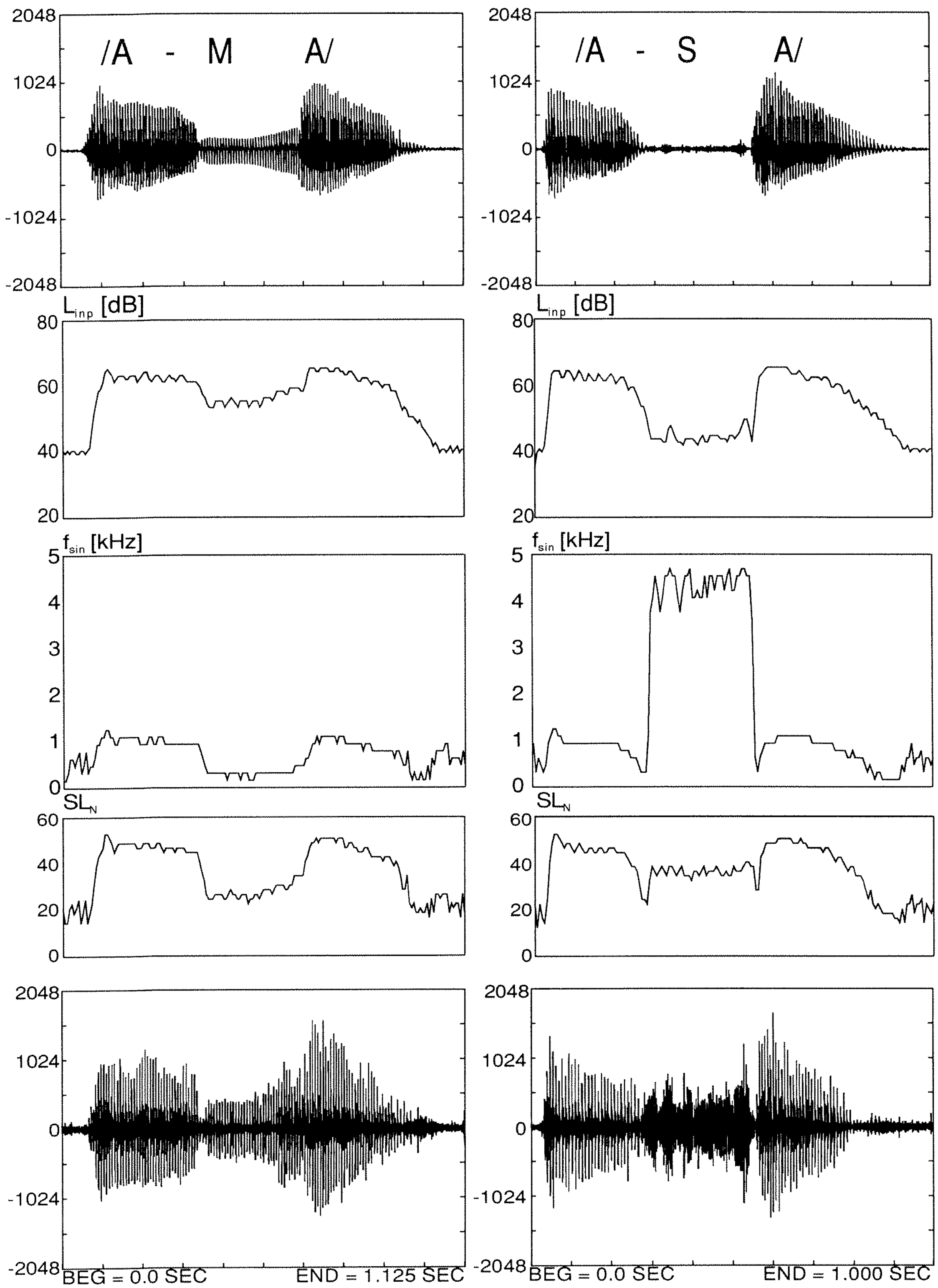

Figure 6.
Input time signals (top), input levels L_{inp} (corresponding to perceived loudness of narrow band signal), spectral gravity f_{sin}, equivalent scaled loudness SL_N, processed output signals (bottom). Speech tokens /A-MA/ (left) and /A-SA/ (right).

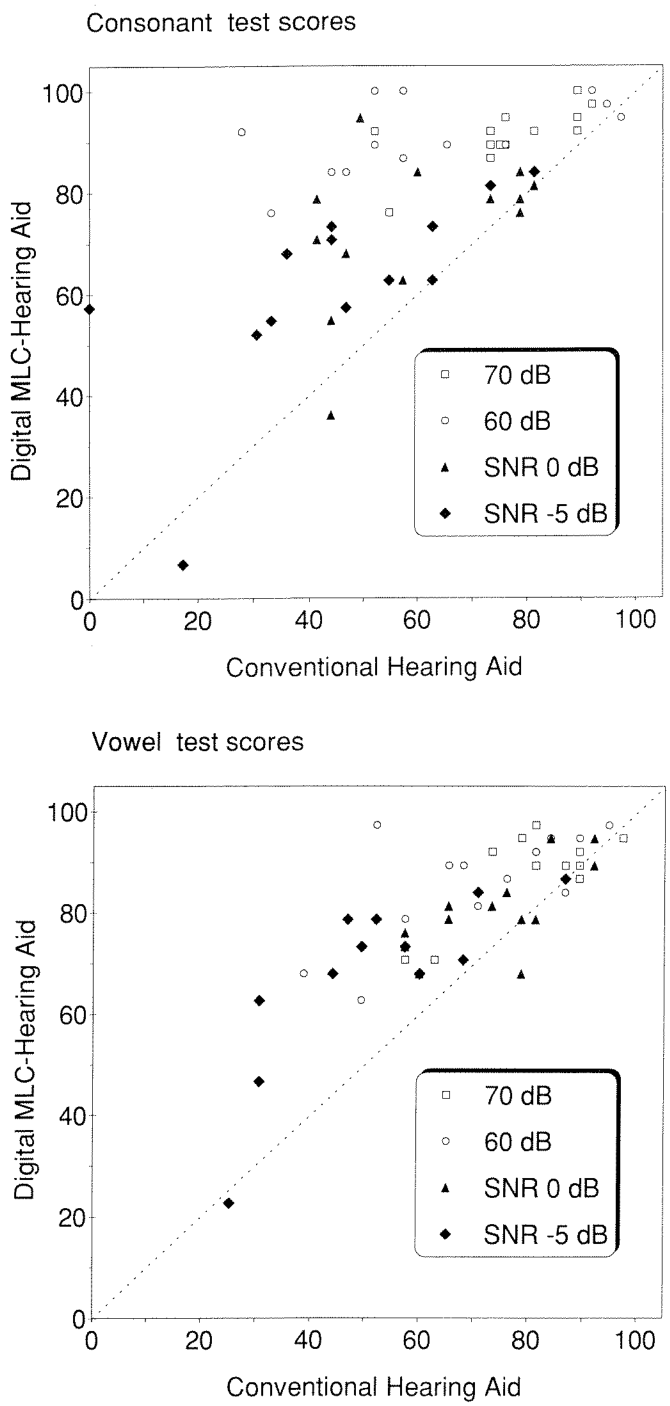

Figure 7.
Comparison of speech intelligibility tests with own conventional hearing aid and digital MLC hearing aid for 13 subjects. Open symbols: tests in quiet; filled symbols: tests in speech spectrum shaped noise. a) Consonant test scores. b) Vowel test scores.

tion improvements in background noise, while others did not. A correlation between the hearing loss functions and the speech intelligibility scores was not found. Other phenomena, such as reduced frequency and temporal resolution, might have to be considered in addition to loudness recruitment. It can be expected that fine tuning of processing parameters via, for example, a modified simplex procedure (25), or prolonged experience with a wearable signal processing hearing aid might further improve the performance of the MLC algorithm.

DIGITAL SIGNAL PROCESSING STRATEGIES FOR COCHLEAR IMPLANTS

Major research and development efforts to restore auditory sensations and speech recognition for profoundly deaf subjects have been devoted in recent years to signal processing strategies for cochlear implants. A number of technological and electrophysiological constraints imposed by the anatomical and physiological conditions of the human auditory system have to be considered. One basic working hypothesis for cochlear implants is the idea that the natural firing pattern of the auditory nerve should be as closely approximated by electrical stimulation as possible. The central processor (the human brain) would then be able to utilize natural ("prewired" as well as learned) analysis modes for auditory perception. An alternative hypothesis is the Morse code idea, which is based on the assumption that the central processor is flexible and able to interpret any transmitted stimulus sequence after proper training and habituation.

Both hypotheses have never really been tested for practical reasons. On the one hand, it is not possible to reproduce the activity of 30,000 individual nerve fibers with current electrode technology. In fact, it is even questionable whether it is possible to reproduce the detailed activity of a single auditory nerve fiber via artificial stimulation. There are a number of fundamental physiological differences in firing patterns of acoustically versus electrically excited neurons which are hard to overcome (26). Spread of excitation within the cochlea and current summation are other major problems of most electrode configurations. On the other hand, the coding and transmission of spoken language requires a much larger communication channel bandwidth and more sophisticated processing than a Morse code for written text. Practical experiences with cochlear implants in the past indicate that some natural relationships (such as growth of loudness and voice pitch variations) should be maintained in the encoding process. One might therefore conceive a third, more realistic, hypothesis as follows: Signal processing for cochlear implants should carefully select a subset of the total information contained in the sound signal and transform these elements into those physical stimulation parameters which can generate distinctive perceptions for the listener.

Many researchers have designed and evaluated different systems varying the number of electrodes and the amount of specific speech feature extraction and mapping transformations used (27). Recently, Wilson, et al. (28) reported astonishing improvements in speech test performance when they provided their subjects with high-rate pulsatile stimulation patterns rather than analog broadband signals. They attributed this effect partly to the decreased current summation obtained by nonsimultaneous stimulation of different electrodes (which might otherwise have stimulated in some measure the same nerve fibers and thus interacted in a nonlinear fashion) and partly to a fundamentally different, and possibly more natural, firing pattern due to an extremely high stimulation rate. Skinner, et al. (29) also found significantly higher scores on word and sentence tests in quiet and noise with a new multipeak digital speech coding strategy as compared with the formerly used F0F1F2-strategy of the Nucleus-WSP (wearable speech processor).

These results indicate the potential gains which may be obtained by optimizing signal processing schemes for existing implanted devices. The present study was conducted in order to explore new ideas and concepts of multichannel pulsatile speech encoding for users of the Clark/Nucleus cochlear prosthesis. Similar methods and tools can, however, be utilized to investigate alternative coding schemes for other implant systems.

Signal Processing Strategies

A cochlear implant digital speech processor (CIDSP) for the Nucleus 22-channel cochlear prosthesis was designed using a single-chip digital signal processor (TMS320C25, Texas Instruments) (30,31). For laboratory experiments, the CIDSP was incorporated in a general purpose computer which provided interactive parameter control, graphical display of input/output and buffers, and offline speech file processing facilities. The experiments described in this paper were all conducted using the laboratory version of CIDSP.

Speech signals were processed as follows: after analog low-pass filtering (5 kHz) and A/D conversion (10 kHz), preemphasis and Hanning windowing (12.8 ms, shifted by 6.4 ms or less per analysis frame) was applied and the power spectrum calculated via FFT; specified speech features such as formants and voice pitch were extracted and transformed according to the selected encoding strategy; finally, the stimulus parameters (electrode position, stimulation mode, and pulse amplitude and duration) were generated and transmitted via inductive coupling to the implanted receiver. In addition to the generation of stimulus parameters for the cochlear implant, an acoustic signal based on a perceptive model of auditory nerve stimulation was output simultaneously.

Two main processing strategies were implemented on this system: The first approach, Pitch Excited Sampler (PES), is based on the maximum peak channel vocoder concept whereby the time-averaged spectral energies of a number of frequency bands (approximately third-octave bands) are transformed into appropriate electrical stimulation parameters for up to 22 electrodes (**Figure 8**, top). The pulse rate at any given electrode is controlled by the voice pitch of the input speech signal. A pitch extractor algorithm calculates the autocorrelation function of a lowpass-filtered segment of the speech signal and searches for a peak within a specified time lag interval. A random pulse rate of about 150 to 250 Hz is used for unvoiced speech portions.

The second approach, Continuous Interleaved Sampler (CIS), uses a stimulation pulse rate that is independent of the fundamental frequency of the input signal. The algorithm continuously scans all frequency bands and samples their energy levels (**Figure 8**, middle and bottom). Since only one electrode can be stimulated at any instant of time, the rate of stimulation is limited by the required stimulus pulse widths (determined individually for each subject) and the time to transmit additional stimulus parameters. As the information about the electrode number, stimulation mode, and pulse amplitude and width is encoded by high frequency bursts (2.5 MHz) of different durations, the total transmission time for a specific stimulus depends on all of these parameters. This transmission time can

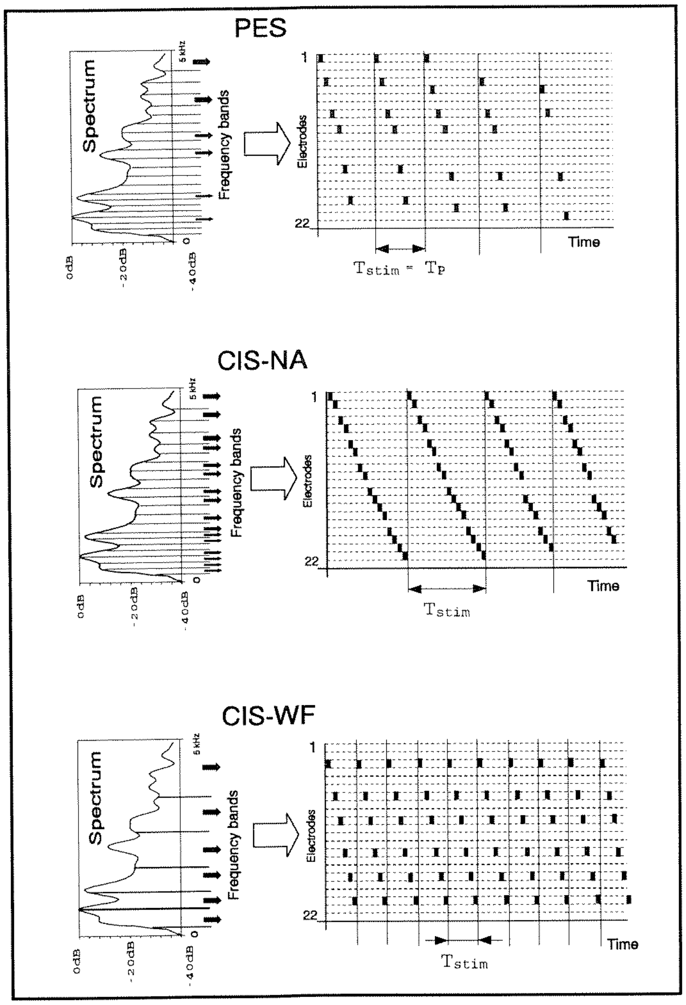

Figure 8.
Three CI-DSP strategies: pitch excited sampler (PES), continuous interleaved sampler with narrow band analysis (CIS-NA), continuous interleaved sampler with wide band analysis and fixed tonotopic mapping (CIS-WF).

be minimized by choosing the shortest possible pulse width combined with the maximal amplitude.

In order to achieve the highest stimulation rates for those portions of the speech input signals which are assumed to be most important for intelligibility, several modifications of the basic CIS strategy were designed, of which only the two most promising (CIS-NA and CIS-WF) will be considered in the following. The analysis of the short time spectra was performed either for a large number of narrow frequency bands (corresponding directly to the number of available electrodes) or for a small number (typically six) of wide frequency bands analogous to the approach suggested by Wilson, et al. (28). The frequency bands were logarithmically spaced from 200 to 5,000 Hz in both cases. Spectral energy within any of these frequency bands was mapped to stimulus amplitude at a selected electrode as follows: all narrow band analysis channels whose values exceeded a noise cut level (NCL) were used for CIS-NA, whereas all wide band analysis channels irrespective of NCL were mapped to preselected fixed electrodes for CIS-WF. Both schemes are supposed to minimize electrode interactions by preserving maximal spatial distances between subsequently stimulated electrodes. The first scheme (CIS-NA) emphasizes spectral resolution while the second (CIS-WF) optimizes fine temporal resolution. In both the PES and the CIS strategies, a high-frequency preemphasis was applied whenever a spectral gravity measure exceeded a preset threshold.

Subjects

Evaluation experiments have been conducted with five postlingually deaf adult (ages 26–50 years) cochlear implant users to date. All subjects were experienced users of their speech processors. The time since implantation ranged from 5 months (KW) to nearly 10 years (UT, single-channel extracochlear implantation in 1980, reimplanted after device failure in 1987) with good sentence identification (80–95 percent correct responses) and number recognition (40–95 percent correct responses) performance and minor open speech discrimination in monosyllabic word tests (5–20 percent correct responses, all tests presented via computer, hearing-alone) and limited use of the telephone. One subject (UT) still used the old wearable speech processor (WSP) which extracts only the first and second formant and thus stimulates only two electrodes per pitch period. The other four subjects used the new Nucleus Miniature Speech Processor (MSP) with the so-called multipeak strategy whereby, in addition to first and second formant information, two or three fixed electrodes may be stimulated to convey information contained in two or three higher frequency bands.

The same measurement procedure to determine thresholds of hearing (T-levels) and comfortable listening (C-levels) used for fitting the WSP or MSP was also used for the CIDSP strategies. Only minimal exposure to the new processing strategies was possible due to time restrictions. After about 5 to 10 minutes of listening to ongoing speech, 1 or 2 blocks of a 20-item 2-digit number test were carried out. There was no feedback given during the test trials. All test items were presented by a second computer which also recorded the responses of the subjects entered via touch-screen terminal (for multiple-choice tests) or keyboard (numbers tests and monosyllable word tests). Speech signals were either presented via loudspeaker in a sound-treated room (when patients were tested with their WSPs) or processed by the CIDSP in real time and fed directly to the transmitting coil at the subject's head. Different speakers were used for the ongoing speech, the numbers test, and the actual speech tests, respectively.

Results and Discussion

Results of 12-consonant (/aCa/) and 8-vowel (/dV/) identification tests are shown in **Figure 9**. The average scores for consonant tests **(Figure 9a)** with the subject's own wearable speech processor were significantly lower than with the new CIDSP strategies. The pitch-synchronous coding (PES) resulted in worse performance compared with the coding without explicit pitch extraction (CIS-NA and CIS-WF). Vowel identification scores **(Figure 9b)**, on the other hand, were not improved by modifications of the signal processing strategy.

The results for subject HS indicated that PES may provide better vowel identification for some subjects, while CIS was better for consonant identification for all tested subjects. It is possible that CIS is able to present the time-varying spectral information associated with the consonants better than can PES. Results from a male-female speaker identification test indicated that no speaker distinctions could be made using CIS, unlike PES which yielded very good speaker identification scores. This suggested that voice pitch information was well transmitted with PES but not at all with CIS.

Hybrids of the two strategies listed above were, therefore, developed with the aim of combining the respective abilities of PES and CIS in transmitting vowel and consonant information, as well as retaining PES' ability to transmit voice pitch information with the resultant hybrids. In one hybrid, the stimulation was switched between PES and CIS respectively, depending on whether the input speech signal was voiced or not. In another hybrid, for voiced portions of the speech, the lowest frequency active electrode was stimulated using PES while the remaining active electrodes were, at the same time, stimulated using CIS. For unvoiced portions, all

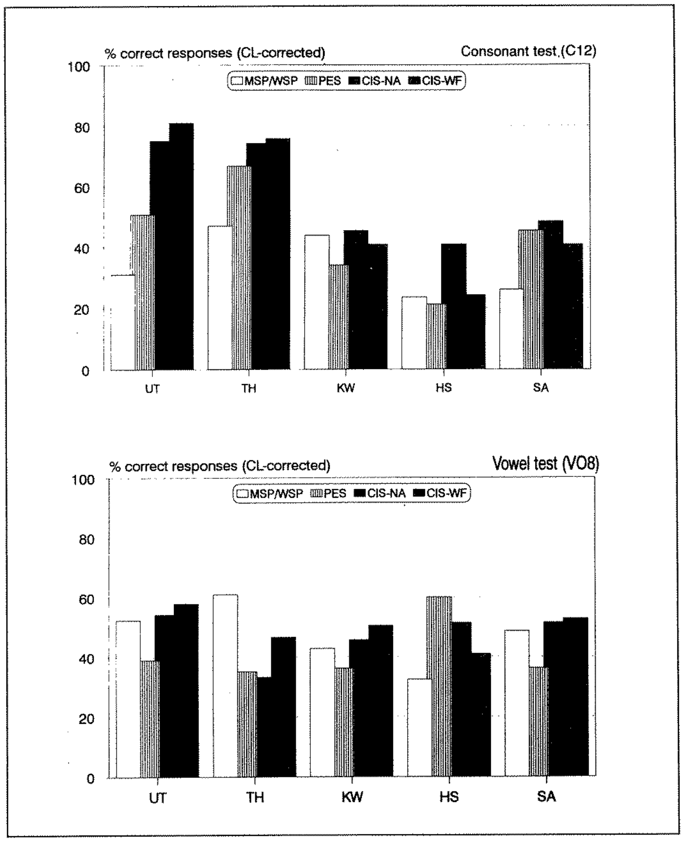

Figure 9.
Speech test results with five implantees, four processing conditions. a) Consonant test results. b) Vowel test results.

active electrodes used CIS stimulation. A third hybrid is a variation on the one above, using PES on the two lowest frequency active electrodes instead of only one.

Figure 10 shows the differences in electrode activation patterns between the Nucleus-MSP **(Figure 10a)** and the PES/CIS hybrid strategy **(Figure 10b)**. The graphs were generated from direct measurements of the encoded high frequency transmission signals emitted by the induction coil of the speech processor. Using the subject's programming map, which determines the relationships between stimulus amplitude and perceived loudness, relative stimulus levels (ranging from 0 to 100 percent) were obtained which are mapped into a color (or black and white) scale for the purpose of illustration. The axes of the graphs have been arranged similar to a spectrogram, with time on the abscissa and frequency (tonotopical order of active electrodes) on the ordinate. A number of striking differences between the two coding schemes became apparent

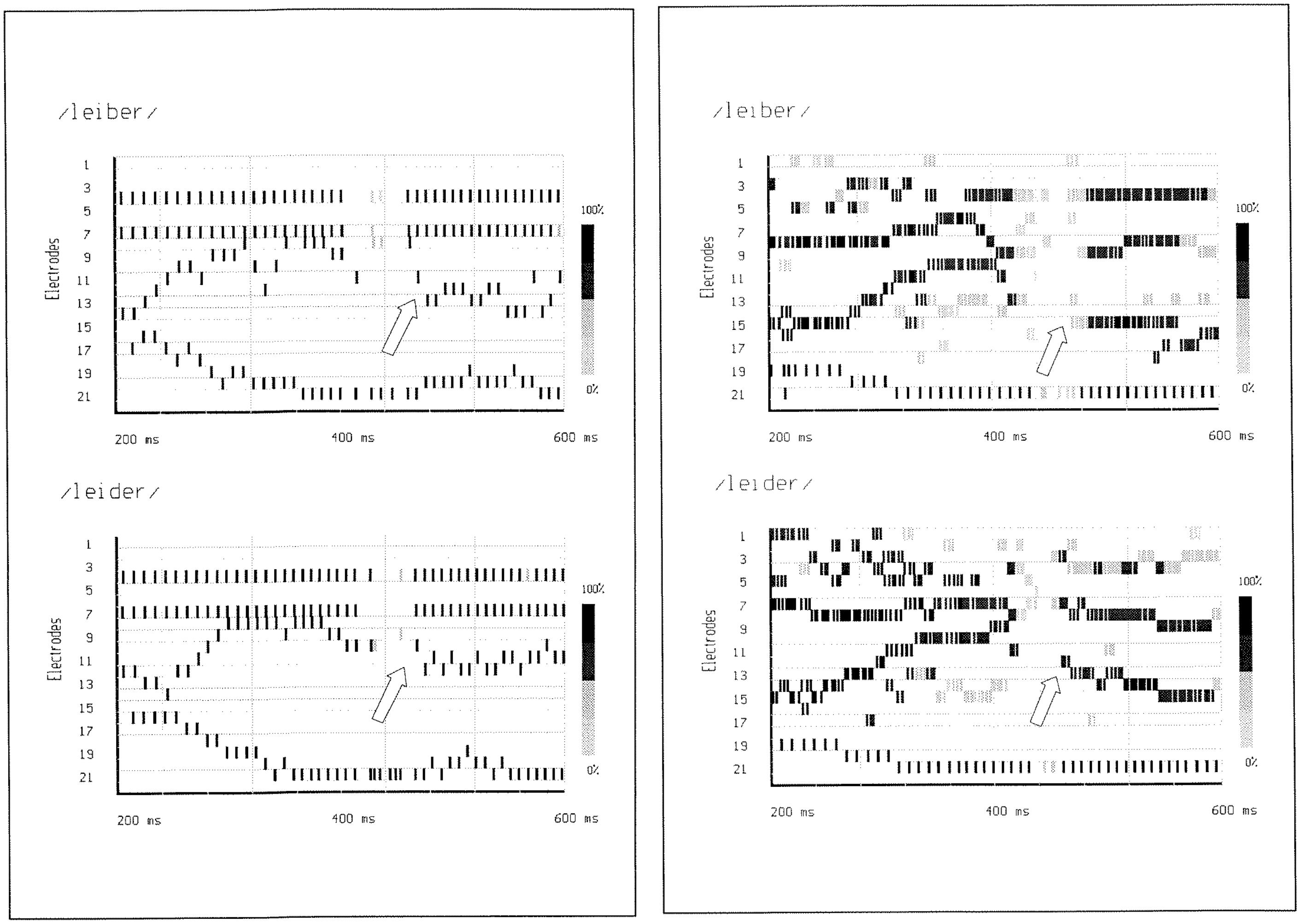

Figure 10.
"Electrodograms" of the phonological consonant minimal pair /leiber - leider/. a) Processed by the Nucleus-MSP. b) Processed by a new hybrid processing strategy where the lowest electrode is excited pitch-synchronously (PES-mode) and the higher electrodes at maximal rate (CIS-mode). The start of the voiced plosive sounds at about 450 ms is indicated by arrows.

from these displays. The first and second formant trajectories can be clearly seen in the MSP examples. Due to the smaller number of electrodes assigned to the first formant region in the case of the hybrid strategy, the variations in first formant frequency are less apparent and cover only three electrodes compared with six electrodes with the MSP. The rate of stimulation, however, is virtually identical for the lowest trajectory in both cases. The two highest stimulus trajectories in the MSP are assigned to fixed electrode numbers (4 and 7), whereas in the hybrid strategy, there is distinctly more variation visible in those areas than would be expected. The most striking difference between the MSP and the hybrid strategies, however, is the high-rate CIS portion with rather large variations in stimulus levels across electrodes and over time, as opposed to the more uniform and scarce stimulus pattern for the MSP. It can be seen that a distinction between the voiced plosive phonemes /b/ and /d/ is virtually impossible for the MSP, whereas with the hybrid (or CIS) the second formant transition between about 460 and 490 msec becomes visible and might be utilized by the implantees.

Preliminary experiments with these hybrid coding strategies have indeed shown some improvement in consonant identification. The results suggested, however, that the presence of voicing within time-varying portions could interfere with the perception of the CIS encoded information. Implantees were able to perform male-female speaker identifications quite well with all hybrids, indicating that the voice pitch information encoded into the PES stimulation within the hybrids can be perceived and used. This implies that by activating one or two low frequency electrodes using PES, it is possible to incorporate voice pitch information into an essentially CIS-like coding strategy.

Conclusions

The above speech test results are still preliminary due to the small number of subjects and test conditions. It is, however, very promising that new signal processing strategies can improve speech discrimination considerably during acute laboratory experiments. Consonant identification apparently may be enhanced by more detailed temporal information and specific speech feature transformations. Whether these improvements will pertain in the presence of interfering noise remains to be verified.

The experiments using hybrid coding strategies combining PES and CIS stimulation indicate that there is even more potential for improvement of the transmission of information regarding particular consonants. Further studies will have to be carried out to investigate in greater detail the perceptual interactions that arise between PES and CIS stimulation within the resultant stimuli. The results also show that voice pitch information, added to a CIS-like coding strategy by encoding it into one or two low frequency electrode(s), can be perceived and used by a cochlear implant user. Further optimization of these processing strategies preferably should be based on more specific data about loudness growth functions for individual electrodes or additional psychophysical measurements.

Although many aspects of speech encoding can be efficiently studied using a laboratory digital signal processor, it would be desirable to allow subjects more time for adjustment to a new coding strategy. Several days or weeks of habituation are sometimes required until a new mapping can be fully exploited. Thus, for scientific as well as practical purposes, the further miniaturization of wearable DSPs will be of great importance.

SUMMARY

The application of digital signal processors for the hearing impaired was demonstrated with three examples:

- Optimal parameters for a two-microphone adaptive beamformer noise reduction algorithm were determined by simulations and measurements for realistic environments. The experimental evaluation was carried out with normal-hearing and hearing-impaired subjects using a real time implementation on a floating point DSP. Results indicated the potential benefit of this method even in moderately reverberant rooms when a speech detection algorithm is included and an adequate filter length for the least mean squares (LMS) algorithm is used.
- A digital hearing aid with a multiband frequency domain modification of the short-time amplitude spectrum was fitted using loudness scaling measurements and evaluated with 13 hard-of-hearing subjects. The measurements that used narrow band signals were converted into gain factors via a simplified loudness estimation model based on the calculation of the spectral gravity point and the total energy within a short segment of the input signal. Significant improvements in speech recognition scores were found in quiet as well as in noisy conditions.
- A variety of high-rate stimulation algorithms were implemented on an experimental digital signal processor for the Nucleus multielectrode cochlear implant and compared with the performance of the patient's own wearable and miniature speech processor. It was found that consonant identification could be significantly improved with the new strategies in contrast to vowel identification, which remained essentially unchanged. Degradation in perceived sound quality due to loss of voice pitch information may be compensated by hybrid strategies.

ACKNOWLEDGMENTS

This study was supported by the Swiss National Research Foundation (Grant Nos. 4018-10864 and 4018-10865), by Ascom-Research (Hombrechtikon), and

Cochlear AG (Basel). We are also very grateful to all the subjects who have participated in sometimes boring and tiring experiments.

REFERENCES

1. Working Group on Communication Aids for the Hearing-Impaired. Speech-perception aids for hearing-impaired people: current status and needed research. J Acoust Soc Am 1991;90:637-86.
2. Boll SF. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process 1979;27:113-20.
3. Graupe D, Grosspietsch JK, Basseas SP. A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation. J Rehabil Res Dev 1987;24(4):119-26.
4. Lim JS, Oppenheim AV. Enhancement and bandwidth compression of noisy speech. Proc IEEE 1979;67:1586-1604.
5. Van Tasell DJ, Larsen SY, Fabry DA. Effects of an adaptive filter hearing aid on speech recognition in noise by hearing-impaired subjects. Ear Hear 1988;9(1):15-21.
6. Vary P. Noise suppression by spectral magnitude estimation—mechanism and theoretical limits. Signal Process 1985;8:387-400.
7. Griffiths LJ. An adaptive lattice structure for noise-cancelling applications. IEEE ICASSP Proc 1978:87-90.
8. Peterson PM, Durlach NI, Rabinowitz WM, Zurek PM. Multimicrophone adaptive beamforming for interference reduction in hearing aids. J Rehabil Res Dev 1987; 24(4):103-10.
9. Chabries DM, Christiansen RW, Brey RH, Robinette MS, Harris RW. Application of adaptive digital signal processing to speech enhancement for the hearing impaired. J Rehabil Res Dev 1987;24(4):65-74.
10. Van Compernolle D. Hearing aids using binaural processing principles. Acta Otolaryngol (Stockh) 1990;Suppl 469:76-84.
11. Greenberg JE, Zurek PM. Evaluation of an adaptive beamforming method for hearing aids. J Acoust Soc Am 1992;91(3):1662-76.
12. Kompis M, Dillier N. Noise reduction for hearing aids: evaluation of the adaptive beamformer approach. Proc IEEE EMBS 1991;13(4):1887-8.
13. Widrow B, McCool JM, Johnson CR. Stationary and nonstationary learning characteristics of the LMS adaptive filter. Proc IEEE 1976;64:1151-62.
14. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 1955;27:338-52.
15. Barfod J. Multichannel compression hearing aids: experiments and consideration on clinical applicability. In: Ludvigsen C, Barfod J, editors. Sensorineural hearing impairment and hearing aids. Scand Audiol 1978;Suppl 6:316-40.
16. Levitt H. Digital hearing aids: a tutorial review. J Rehabil Res Dev 1987;24(4):7-20.
17. Bustamante DK, Braida LD. Principal-component amplitude compression for the hearing impaired. J Acoust Soc Am 1987;82:1227-42.
18. Dillier N. Programmable master hearing aid with adaptive noise reduction using a TMS320-20. IEEE ICASSP Proc 1988:2508-11.
19. Levitt H, Neuman AC. Evaluation of orthogonal polynomial compression. J Acoust Soc Am 1991;90(1):241-52.
20. Lippmann RP, Braida LD, Durlach NI. Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. J Acoust Soc Am 1981;69:524-34.
21. Staab WJ. Review article: digital/programmable hearing aids—an eye towards the future. Brit J Audiol 1990;24:243-56.
22. van Dijkuizen JN, Festen JM, Plomp R. The effect of frequency-selective attenuation on the speech-reception threshold of sentences in conditions of low-frequency noise. J Acoust Soc Am 1991;90(2):885-94.
23. Villchur E. The rationale for multi-channel amplitude compression combined with post-compression equalization. Scand Audiol 1978;Suppl 6:163-77.
24. Levitt H, Neuman A, Mills R, Schwander T. A digital master hearing aid. J Rehabil Res Dev 1986;23(1):79-87.
25. Neuman AC, Levitt H, Mills R, Schwander T. An evaluation of three adaptive hearing aid selection strategies. J Acoust Soc Am 1987;82:1967-76.
26. Evans EF. How to provide speech through an implant device. Dimensions of the problem: an overview. In: Schindler RA, Merzenich MM, editors. Cochlear implants. New York: Raven Press, 1985: 167-83.
27. Clark GM, Tong YC, Patrick JF. Cochlear prostheses. New York: Churchill Livingstone, 1990: 1-264.
28. Wilson BS, Lawson DT, Finley CC, Wolford RD. Coding strategies for multichannel cochlear prostheses. Am J Otol 1991;12(Suppl 1):55-60.
29. Skinner MW, Holden LK, Holden TA, Dowell RC, et al. Performance of postlingually deaf adults with the wearable speech processor (WSP III) and mini speech processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant. Ear Hear 1991;12(1):3-22.
30. Dillier N, Senn C, Schlatter T, Stöckli M. Wearable digital speech processor for cochlear implants using a TMS320C25. Acta Otolaryngol (Stockh) 1990;Suppl 469:120-7.
31. Bögli H, Dillier N. Digital speech processor for the Nucleus 22-channel cochlear implant. Proc IEEE EMBS 1991;13(4):1901-2.