Feature extracting hearing aids for the profoundly deaf using a neural network implemented on a TMS320C51 digital signal processor.

J R Walliker *+, J Daley *+, A Faulkner • & I Howard*

*Dept. of Phonetics & Linguistics, University College London,
London, NW1 2HE, UK
Phone +44 71 380 7401
FAX +44 71 383 0752

+Dept. of Clinical Physics & Bioengineering, Guys Hospital,
London SE1 9RT, UK
Phone +44 71 955 4212
FAX +44 71 955 4532

Many people with profound hearing impairment, while able to detect amplified sound are often unable to make sense of what they hear. Conventional hearing aids which amplify, filter and compress the speech signal are of little use to them. It has been demonstrated that some profoundly deaf listeners are able to make better use of speech features such as voice fundamental frequency (Fx) and frication when they are presented in a simplified form matched to their residual hearing than when conventionally presented.

An algorithm is required for this purpose which can extract Fx in real-time, on battery powered portable hardware. The method should preferably operate on a period-by-period basis both to minimise the processing delay and because there may be useful information in the period-by-period irregularities of a speakers Fx. For the technique to be useful, the extraction method must perform better in a noisy and reverberant environment than the conventionally aided listener.

A number of neural network based algorithms were investigated which involved pre-processing the speech pressure waveform with a wide-band filterbank. The output from each filter band is fed into a digital delay line, storing at least 20ms of data. The taps of the delay lines form the input to a pattern classifier based on a multi-layer perceptron (mlp). The fully connected mlp, which consists of two hidden layers with one output unit is trained by a modified back-propagation method using simultaneous recordings of speech and Laryngograph data. The Laryngograph provides a target corresponding to the vocal fold closures which are the primary excitation to the vocal tract in voiced speech. For the first real-time implementation an algorithm was selected which uses most of the processing power of a TMS320C25 processor. This has 6 filter bands, followed by down-sampling to 2kHz before the data enters the delay lines. Six units are used in each hidden layer, and the sigmoid output function of each unit in the network is generated by table lookup.

The algorithms were initially implemented in assembler on a Loughborough Sound Images TMS320C25 development board, then ported to a Texas Instruments TMS320C50 software development system board before downloading to the portable hardware.

The algorithm has been tested both with digitised speech recordings on a minicomputer and in real-time using the LSI development board. Comparisons were made against several established Fx estimation algorithms including a time domain peak-picker, cepstral analysis and autocorrelation analysis. Its performance in competing noise was found to be better than that obtainable by peak-picking and as good as or better than the established techniques.

Aided lipreading for speech in noise was assessed in 5 profoundly hearing impaired listeners to compare the benefits of conventional hearing aids with those of the speech pattern element aid. An early implementation running on a development board was used which provides mlp based Fx information together with speech+noise amplitude information. The mlp based speech pattern element aid gave significantly better reception of consonantal voicing contrasts from speech in pink noise compared with that achieved with conventional amplification and led to better overall performance in audio visual consonant identification. In these listeners the direct presentation of noise not only obscured the speech information but also led to discomfort. The signal based on response to noise of the mlp based Fx extractor was regarded as much less distracting than the noise itself by all 5 listeners.

The shirt-pocket sized portable speech processor is based on a TMS320C51 processor. This was chosen for several reasons, notably its high speed (50ns cycle time), improved power-down modes and enhanced architecture and instruction set. The system is built on two 6 layer printed wiring boards. One holds the processor and memory while the other holds a microphone amplifier, 8kHz codec, and earphone power amplifier. The Analog Devices AD28MSP02 16 bit linear sigma delta codec integrates anti-aliasing and reconstruction filters with switchable high-pass filtering and a serial dsp interface. Two 128k x 8 bit 120ns Intel 28F001BX-B CMOS flash eproms store program code (which is downloaded to on-chip ram for execution) and lookup tables. These devices were selected because of their protected boot area, reasonably high speed (120ns) and low power consumption. They can be reprogrammed to suit the requirements of each individual patient in the audiology clinic using an interface card in a PC compatible computer. Two 8k x 8 bit 25ns CMOS static rams are used for fast access to those mlp weights which cannot be fitted into the 2k words of on-chip ram. They will be unnecessary when the 320C50 processor becomes available, as this has 10k words of ram. Four Sanyo 1.2Ah NiCd rechargeable cells power the system.

The main processing steps are: analog-to-digital-conversion (sampling rate 10 kHz, after analog low-pass filtering at 5 kHz), preemphasis and windowing (Hanning window of 12.8 ms, shifted by 6.4 ms or less per analysis cycle), calculation of the power spectrum, feature extraction and transformation according to selected encoding strategy, generation of stimulus parameters (electrode position, stimulation mode, pulse amplitude and duration), transmission via inductive coupling to the implanted receiver.

Two processing strategies have been implemented on this system: The first approach (PES, Pitch Excited Sampler) is based on the classical channel vocoder concept whereby the time-averaged spectral energy of a number of logarithmically spaced frequency bands is transformed into appropriate electrical stimulation parameters for up to 22 electrodes. The pulse rate at any electrode is controlled by the voice pitch of the input speech signal. The 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 window. A random pulse rate of about 150 to 250 Hz is used for unvoiced speech portions. The second approach (CIS, continous Interleaved Sampler) uses a stimulation pulse rate which is independent of the input signal. The algorithm scans continuously all frequency bands and samples their energy levels. As only one electrode can be stimulated at any instance of time the rate of stimulation is limited by the required stimulus pulse widths (as determined individually for each subject) and some additional constraints and parameters.

The temporal sequence of stimulated electrodes is probably an important parameter for high-rate stimulation algorithms. A number of different rules have therefore been implemented which specify minimal temporal and spatial distances. Other modifications and extensions of the basic PES- and CIS-strategies include enhancement of speech feature contrasts such as peak-to-valley relations for vowel formants and spectral gravity in high-frequency consonants.

Evaluation experiments have been conducted with 5 patients until now. Performance in consonant identification tests was significantly better with the new processing strategies than with the patients own wearable speech processors whereas improvements in vowel identification tasks were less obvious. Several modifications of the basic PES- and CIS-strategies were tested resulting in large variations of identification scores. Information transmission analysis of confusion matrices revealed a rather complex pattern across conditions and speech features. Optimization and fine-tuning of processing parameters for these coding strategies requires more data both from speech identification and discrimination as well as psychophysical experiments.

A similar experimental DSP-System has been used to implement a multiband loudness correction (MLC) algorithm for a digital hearing aid. Magnitude estimation procedures are used to determine loudness growth functions in 5 frequency bands. Gain tables are calculated and downloaded into the DSP-memory. Windowed blocks of input signals are transformed to the frequency domain via FFT, the amplitude spectra are loudness-corrected according to estimated iso-loudness contours and the modified spectra inversely transformed and overlap-added. Speech tests in quiet and noise with 5 users of conventional hearing aids resulted in significant improvements of discrimination scores with the MLC-algorithm.