# HEARING AIDS FOR THE PROFOUNDLY DEAF BASED ON NEURAL NET SPEECH PROCESSING

**Manfred Leisenberg**
University of Southampton
Institute for Sound and Vibration Research
Southampton, UK - SO9 5NH, England
Email:LEISEN@MAIL.SOTON.AC.UK

## ABSTRACT

A new speech processing concept for Cochlear Implant (CI) - systems has been developed. It is based on robust feature extraction and a neural net classifier: Feature coefficients, extracted either by relative spectral perceptual linear predictive technique or regular CI-filtering, are classified into 'auditory related units '. The classifier is based on an adapted self-organizing Kohonen algorithm which finds representative clusters in the input feature vector space. These clusters are closely related to the statistical distribution of the feature coefficients and represent phonetic units. Firing neural net output nodes control the synthesis of a limited 'stimulus pattern alphabet'. Each 'letter' represents a sub-phoneme and is linked to a highly distinguishable complex stimulus pattern. The concept has been implemented with CINSTIM V2.0. First experimental results confirm the new CI speech processing strategy.

## 1. Introduction

Consider the problem of the huge amount of information a cochlear implant (CI) patient has to process shortly after being implanted with one of the common CI- systems. Wouldn't it be a good idea to supply the patient first with a limited 'alphabet' of stimulus patterns and increase the number of 'letters' step by step during the rehabilitation process until he will be able to recognize the full continuous information stream sufficiently?

CI- systems are based on the principle of coding acoustical information into electrical stimulus patterns. Fig.1 illustrates the principle of such a system: First, the acoustical input signal is preprocessed. Digitised input speech is analysed by an appropriate method (filtering, FFT). Then, features like formant frequencies and related amplitudes are extracted. Finally, frame by frame the resulting data are encoded into stimulus parameters in order to produce stimulus patterns which are consisting of

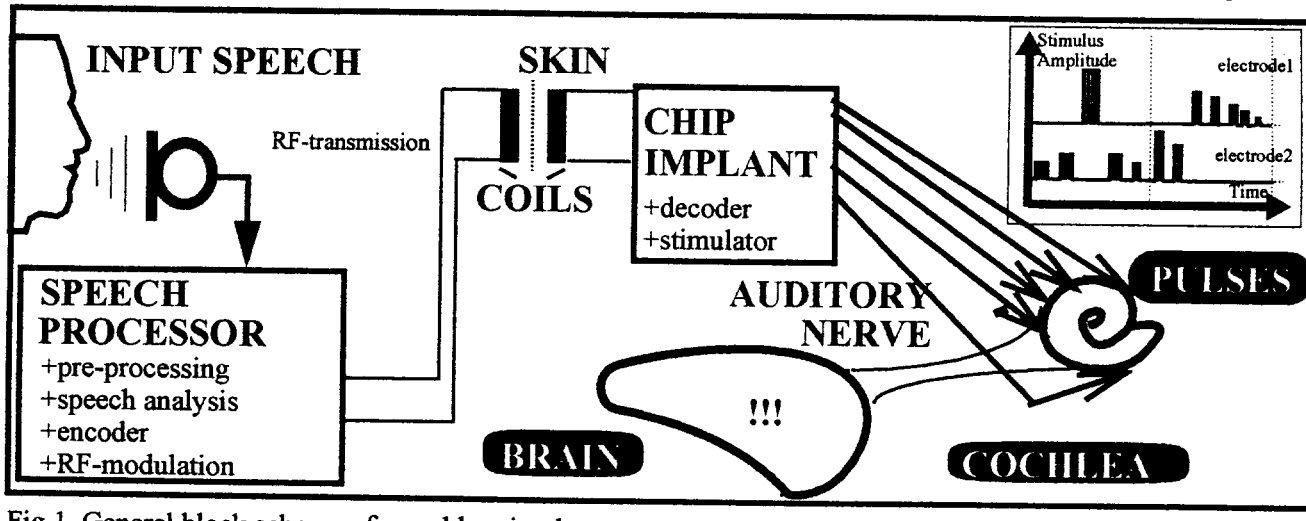


Fig.1. General block scheme of a cochlear implant system



0-7803-2431-5/95 $4.00 

bursts of pulses. The parameter data are transferred to the implant transcutanieously via RF- modulation. Finally, the mentioned pulses stimulate the auditory nerve of the deaf patient in order to provide him some sort of sound impression.

Existing commercial systems are providing patterns nearly unlimited in their variety. The same word spoken twice by one person will never produce exactly the same pattern sequences. The influence of different speakers or inadequate acoustical conditions produces often an unrecognizable stream of information at the auditory nerve. Especially for the initial rehabilitation period and for prelingually deafened patients it seems to be promising to limit the information to be provided to a relatively small amount of different stimulation patterns. A limited number of such patterns, the 'alphabet', and their sense might be learnt by the hearing impaired quickly and easily. For this purpose we have developed a new CI- speech processing concept. It is based on neural net classification [1] and can be combined with the known CI-processing methods. This paper describes the technical idea behind our concept, implementation advances and first experimental results.

## 2. Neural net CI speech processing

The concept of neural net CI speech processing strategy is based on following approaches:

- Input sound (speech) is transformed into sequences of discrete, distinguishable stimulus patterns.

- Each pattern represents a group of similar acoustical parameters (e.g. it represents a phoneme)

- The number of different patterns is limited and may be increased with the success of the rehabilitation process.

- Optional, sound may be processed independent of characteristics of different speakers and robust against noise.

Fig.2. illustrates CI speech processing when incorporating the above concept: After pre-processing, input speech is processed by feature extraction. Each frame of speech is represented by a m-dimensional feature vector $\chi_i$,

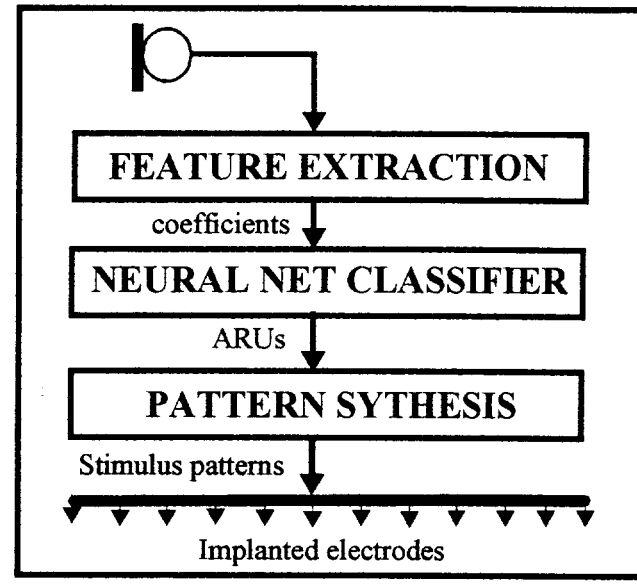


Fig.2. Block scheme of neural net CI processing

$\chi_i=\{X1_{1i},X_{2i},...,X_{k,i},...,X_{mi}\}$, i= 1,2,...n, $1\le k\le m$.

For feature extraction we currently employ either RASTA-PLP [6] or a regular CI filtering algorithm[7]. RASTA-PLP provides robust, speaker independent feature coefficients.

The classifier assigns each $\chi_i$ to so called 'auditory related units (ARUs)'. It is based on a neural net algorithm which is able to learn properties of the relation between acoustical input features and the ARUs.Artificial neural nets are either working supervised or unsupervised. In case of supervised neural networks during the learning process the connectionist system has to know which nodes of the output layer have to be activated related to a specific input pattern. In terms of our classification problem we would need to know the targets in order to classify the features $\chi_{k,i}$. We could find such targets empirically or by employing a auditory nerve stimulation model. Our investigations [2] show that the supervised method is not very effective to process speech in CI- systems. In particular, there is no effective model available to derive the appropriate training data.

Unsupervised neural nets do not need any external adjustment to determine the desired input/output transformation. Because of it's special property of effectively creating spatially organized 'internal representations' of various speech features we choose

the unsupervised Kohonen feature map [3]. During the training process Kohonen's algorithm finds clusters in the input feature vector space. These clusters are closely related to the statistical distribution of the input feature vectors and resemble very closely the topographically organized maps found in the cortices of more highly developed animal brains[4].In addition to that output nodes representing such clusters can be interpreted as phonetic units in tonotopical maps [5]. Depending on the frame length $t_F$ the ARUs represent words, phonemes or sub-phonemes (our experiments: $t_F$ = 40ms -> sub-phonemes).

Fig.3 shows the topological structure of the Kohonen map. During the training process sequences of feature vectors $\mathcal{Z}_i$ are presented to the map. For explanation, the feature vectors $\mathcal{Z}_i$ are 4- dimensional;

$$\mathcal{Z}_i=\{F1_i,F2_i,B1_i,B2_i\} i= 1,2,...,k,...n.$$

Components of $\mathcal{Z}_i$ represent extracted speech features with $F1_i$,$F2_i$ peak frequencies of the first and second formant, $B1_i$,$B2_i$ related bandwidths; i denotes the frames number. The output layer consists of a 2-dimensional array. Every input node is connected to

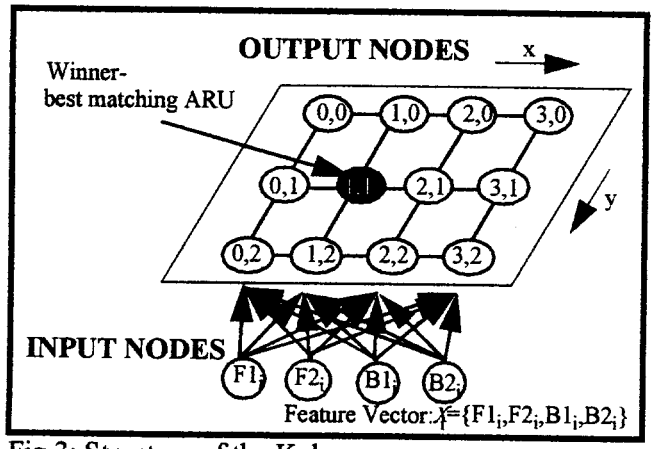

Fig.3: Structure of the Kohonen map

every output node via the weighted links. During the training process the weights $w_{xyf}$ are adapted to the input vectors. At the end of the learning process the density function of the weights closely represents the probability density function of the input vectors $\mathcal{Z}_i$. In other words: each weight vectors $\omega_{xy}$ represents a cluster found in the feature vector space. After finishinging training output nodes are labelled. A label consists of a pointer to the analytical description of a complex stimulus pattern and denotes a 'letter' of the above mentioned 'stimulus pattern alphabet'. The number of 'letters' corresponds to number of output nodes. Sub-alphabets can be derived from maps with less output nodes if maps are trained in a hierarchical order [5].

We developed two labelling strategies:1.- Labels are assigned to the ARUs arbitrarily. That creates completely artificial 'sounds'. This method is only applicable to preligually deafened patients. 2.- A regular CI system runs in parallel and produces reference patterns for comparision. With this methods we achieve more natural 'sounds'.

The resulting "alphabet" will be used to encode the acoustical input information during the following test steps.When testing the map, the minimal Euclidean distance between the presented input feature vector $\mathcal{Z}_i$ and all nodes represented by their weight vectors $\omega_{x,y}$

$$\omega_{xy}=\{w_{xyF1},w_{x,y,F2},w_{x,y,B1},w_{x,y,B2}\}$$

denotes the best matching ARU. The mentioned Euclidean distance is measure of quantization error $Q_{err}$. Finally, for individual adaptation the map is fine tuned by employing Kohonen's 'learning vector quantisation'[3].

The subsequent synthesizer generates stimulus patterns by processing the labelled ARUs. The psychophysical implications of this approach are discussed in [8].

## 3.IMPLEMENTATION AND EXPERIMENTS

### 3.1. CINSTIM implementation

The CINSTIM (Cochlear Implant Neural net Simulation and sTIMulation framework [9] is based on PC, a hierarchically organised graphical user interface and a COCHLEAR 22 MSP™ CI-system[7]. The objective is to provide a powerful, block oriented experimental tool including different feature extraction algorithms, artificial neural net (ANN) processing and access to the CI- system in order to validate the above concept and to conduct patient tests.Figure 4. illustrates CINSTIMs processing blocks:Input speech samples are provided by the TIDIGITS database. The feature extraction block implements the RASTA-PLP- algorithm [6] and, optional a MSP filtering algorithm [7].Feature coefficients are either directly sent to the stimulation kernel or to the neural net. In the first case the neural

net will be bypassed to run the stimulation for reference purposes. Otherwise, the feature coefficients are fed to the classifier. Each set of feature coefficients at the ANN input results in firing of one of the output nodes (ARUs). Subsequent dictionary is the first synthesis stage. It consists of a database with stimulus parameters of all 'letters' of the 'stimulus pattern alphabet'. ARUs are the database access keys, output are stimulus parameters. The definition and creation of the 'letters' is done by employing the off-line stimulation tool. Examples of different 'letter'- types are presented in [2]. The stimulation kernel controls the actual stimulation. The 'off- line stimulation tool' enables the user to manipulate off-line stimulations and to develop distinguishable 'letters' of the 'alphabet'. Individual stimulus patterns files can be edited, and reloaded to repeat a test, and to be converted into a unique dictionary entry.

### 3.2. Experimental results

Extensive experiments have been conducted in order to simulate the speech analysis/ classifier complex. Simulation results [2] validated the technical concept. After training with the TIDIGITS database quantization error remained below 5%. First promising experiments with deaf implanted patients were achieved with CINSTIM V2.0 [9]. These experiments concentrated on the definition and test of distinguishable complex stimulus patterns. We found [8] that, with training, patients are able to recognise up to 80% of such artificial patterns.

## 4.CONCLUSIONS

An alternative concept for speech processing in CI-systems has been proposed. Computer simulation results validate the feature extraction and neural net classifier principle. The new CI speech processing strategy has been implemented with CINSTIM V2.0.This software package provides a user friendly graphical user interface. The implementation is block oriented. That provides flexibility and the option of replacing or exchanging particular blocks for later modifications or extensions. With first successful and promising patient test we have demonstrated the possibility of applying artificial neural networks to cochlear implant speech processing. Further experiments will be conducted in order to find a common 'stimulus pattern alphabet' for a number of subjects and to derive sub-alphabets by employing hierarchically organised Kohonen maps.

## 5.REFERENCES


[1] Leisenberg, M., A concept of an adaptive, neural net based cochlear- implant- system using speaker independent signal representation, Proc. International Symposium on cochlear implants, Toulouse, Fr.1992

[2] Leisenberg,M.: Application of Artificial Neural Networks for a New Concept of Cochlear Implant Systems, in: Hochmair- Desoyer, I.J., Hochmair, E.S.: Advances in Cochlear Implants, International Interscience Seminars 7/93, Innsbruck, Austria

[3] Kohonen, T.: The self-organizing map, Proc. IEEE, 78(1990)9, p. 1464 ff.

[4] Tavan, P., Grubmüller, H., Kühnel, H., Self-organization of associative memory and pattern classification: Recurrent signal processing on topological feature maps, Biological Cybernetics, 64, p. 95-105, 1990

[5] Windheuser,C., Competitive Sequence Learning, Diplomarbeit Rheinische Friedrich- Wilhelms-Universität Bonn, Bonn 1991

[6] Hermansky,H., Morgan, N., RASTA processing of speech, IEEE Trans. on speech and audio processing, 2(1994)4, p.578 ff.

[7] MINI System 22 - Audiologist's Handbook, Cochlear Pty., Sydney, Australia1992

[8] Leisenberg,M. ,Southgate,J.: First results on patient experiments with CINSTIM - The Southampton Cochlear Implant/ Neural network STIMulation framework, Proc. International Cochlear Implant,Speech and Hearing Symp., October 1994, Melbourne, Australia

[9] Leisenberg,M. , Downes,M.: CINSTIM: The Southampton Cochlear Implant/ Neural network STIMulation framework - implementation advances of a new, neural net based speech processing concept,International Cochlear Implant,Speech and Hearing Symp.,October 1994, Melbourne, Australia