16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SPARSE STIMULI FOR COCHLEAR IMPLANTS

Guoping Li, Mark E. Lutman

Institute of Sound and Vibration Research
University of Southampton, SO17 1BJ, Southampton, UK
phone: +(44)2380592921, fax: +(44)2380593190, email: lgp@isvr.soton.ac.uk, mel@isvr.soton.ac.uk

ABSTRACT

The aim of the present research is to explore the application of sparse coding principles to the processing within a cochlear implant. These principles would determine what information in noisy speech should be extracted and used to excite the electrode array within the cochlea. The hypothesis is that reducing redundancy in the signal, making it more sparse, would improve speech recognition scores in noise.

The proposed sparse coding strategy was based on a combination of ICA (independent component analysis) and PCA (principal component analysis), both operating on the spectrotemporal envelope of the speech signal. The algorithm is tested for speech in quiet and modulated babble noise conditions (signal-to-noise ratios, SNR=15 dB, 10 dB, 5 dB). Results show that the algorithm is beneficial, particularly when baseline performance of listeners is poor. This approach is applicable both to acoustical hearing aids and cochlear implants.

1. INTRODUCTION

A cochlear implant (CI) is an electrical device that helps to restore partial hearing to the profoundly deaf. The main principle of cochlear implants is to use electrodes, inserted in the inner ear, stimulating the auditory nerves. Electrodes at different places correspond to different frequencies. Cochlear implants transfer acoustical information to the auditory perceptual system via electrical pulses representing modulation of the speech spectrum. Although the speech information sent through cochlear implants is quite crude, the performance of CI users has seen increases with the new speech processors and algorithms. A majority of implant users have benefited from this device. Many of them can converse using the telephone without difficulty. Some top CI users even get similar performance in quiet to normal hearing subjects using clinical speech recognition tests [1].

However, the average performance of most cochlear implant users still falls below normal hearing, especially some CI users perform much worse in a noisy environment. How to help the lower performance users is still a key research question in CI research. Normal hearing people can understand speech very well in a moderately noisy environment, but this is a very challenging situation for cochlear implant users to deal with. Normal hearing subjects are able to obtain masking release by exploring the characteristics of noise, such as single same talker or single different talker. CI users are unable to take advantage of these masking characteristics and no masking release has been observed when a different masking voice was used [2]. Most CI users, compared with normal hearing subjects, are not able to fully use the spectral information provided by their implant in noisy situations [3].

One idea to improve the speech recognition for the lower end performance users in noise is to minimise redundant stimuli, from which CI users will be unable to get any benefit. The lower performance users can be considered as having more severely limited resources to encode the electrical stimuli provided by CI. In noisy situations, the non-useable stimuli would compete for the encoding resources with the useful information for speech understanding. These stimuli in noisy condition could even harm the speech perception performance of CI users. CI users already have a very limited dynamic range [4] to encode the electrical stimuli. Giving stimuli that can not be used by the auditory system of CI users can be thought as a waste of encoding resources in general. There is clearly a bottleneck between electrical stimuli and acoustical information. This bottleneck is probably worse for CI users whose performance is poor.

In order to achieve higher speech recognition performance, a cochlear implant has to transfer most essential information into the limited dynamic range of impaired auditory neurons. The limited dynamic range problem is essentially an information transmission problem. The CI processor has to find a way to optimally transfer most relevant speech information to CI users. Looking for the most essential information for CI speech processing has been the key to research from the start of CI and has important implication for speech perception research. The other way to look at this problem is to consider information theory and quantify the problem with the corresponding mathematical tools.

Most recently sparse coding [5] based on information theory has been applied to the auditory filter modelling process and achieved great success [6]. Sparse coding theory states that only a few neurons fire at the same time, and the redundancy of data can be explored through higher order statistics [7, 8]. Cochlear implants cause neuron firing patterns through electrical stimuli. The firing patterns are very well synchronized with the electrical stimuli [9]. To implement sparse coding principles for electrically stimulated auditory neuron firing requires sparse stimuli: a sparse speech spectrum.

2. SPARSE ALGORITHM

In order to produce a sparse spectrum based on information theory, a combination of PCA and ICA is used. Fig: 1 shows the structure and key concept of the SPARSE algorithm.

Speech is first fed into a conventional Advanced Combination Encoder (ACE) speech processing strategy shown in the right panel to get a spectrum envelope. The envelopes in different frequency channels are estimated from each spectral band of the audio signal. ACE strategy then select N channels out of M frequency band to stimulate auditory neurons of CI users.

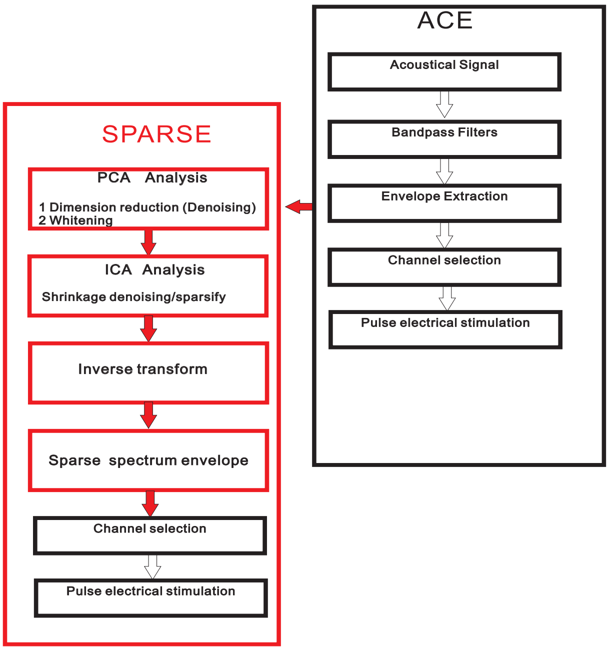

Figure 1: *Acoustical signal first is processed by filter banks and then envelopes are extracted. The conventional ACE algorithm will select a few channels out of 20 to 22 channels and stimulate the auditory neurons non-simultaneously. The SPARSE algorithm uses PCA and ICA analysis on the envelope signals, reducing the redundancy and making the spectrum envelope representation sparse.*

For the new SPARSE algorithm, the envelope signals across different frequency bands from ACE are then fed into PCA to reduce dimensions and some noise. The dimension reduction is especially important for an encoding system that has limited dynamic range [7]. The output of PCA after whitening (X) can be sent to ICA and transformed to the independent space by transformation W, where channels are independent based on Equ. 1. Thresholding can be performed on the independent channels. The thresholding fucntion in Equ. 2 is based on the assumption that the independent components S are non-gaussian and noise is more gaussian [10].

$$S = W * X \tag{1}$$

Where W is the ICA transformation matrix and X is the output after PCA. 12 out of 22 eigenvectors of the envelope signals are selected.

$$S\_Essential = \frac{sign(S. * max(0, |S| - A. * \sigma^2)}{(1 + \sigma^2 * B)} \tag{2}$$

Where $S\_Essential$ represents the essential components in the independent space S. S represents the independent components and σ is the standard deviation of the Gaussian noise. Here it can be replaced by the standard deviation of the independent components in order to reduce the independent components with smaller amplitude, assuming that the small absolute values are purely noise. A and B are constants: for modulated noise in the experiment, A=3, B=8. The selection of A and B is based on the trials of subjective listening to the

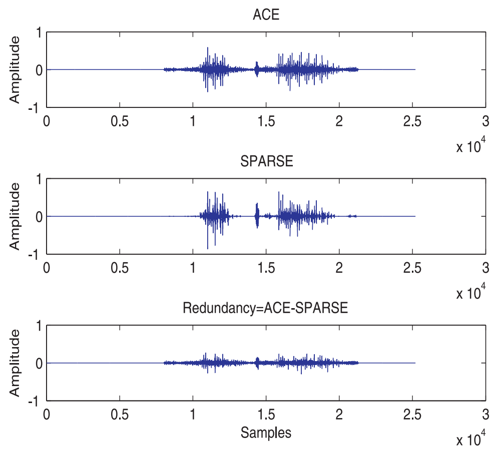

Figure 2: *The simulation of /ata/ from ACE output with SNR=10 dB. The middle panel shows the output of SPARSE. The lower part is the difference between ACE and SPARSE. The difference can be seen as redundant components hidden in the ACE output, as the SPARSE output can represent the /ata/ sound. It can be seen that the SPARSE output is more clean and sparse.*

vocoder of sparse envelope signals [12]. A and B could also be optimized for each individual users.

Many sparse denoising shrinkage algorithms derive W based on a clean set of data [10, 13]. Here W is derived based on the noisy data directly. The SPARSE algorithm applies ICA on noisy data for two reasons. One is that clean speech will not be available in the real time application; the other is that PCA can help to reduce part of noise through dimension reduction. The SPARSE algorithm is designed not only to reduce noise but also to reduce some speech components that are redundant. Inverse transformation can be performed on the selected independent components after thresholding.

$$\hat{X} = W^{-1} * S\_Essential \tag{3}$$

$\hat{X}$ then contains the essential parts of the spectrum envelope and is a sparse version of X, the speech spectrum output of ACE. $\hat{X}$ is used to generate the electrical pulses sequence for CI stimuli to produce a sparse pattern of auditory neuron firing.

Fig. 2 shows an example of the output of ACE and SPARSE algorithms with SNR = 10 dB. The SPARSE output is much more clean. The difference between ACE and SPARSE can be seen as the redundant parts hidden in the ACE output. The SPARSE output can be heard more clearly and sounds more crisp than the ACE output in the simulation.

3. SPARSE EVALUATION

One important factor considered in the proposed algorithm is the sparseness of the reconstructed signal. An important objective is to transform the CI stimuli to more sparse stimuli, which could potentially make the neurons fire sparsely and implement the sparse coding theory for CI. Sparseness can be quantified by kurtosis of the signal [7].

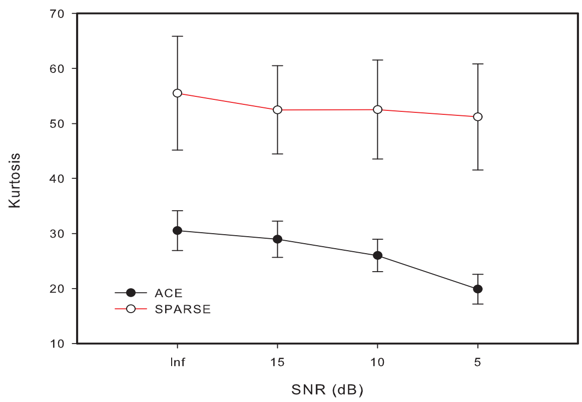

Figure 3: *The increased sparseness with SPARSE output. ACE and SPARSE algorithm outputs can be simulated as acoustical signals using a vocoder, combining envelopes from different frequency bands after being modulated by their corresponding narrow band frequency noises. The sparseness of these simulated waveforms can thus be measured by kurtosis. The output with the SPARSE algorithm is much more sparse than the ACE output. The objective of making the spectrum sparse is achieved, even for the clean speech (SNR=Inf).*

$$K = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu)^4}{\sigma^4} - 3 \tag{4}$$

where x is the amplitude of the signal, μ is the mean and σ is the standard deviation. For a normalized Gaussian (non-sparse) distribution with $\mu = 0$ and $\sigma = 1$, the kurtosis is (by definition) $K = 0$, for other signals the kurtosis may be super-Gaussian ($K > 0$) or sub-Gaussian ($K < 0$).

3.1 Speech materials

Speech tokens were drawn from 9 VCV (Vowel Consonant Vowel) words: (/aba/, /ada/, /aga/, /aka/, /ala/, /ama/, /ana/, /apa/, /ata/). Modulated noise was used in four noisy conditions (Quiet, SNR=15, 10, 5 dB). The modulated noise was also used in [11]. The number of talkers for modulated noise was eight.

The evaluation of sparseness takes the simulated output waveforms as a whole and calculates the kurtosis of the entire time series based on Equ. 4. If the kurtosis increases then the sparseness of the stimuli is increased. Fig. 3 shows that the kurtosis of the output of the SPARSE algorithm is much higher than the sparseness of the output of ACE algorithms. The kurtosis of the enhanced signal after the SPARSE algorithm is higher than that of the output of ACE. The enhanced signal is therefore more sparse than the output of the ACE algorithm.

4. LISTENING EXPERIMENTS

The increased kurtosis of the enhanced signals shows that the SPARSE algorithm can make the stimuli sparse. In order to test the intelligibility of the sparse speech, both for CI users and normal hearing listeners, listening experiments

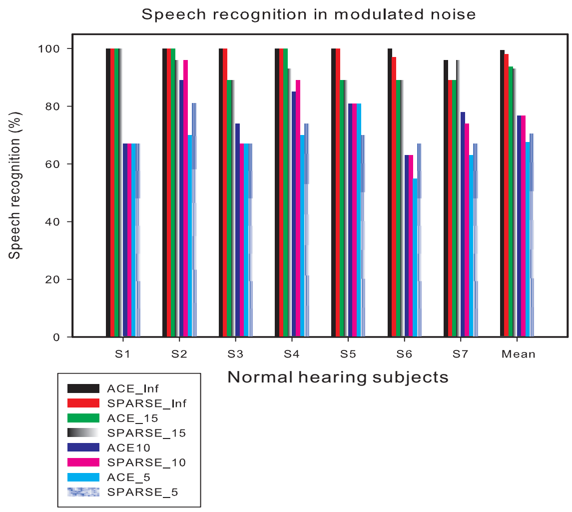

Figure 4: *Normal hearing subject recognition scores in modulated babble noise with ACE and SPARSE in four signal noise ratio conditions. 'ACE_Inf' represents the condition of quiet with ACE algorithms; 'sparse_15' represents the 15 dB SNR condition with SPARSE.*

were performed to test the algorithm. Both normal hearing (7 subjects) and CI users (10 subjects) participated in the listening experiments. The normal hearing subjects listen through TDH-39 earphones. The stimuli of CI users were saved in the computer and presented through Nucleus Implant Communicator software connected to their cochlear implant device (NIC streaming).

4.1 Experiment I: Normal hearing subjects and sparse stimuli

Normal hearing subjects can listen through vocoder simulation of the output of CI processing. The vocoder was produced based on the envelope signals of the ACE and the SPARSE strategy. The speech materials used are the same as the speech materials used in the kurtosis evaluation: 9 VCV words, babble modulated noise, four different noise conditions (Quiet, SNR=15, 10, 5 dB). Each item was presented four times to each subject and only the last three times were counted in the average score. All stimuli presented to the normal hearing listeners were set to equal loudness.

Fig. 4 shows the individual results for speech recognition scores in the modulated babble noise. Most subjects show improvement when the SNR is lower, especially at +5 dB SNR. The difference between SPARSE and ACE is statistically significant ($P = 0.021$) for $SNR = 5\ dB$ in the babble noise condition. The difference between SPARSE and ACE is not statistically significant for other conditions. There is possibly a ceiling effect for normal hearing subjects. Fig. 5 plots the differences between the scores of subjects with ACE and with SPARSE. It can be seen that the speech recognition score improves when baseline performance of subjects is poor, say when the speech recognition score is below 70%.

The results in normal hearing subject show that the SPARSE algorithm helps when SNR is low and speech recognition performance is lower with ACE. Fig. 5 shows that when the speech recognition score is below 70% in noise

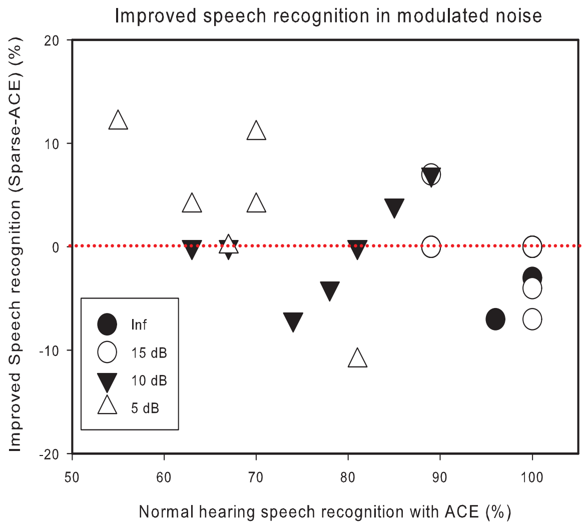

Figure 5: *Normal hearing subject recognition score in modulated noise with ACE and SPARSE in four SNR conditions. X-axis is the speech recognition score with ACE. Y-axis is the score difference between ACE and SPARSE in modulated babble noise. The improvement is smaller in the area of good performance. The dotted line represent a score difference of zero. For symbols under the dotted line, the recognition score decreases with SPARSE.*

conditions, the SPARSE algorithm can help improve performance. While for the listeners who have already have high performance, the SPARSE algorithm could make the performance worse.

One possible reason is that when the speech recognition score is high, the listeners are able to resolve most information presented and reach ceiling performance. Reducing redundant information may reduce speech recognition score. While for people who could not resolve the information in noisy conditions, the SPARSE algorithm can help to improve their performance by reducing the noise and redundant information.

Another possibility is simply the statistical artefact of regression to the mean, where a listener who scores low by chance on one test will tend by chance to show better performance on the other.

4.2 Experiment II: CI users and sparse stimuli

The same speech materials were used as for normal hearing subjects: nine VCV (Vowel Consonant Vowel) words: (/aba/, /ada/,/aga/,/aka/,/ala/, /ama/, /ana/, /apa/, /ata/) and modulated noise. Four noise conditions (Quiet, SNR=15, 10, 5 dB), four repeats per item are presented and only the last three times counted in the average score. The stimuli were presented to the subjects through the NIC stream software, which can deliver the electrical pulse sequence to the CI. The sequences were produced based on CI settings that subjects used daily. The sequences were saved in the computer before presentation.

Fig. 6 shows the individual results of speech recognition scores in the modulated babble noise. The variance among subjects in speech recognition performance is quite large. For example, subjects S4 and S6, have great improvement on speech recognition score using the SPARSE algorithm across different SNR; while for listeners whose performance is relatively high in quiet, the improvement is less or no improvement (S9 or S10 for example). This is consistent with the normal hearing listeners, but the improvement for CI is larger than for normal hearing listeners. This suggests that sparse stimuli are more important for CI users.

Fig. 7 plots the difference between the scores of subjects with ACE and SPARSE in modulated babble noise. The main improvement is again shown in the left upper corner as the algorithm helps listeners whose performance with ACE is poor. The general conclusion is that the SPARSE algorithm helps those whose baseline performance is low. This is interesting as the improvement for users with lower performance is a challenging task for CI research.

The SPARSE algorithm could help these subjects to achieve relatively higher speech recognition scores, which are comparable with the high end performance CI users. As shown in Fig. 6, subject S7 has a very low speech recognition performance with ACE at 5 dB SNR (19%), compared to the score of S6 in the same condition (44%). With SPARSE, the performance of S7 at 5 dB SNR can be as high as S7: 48%. S5, S7 and S8 had missing data points at 5 or 10 dB as the subjects reported tiredness and less data were collected. Possible reasons for the improvement of CI users could be: (1) the new strategy reduces interaction between channels; (2) it selects only the essential information in speech for simulating auditory neurons; (3) as only the most essential information is selected, the limited dynamic range of cochlear implant users can be used efficiently; (4) it might force neurons to fire more sparsely, and hence more physiologically, compared to neurons stimulated by the present commercial algorithms. Incidentally, SPARSE will reduce energy consumption and therefore prolong battery life.

Limitations on improvement were undoubtedly caused by lack of familiarity of users with the novel processing. CI users typically require days or weeks of familiarity to benefit from speech processor improvement. Although the results here were based on an off line SPARSE strategy, a real time SPARSE strategy is still being investigated. There are already some online algorithms which combine PCA and ICA for blind source separation [14]. Also high order statistics was also shown have been helpful to extract speech features even in a very short time window (several milliseconds) [15]. Further online implementation of the SPARSE algorithm is to be investigated.

5. SUMMARY AND CONCLUSIONS

A new algorithm named SPARSE has been developed based on current CI speech processing and the principle of sparse coding.

SPARSE appears to improve speech recognition in noise compared to conventional processing. Improvements are evident in noisy conditions and for listeners with poor baseline performance, where need is greatest. Greater familiarities with SPARSE should reveal greater improvements for cochlear implant users. Further evaluation is desirable.

6. ACKNOWLEDGEMENTS

Dr. Stefan Bleeck assisted with the Matlab interface for NIC experiments for CI subjects and provided invaluable critical insight to the work. Cochlear Europe Ltd. provided the

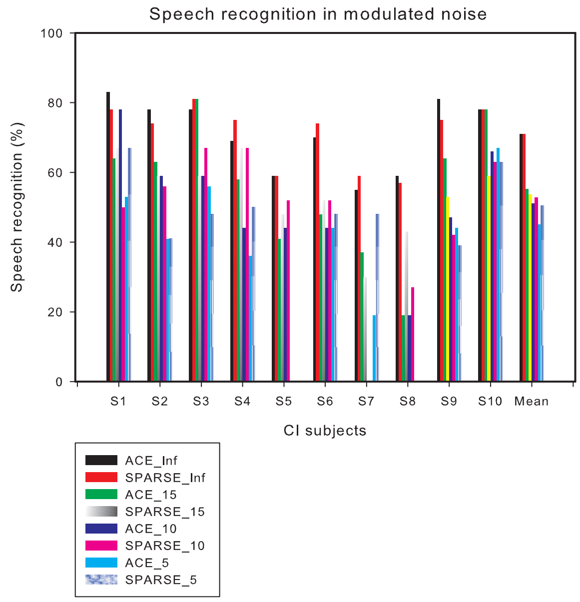

Figure 6: *CI user speech recognition scores in modulated noise. The 'ACE_Inf' represents the condition of quiet with ACE. 'SPARSE_15' represents the 15 dB SNR condition with SPARSE.*

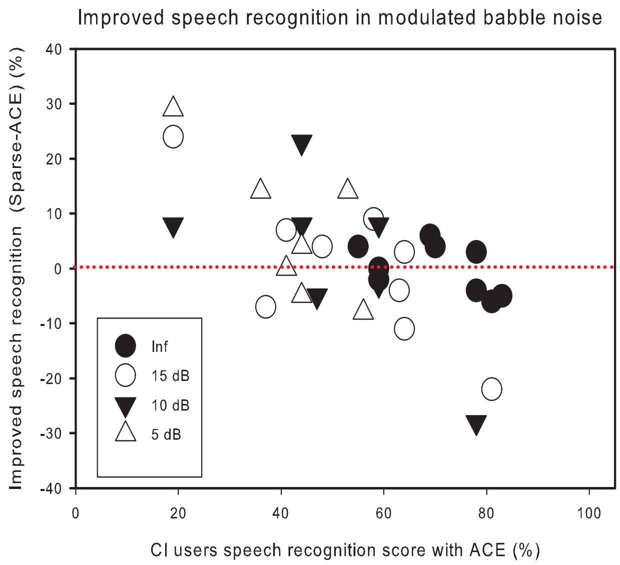

Figure 7: *CI user speech recognition scores in modulated noise with ACE and SPARSE in four SNR conditions. X-axis is the speech recognition score with ACE. Y-axis is the score difference between ACE and SPARSE in modulated noise. The dotted line represents the score difference of zero. For the differences under the dotted line, the recognition score decreases with SPARSE.*

NIC Matlab code and devices, plus partial support for a studentship.

REFERENCES

[1] B. S. Wilson and M. F. Dorman, "The Surprising Performance of Present-Day Cochlear Implants", *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 969–972, 2007.

[2] G. S. Stickney, F. Zeng, R. Litovsky and P. Assmann, "Cochlear implant speech recognition with speech maskers", *J. Acoust. Soc. Am*, vol. 116, pp. 1081–1091, 2004.

[3] L. M. Friesen, R. V. Shannon, D. Baskent and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants", *J. Acoust. Soc. Am*, vol. 110, pp. 1150–1163, 2001.

[4] S. Greenberg, W. A. Ainsworth, A. N. Popper and R. R. Fay, *Speech processing in the auditory system*. Springer, 2004.

[5] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs", *Curr Opin Neurobiol*, vol. 14, pp. 481–487, 2004.

[6] M. S. Lewicki, "Efficient coding of natural sounds", *Nature Neuroscience*, vol. 5, pp. 356–363, 2002.

[7] D. J. Field, "What is the goal of sensory coding?", *Neural Comput.*, vol. 6, pp. 559–601, 1994.

[8] H. B. Barlow, "Sensory mechanisms, the reduction of redundancy and intelligence" , *In National Physical Laboratory symposium No. 10*, 1959.

[9] R. Hartmann, G. Topp and R. Klinke, "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea", *Hearing Research*, vol. 13, pp. 46–72, 1984.

[10] A. Hyvarinen, "Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation", *Neural Comput.*, vol. 11, pp. 1739–1768, 1999.

[11] M. Cooke, "A glimpsing model of speech perception in noise", *J. Acoust. Soc. Am.*, vol. 119, pp. 1562-1573, 2006.

[12] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues ", *Science*, vol. 270, pp. 303–304, 1995.

[13] I. Potamitis, N. Fakotakis and G. Kokkinakis, "Speech enhancement using the sparse code shrinkage technique", *Acoustics, Speech, and Signal Processing, 2001. Proceedings.*, ICASSP '01. 2001 IEEE International Conference, 2001.

[14] M. Berg and E. Bondesson, "Real-Time implementation of a combined PCA-ICA algorithm for blind source separation", *Bleikinge Institute of technology*, Master thesis, 2005.

[15] M. C. Orr, B. J. Lithgow, and R. Mahony, "Speech features found in a continuous high order statistical analysis of speech", *Engineering in Medicine and Biology, EMBS/BMES Conference*, vol. 1, pp. 180-181, 2002.