# Independent component analysis: a new framework for speech processing in cochlear implants?

Guoping Li, Mark E. Lutman

*ISVR, University of Southampton,*
*Highfield, Southampton, SO17 1BJ*
{lgp,mel}@isvr.soton.ac.uk

***Abstract*— In this paper, we described the experiments we are going to do on the sparseness feature of speech signal and its relationship with the speech perception. And then we give a brief introduction on the potential new framework for speech processing for cochlear implants.**

***Index Terms*—Cochlear implant, Speech perception, Blind Source Separation, Frequency domain ICA, Convolutive mixtures**

## I. Introduction

A cochlear implant is an electric device, stimulating the auditory nerves through electrodes which are implanted in the inner ear through surgery. It can help profoundly hearing impaired people achieve a score of 80 percent correct on high-context sentences according to the (NIH) consensus statement on cochlear implants (1995). However, a common problem encountered by cochlear implant users is the cocktail party problem, where different sounds are overlapped both in time and frequency domains.

How humans solve the cocktail party problem is still a puzzle although there has been lots of research, shining some lights on it (Monica, Ruth et al. 2004; Stickney, Zeng et al. 2004). One important theory for speech perception in the cocktail party is called glimpsy theory, stating that listener can focus on the non-overlapping area of the mixed signals (Cooke, 2005). This theory in principle is the same as the assumption that there is always a dominant signal at most parts of the time frequency matrix of the mixed signals ; i.e. the signals are sparse. One of question is how the sparseness of a specified signal will affect the performance of speech perception. And is there any sparseness-based ICA algorithm which could be implemented in the cochlear implant?

This paper is organized as follows. In section 2, we focus on the first question and an example of controlling the sparseness by kurtosis. Section 3 will give a general description of signal processing in a cochlear implant and propose a new framework for speech processing in cochlear implants. We then discussed the possible outcome of the experiments in section 4.

## II. Sparseness and Speech Perception

Barlow observed that in many sensory nervous systems there are less neurons at later stages of processing than those at earlier stages. This is referred as ‘sparse coding’. What would happen if we make the signal representation more sparse? A standard way to measure the sparseness is kurtosis, measuring the $4^{th}$ moment relative to the variance squared:

$$k = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(r_i - \overline{r}\right)}{\sigma^4} - 3 \qquad (1.1)$$

If the variance is normalized to 1 and mean is 0, we have:

$$k = \frac{1}{n}\sum_{i=1}^{n}\left(r_i\right)^4 - 3 \qquad (1.2)$$

Using simple kurtosis gradient ascent projection pursuit algorithm (JV Stone, 2005), we can get a series signals with kurtosis ascending.

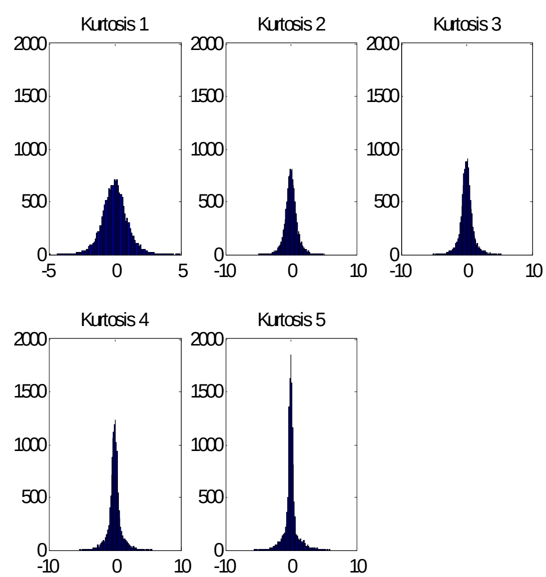


*Figure 1 Signal with kurtosis increasing.*

The input to the projection pursuit algorithm is a mixture of babble noise and VCV (vowel consonant vowel) speech signal /aba/ made by the head related transfer function measured by Gardner and Martin (1994) from a KEMAR dummy head under anechoic conditions. The output is the estimation of signal /aba/ with different kurtosis.

The psychoacoustic experiments are being prepared for normal hearing subjects. The assumption is that the speech perception score will improve with the increase of the kurtosis of the signal. A function of kurtosis and speech recognition score will be estimated.

## III. SIGNAL PROCESSING IN COCHLEAR IMPLANT

The speech processing algorithm of a cochlear implant is to extract the speech envelop information in different frequency bands, and its amplitude will modulate a series of pulses to stimulate the auditory nerves (see Fig 2). A faithful expression of such mixture in a cocktail party environment will make the cochlear implant users confused, and the speech recognition could drop to zero, while normal hearing ears can still manage a score above 90% percent (Stickney, Zeng et al. 2004).

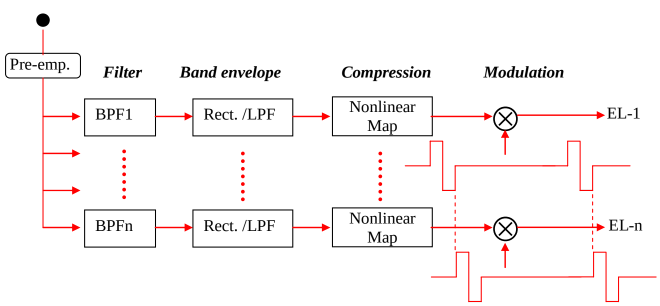


*Figure 2 A schematic overview of a cochlear implant.*

A key issue for cochlear implants is the speech processing strategy: what kind of speech features should be sent to the auditory neurons, via the electrodes? After much research, several envelope-based speech processing strategies were developed and achieved great success for the improvement of speech understanding performance for cochlear implant users (Wilson, Finley et al. 1991; Mcdermott, Mckay et al. 1992; Whitford, Seligman et al. 1995; Skinner, Holden et al. 1997; Wilson, Finley et al. 1997; Skinner, Fourakis et al. 1999). But still, there is much difficulty to understand speech when the users are in an adverse environment, such as cocktail party or noisy situation due to the overlapped area in the time frequency space.

Considering how the sound is coded in the auditory system and the development of independent component analysis techniques, we propose a new speech processing framework which could separate mixed sounds presented to the cochlear implant users simultaneously. Our aim is to incorporate the independent component analysis technique into the present speech processing algorithms. There are two methods, one is to put the independent component analysis as a pre-processor, and the other is to integrate the algorithm as part of speech processing algorithms, based on a commercial cochlear implant algorithm available. In principle, any algorithm which can separate convolutive audio signals could be helpful for the cochlear implant users. Once separated, the listener would be able to select the separated channel he is interested in. Here we are at the beginning of experiments.

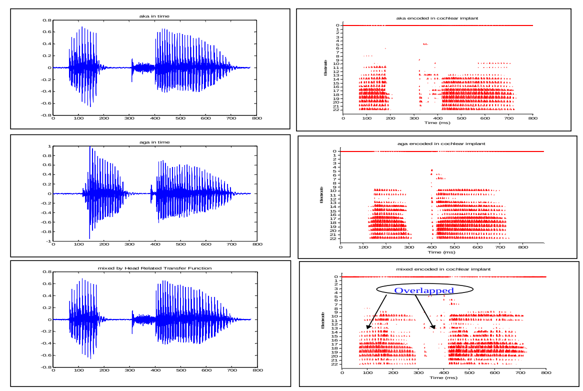


*Figure 3 An example of overlapped speech in the stimulation of a cochlear implant. The top two panels are the single speech signals /aga/, /aka/ and their 'spectrum stimulation' graphs. The bottom is the mixed signals and the overlapped areas are shown by the arrows.*

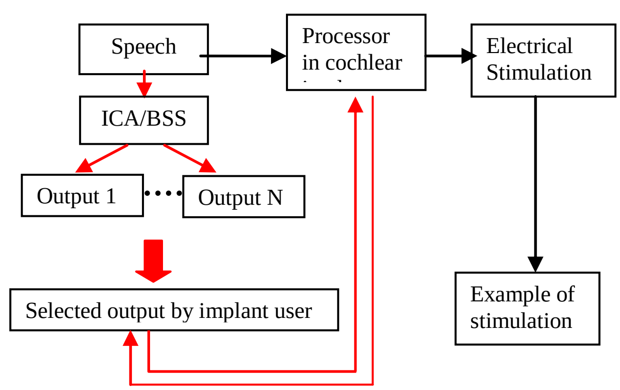


*Figure 4 A new framework for signal processing in cochlear implants.*

## IV. DISCUSSION

### A. General

Trying to understand and mimicking the human speech understanding ability in a cocktail party like environment is a difficult but rewarding process. It is an interesting question both for signal processing and hearing science community.

There has been lots of research on speech perception varying features of speech, such as spectrum detail, envelope and continuity. But there are few researches investigating the speech perception by manipulating the statistic features of speech, such as sparseness. Such analysis could provide a bridge between pure psychoacoustic experiments in hearing sciences and signal processing. The psychoacoustic experiments on the sparseness feature of speech does not necessarily prove that the human speech perception uses higher order statistics but it could show the higher order statistics are important for speech perception and it could be one of the key indexes to predict the speech perception performance in a cocktail party environment.

### B. The proposed framework

The critical part of this framework is which ICA/BSS algorithm we are going to use. There have been several algorithm candidates (Ikeda and Murata 1998; Mitianoudis and Davies 2003; Yilmaz and Rickard 2004), which could help to improve the speech recognition performance of cochlear implant users. Now we are working on these algorithms and trying to develop one suitable for the further objective and subjective evaluation for cochlear implant users.

So how far are we from the real implementation of ICA/BSS algorithms into the hearing aid devices, such as cochlear implant?

There has been some discussions whether ICA will help solve the practical problems (Haykin and Chen 2005). And certainly this debate will continue. But we believe this question can only be answered convincingly after both objective and subjective evaluation experiments.

## V. REFERENCES


Haykin, S. and Z. Chen (2005). "The cocktail party problem." Neural Computation **17**(9): 1875-1902.

Ikeda, S. and N. Murata (1998). An approach to blind source separation of speech signals, Skovde, Sweden, Springer-Verlag London.

Mcdermott, H. J., C. M. Mckay, et al. (1992). "A New Portable Sound Processor for the University-of-Melbourne Nucleus Limited Multielectrode Cochlear Implant." Journal of the Acoustical Society of America **91**(6): 3367-3371.

Mitianoudis, N. and M. E. Davies (2003). "Audio source separation of convolutive mixtures." **11**(5): 489.

Monica, L. H., Y. L. Ruth, et al. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer." The Journal of the Acoustical Society of America **115**(2): 833-843.

Skinner, M. W., M. S. Fourakis, et al. (1999). "Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies II. Consonants." Ear and Hearing **20**(6): 443-460.

Skinner, M. W., L. K. Holden, et al. (1997). "Parameter selection to optimize speech recognition with the Nucleus implant." Otolaryngology-Head and Neck Surgery **117**(3): 188-195.

Stickney, G. S., F. G. Zeng, et al. (2004). "Cochlear implant speech recognition with speech maskers." Journal Of The Acoustical Society Of America **116**(2): 1081-1091.

Stone, JV (2005) http://www.shef.ac.uk/~pc1jvs

Whitford, L. A., P. M. Seligman, et al. (1995). "Evaluation of the Nucleus Spectra-22 Processor and New Speech Processing Strategy (Speak) in Postlinguistically Deafened Adults." Acta Oto-Laryngologica **115**(5): 629-637.

Wilson, B. S., C. C. Finley, et al. (1991). "Better Speech Recognition with Cochlear Implants." Nature **352**(6332): 236-238.

Wilson, B. S., C. C. Finley, et al. (1997). "Temporal representations with cochlear implants." American Journal of Otology **18**(6): S30-S34.

Yilmaz, O. and S. Rickard (2004). "Blind separation of speech mixtures via time-frequency masking." Ieee Transactions on Signal Processing **52**(7): 1830-1847.