Applied Acoustics 76 (2014) 300–309

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Influence of coding strategies in electric-acoustic hearing: A simulation dedicated to EAS cochlear implant, in the presence of noise

Fabien Seldran [a,b,c,d,*], Stéphane Gallego [b,c,d], Hung Thai-Van [b,c,d,e], Christian Berger-Vachon [b,c,d]

[a] Vibrant Med-El Hearing Technology GmbH, 400 avenue Roumanille, CS 70062 - Bât 6, Sophia-Antipolis 06902, France
[b] INSERM U1028, Lyon Neuroscience Research Center, PACS Team (Speech, Audiology, Communication Health), Lyon F-69000, France
[c] CNRS UMR5292, Lyon Neuroscience Research Center, PACS Team (Speech, Audiology, Communication Health), Lyon F-69000, France
[d] University Lyon 1, Lyon F-69000, France
[e] Audiology and ORL Department, Edouard Herriot Hospital, Lyon F-69437, France

ARTICLE INFO

Article history:
Received 26 June 2012
Received in revised form 5 August 2013
Accepted 7 August 2013
Available online 26 September 2013

Keywords:
Electric-acoustic stimulation
Cochlear implant
Speech-coding strategies
N-of-M
Vocoder simulation

ABSTRACT

This work deals with electric-acoustic stimulation (EAS), which keeps the low frequency acoustic information and electrically codes the high frequencies of the signal. One of the goals of the coding strategies is to limit the phenomenon of channel interaction, which can occur in CIs. The "N-of-M" strategy, where only a subset of electrode channels is stimulated, may be of advantage. Generally, this processing is associated with a pre-emphasis filter. Two important parameters for the N-of-M strategy are the number of active channels (N) and the updating rate; the latter corresponds to the stimulation rate. M is the number of electrical channels.

The goal of this study was to investigate the influence of these parameters on speech intelligibility in EAS. The signal was presented, in simulation, to normal-hearing (NH) subjects in acoustic (A), electric (E) and electric-acoustic conditions. Recognition performance was measured in quiet and in the presence of background noise (cafeteria noise).

Signal-to-noise ratios (SNRs) ranged from 0 to +12 dB. Fifteen listeners participated in the experiment. The N values ranged from 2 to 10 (out of 10); M was 10. The frame updating rate was 250 updates per second (ups) and 1000 ups.

Results showed that increasing N from 2 to 10 improved speech intelligibility, especially in the presence of the background noise, under E and EAS conditions. In noisy situations, 2/10 coupled with a high-pass pre-emphasis filter led to results similar to the 10/10 condition. Changing the frame rate from 250 ups to 1000 ups did not modify the performance.

Future investigations on patients using EAS are now needed to validate the performance seen with NH listeners. Above all, in the strategy 2 out of 10, the number of pulses per second can be divided by 20, and when the pre-emphasis is used only a slight decrease in performance is expected; this is of interest when interaction between the electrodes corrupts the performance.

1. Introduction

Owing to advances in surgical techniques it is now possible, in patients with partial deafness, to preserve residual hearing in the low frequencies [1,2]. Preservation of hearing in these patients allows hearing with two simultaneous stimulations: acoustical and electrical. This bimodal hearing is called electric-acoustic stimulation (EAS) [3–6]. Using this approach, a short electrode array is inserted into the basal region of the cochlea which normally responds to high-frequency sounds, while attempting to preserve the apical region which responds to lower frequencies. The latter can be stimulated acoustically.

The development of EAS had led CI manufacturers to design short electrode arrays. For example, the Iowa/Nucleus Hybrid CI

Abbreviations: A, acoustic stimulation (only); ACE, advanced combination encoder; CI, cochlear implant; CIS, continuous interleaved sampling; E, electric stimulation (only); EAS, electric-acoustic stimulation; FFT, fast Fourier transform; HP, high-pass; LP, low-pass; NBN, narrow-band noise; NH, normal-hearing; N-of-M, N out of M; PC, percentage of correct responses; RMS, root-mean-square; SMSP, spectral maxima sound processor; SPEAK, spectral peak; SNR, signal-to-noise ratio; SW, sine wave; ups, updatings per second.

* Corresponding author. Address: INSERM U1028, CNRS UMR5292, Centre de Recherche en Neurosciences de Lyon, Equipe Audition, Hôpital Edouard Herriot, Pavillon U, Place d'Arsonval, F-69437 Lyon Cedex 03, France. Tel.: +33 4 72 11 05 03; fax: +33 4 72 11 05 04.

E-mail addresses: fseldran@yahoo.fr (F. Seldran), sgallego@hotmail.fr (S. Gallego), hthaivan@gmail.com (H. Thai-Van), christian.berger-vachon@univ-lyon1.fr (C. Berger-Vachon).

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.08.003

System has a 6 or 10 mm electrode array [4]. MED-EL's FLEX[24] EAS array features a 20.9 mm array—compared to 26.4 mm for the standard (non-EAS) array [7,8]. The limitations of short electrode arrays are that they either provide a small number of channels compared to longer arrays or that the inter-electrode spacing is shorter. For instance, the Iowa/Nucleus Hybrid 10-mm CI System features six channels (compared to 12–22 for standard arrays) and the maximum inter-electrode spacing is about 1.6 mm [9]. MED-EL's FLEX[24] provides 12 stimulation channels, with an inter-electrode spacing of 1.9 mm (compared to 2.4 mm for MED-EL's standard array). A smaller inter-electrode spacing is likely to increase interaction between the electrodes. This interaction is a factor limiting speech intelligibility in CI listeners (e.g., [10–18]). It may explain why, in some patients, speech-recognition saturates, and even decreases, as the number of active electrodes is increased [19,20]. It is reminded that each frequency channel is connected to an electrode.

An approach to limit channel interaction in CIs is to use non-simultaneous pulses [21,22]. One of the earliest examples of sound-coding strategies of this type is the 'continuous interleaved sampling' (CIS) strategy [23]. In this strategy, electrodes are sequentially stimulated, limiting the temporal overlap. Advances in non-simultaneous stimulation have led to the development of 'N-of-M' strategies, such as the 'spectral maxima sound processor' (SMSP) [24], 'spectral peak picking' (SPEAK) [25,26], or 'advanced combination encoder' (ACE) [27]. The principle behind the N-of-M strategies is that only a subset (N out of M) of electrodes is stimulated during a given time frame [28]. The N electrodes that are stimulated correspond to the most-energetic spectral bands or 'spectral maxima'. Several studies have suggested that use of N-of-M strategies yields higher speech recognition than with CIS [24,26,29–31]. In addition, previous simulation studies have indicated that intelligibility with EAS is greater than when acoustic or electric stimulation alone are used [32,33]. It was shown, at least in quiet, that with EAS only a few active channels are needed to provide good speech intelligibility when residual hearing in low frequencies remains. Thus, N-of-M strategies may provide an efficient way of limiting channel interaction without sacrificing intelligibility. However, to our knowledge, the potential benefits of using an N-of-M strategy with a relatively small number of available channels (e.g., M = 10), in the presence of unprocessed low-frequency acoustic information (EAS), has not been thoroughly investigated. Thus, it is interesting to investigate the influence of the N-of-M strategy in the context of EAS, particularly in the presence of noise. The influence of the usual pre-emphasis filtering in this situation is also worth considering. In future investigations, when enough EAS implantees will be available, the results of the present study can be validated.

With N-of-M strategies, CI manufacturers apply a pre-emphasis filter (1200-Hz high-pass filter) to preserve the high frequency components. It is well known that in the long term average spectrum of speech most of the energy is located below 800 Hz; consequently channels up to 800 Hz are essentially used if no pre-emphasis filter is applied. With EAS, the electric stimulation starts near 500 or 700 Hz and the presence of a pre-emphasis filter for N-of-M coding can be discussed.

Another important parameter which is likely to influence channel interaction is the stimulation rate. High stimulation rates provide more information, but also increase channel interaction [34–37]. High stimulation rates better represent the temporal fine structure and fast amplitude fluctuations in the speech signal temporal envelope [21,38,39]. Nowadays, there is a tendency toward the use of high stimulation rates. However, the effect of high stimulation rates on speech recognition performance is not clear. In some cases increasing the rate of stimulation had a positive effect [29,40–45] and sometimes it had no effect on speech recognition [39,46–49]. With EAS techniques in cochlear implant systems a recent patent suggests that a low stimulation rate for the high frequency components would be better [50].

Numerous studies have examined how speech-recognition performance is influenced by the stimulation rate in CI patients and in normal hearing (NH) subjects listening to vocoded stimuli (simulated CI processing) [29,39,42,44–49]. Few studies have investigated how the frame rate and the choice of N, in N-of-M strategies, affect speech intelligibility [47,51–55]. Usually, these parameters are not manipulated independently.

The present study aims to investigate the effects of N, the influence of the pre-emphasis filter and the updating rate on speech intelligibility, with electric-only, acoustic-only and electric-acoustic stimulation combined.

CI processing was simulated for NH listeners using a vocoder and a residual low-frequency hearing simulated by a low-pass filtering as described previously in several studies [4,32,33,56–63]. In the current study, channels were dynamically updated to simulate the N-of-M strategy [64–66]. Strategies were evaluated in quiet and in the presence of background noise (cafeteria noise).

2. Methods

2.1. Listeners

Fifteen NH subjects (7 females, 8 males) aged 19–29 years (mean = 23.9 years) took part in this study. NH was defined by hearing thresholds below 20 dB HL at octave frequencies ranging from 250 to 8000 Hz.

Testing took place in a double-walled sound-attenuated booth at the Department of Otorhinolaryngology, Edouard Herriot Hospital, Lyon, France.

In accordance with the Declaration of Helsinki, written informed consent was obtained from the subjects prior to their inclusion in this study. The study was approved by the local Ethics Committee (CPP Sud-Est IV, Leon Berard Centre of Lyon, France. Number ID RCB: 2008-A01479-46).

2.2. Stimuli

Stimuli were dissyllabic French words uttered by a single male speaker [67]. Recordings of "cafeteria noise" contained a mixture of chatter with background noises produced by cutlery, plates, and glasses, typical of a crowded restaurant or cafeteria. Fournier's dissyllabic words were chosen because they are similar, in the French language, to the American Spondaic lists [67]. In France, this test is widely used for the evaluation of speech understanding in CI patients [68].

With the exception of the "quiet" condition in which the words were presented without any background noise, the word and noise signals were mixed prior to processing. SNRs were: 0, +6, and +12 dB. The "quiet" condition, included for comparison purposes, was treated as one of the SNR conditions. Stimuli were presented in such a way that the intensity of the speech signal in condition "acoustic +10 activated channels in quiet" would be at 65 dB SPL. Consequently, the signal was less energetic with an N-of-M strategy (N < M) than when using all the electric channels (M-of-M). Also, the signal was more energetic when noise was added.

2.3. Signal processing

All signal processing was implemented in Matlab® (MathWorks, Natick, MA). The signals (recorded at 44.1 kHz) were down-sampled to 16.7 kHz, yielding a Nyquist frequency of 8.3 kHz. This frequency corresponds, approximately, to the upper limit of the