NeuroImage: Clinical 2 (2013) 229–238

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Electrophysiological responses to emotional prosody perception in cochlear implant users

D. Agrawal [a,*], J.D. Thorne [b], F.C. Viola [b], L. Timm [a], S. Debener [b], A. Büchner [c], R. Dengler [a], M. Wittfoth [a,d]

[a] Department of Neurology, Hannover Medical School, Hannover, Germany
[b] Department of Psychology, Carl von Ossietzky Universität, Oldenburg, Germany
[c] Department of Otolaryngology, Hannover Medical School, Hannover, Germany
[d] NICA (NeuroImaging and Clinical Applications), Hannover, Germany

ARTICLE INFO

Article history:
Received 22 August 2012
Received in revised form 31 December 2012
Accepted 3 January 2013
Available online 14 January 2013

Keywords:
Emotional prosody
Cochlear implants
ERP
P200
Gamma band power

ABSTRACT

The present electroencephalographic (EEG) study investigated the ability of cochlear implant (CI) users to recognize emotional prosody. Two CI speech-processing strategies were compared: the ACE (Advance Combination Encoder) and the newly developed MP3000. Semantically neutral sentences spoken in three different emotional prosodies (neutral, angry, happy) were presented to 20 post-lingually deafened CI users and age-matched normal-hearing controls. Event related potentials (ERPs) were recorded to study the N100 and the P200 responses. In addition, event-related spectral power modulations were calculated to study the brain activity corresponding to the recognition of prosody in earlier (0–400) as well as later (600–1200) part of the stimuli where the prosodic features differed maximally. CI users with MP3000 strategy showed a higher proportion of correctly recognized prosodic information compared to the ACE strategy users. Our ERP results demonstrated that emotional prosody elicited significant N100 and P200 peaks. Furthermore, the P200 amplitude in response to happy prosodic information was significantly more positive for the MP3000 strategy compared to the ACE strategy. On spectral power analysis, two typical gamma activities were observed in the MP3000 users only: (1) an early gamma activity in the 100–250 ms time window reflecting bottom–up attention regulation; and (2) a late gamma activity between 900 and 1100 ms post-stimulus onset, probably reflecting top–down cognitive control. Our study suggests that the MP3000 strategy is better than ACE in regard to happy prosody perception. Furthermore, we show that EEG is a useful tool that, in combination with behavioral analysis, can reveal differences between two CI processing strategies for coding of prosody-specific features of language.

1. Introduction

In spoken language, emotionally salient information can be provided by variations in speech melody (emotional prosody) or by emotional semantics (verbal emotional content). Emotional prosody is the ability to express emotions through variations of pitch, intensity and duration (Scherer, 2003). Individuals with severe to profound hearing loss have a limited dynamic range in these parameters, thus affecting prosody recognition (Hopyan et al., 2011). Cochlear implants (CI) are thought to improve not only language perception per se, but also specific aspects of language. However, it is still an open question whether such improvements include the recognition of emotional prosodic information.

The CI encodes sounds electronically, bypassing the damaged cochlea, and electrically stimulates the auditory nerve. Speech coding strategies are extremely important in CI processing as they decompose audio signals into different frequency bands and deliver the stimulation pattern to electrodes, maximizing the user's communicative potential. A number of speech processing strategies have been developed over the past two decades mimicking firing patterns inside the normal cochlea (Loizou, 1999). The Advance Combination Encoder (ACE) was developed in the 1990s. This strategy separates speech signals into a number of sub bands (M) and derives the envelope information from each band signal. A subset of these (N) with the largest amplitude is then selected for stimulation (N out of M). In 2005, a new strategy, Psychoacoustic Advanced Combination Encoder (PACE), commercially known as MP3000 (this term will be used here), was developed. This strategy is based on a psychoacoustic-masking model that neglects redundant signals, thus saving valuable bandwidth for those components that are usually perceived by normal-hearing (NH) individuals. The strategy is similar to the MP3 compression algorithm (Nogueira et al., 2005).

There are reports highlighting the performance of CI users in understanding speech, in particular from phoneme to sentences, in quiet as well as in noisy environments. For example, researchers (Fu

* Corresponding author at: Department of Neurology, Hannover Medical School, Carl-Neuberg Strasse 1, Hannover, Germany. Tel.: +49 511 532 3737; fax: +49 511 532 3115.
E-mail address: Agrawal.deepashri@mh-hannover.de (D. Agrawal).

2213-1582
http://dx.doi.org/10.1016/j.nicl.2013.01.001

et al., 2005) reported that CI users' voice gender identification was nearly perfect (94% correct) when large differences in the fundamental frequency (F0) existed between male and female talkers. House and colleagues used semantically neutral Swedish utterances of four target emotions (angry, happy, sad, and neutral), and found that for CI users, the performance of mean vocal emotion recognition increased from 44% to 51% one year after processor activation (House, 1994). Comparably, another study (Luo and Fu, 2007) investigated the ability of NH listeners and CI users to recognize vocal emotions. CI users performed poorly with their own conventional processor, but their performance significantly improved as the number of channels was increased. Taken together these studies reflect mixed results, which might be due to the lack of objective dependent variables used. To overcome this issue event-related potentials (ERPs) could be used, as they do not rely on subjective and behavioral output measures. In addition, ERP methodology may prove useful in narrowing the range of candidate variables that must be analyzed in order to understand the phenomenon associated with improved behavioral performances (Deutsch et al., 2008).

In the last decade, studies have used ERPs to study emotion recognition. There is evidence that ERPs are an important objective measure of auditory emotional prosody differentiation and recognition (Pinheiro et al., 2011). ERP differences across emotions can be found as early as 200 ms after stimulus onset during both visual and auditory emotional processing (Agrawal et al., 2012; Kotz et al., 2006; Schapkin et al., 2000). The traditional ERP methodology reveals the phase-locked neural activities evoked by a particular cognitive process. Time-frequency (TF) analysis, on the other hand, can reveal the non-phase-locked (induced) neural activity that is hidden in standard averaged ERPs (Makeig et al., 2004; Tallon-Baudry et al., 1999). While ERPs and time-frequency analysis both provide a view of the events in brain's information processing stream, an advantage provided by TF analysis, relative to ERPs, is its potential to understand the brain's parallel processing of information. The oscillations at various frequencies reflect the interaction of multiple neural activities in the service of integrative and dynamically adaptive information processing (Roach and Mathalon, 2008).

This additional benefit of TF analysis, relative to ERPs, may also be manifested in greater sensitivity to the neuronal processes underlying prosody perception in CI users. Additionally, non-phase locked brain activity in the gamma-band range should indicate facilitated prosody recognition modulated by strategy (Hannemann et al., 2007; Lenz et al., 2007). However, although frequency-domain (spectral) analysis has been applied by various researchers in memory and speech perception paradigms (Fuentemilla et al., 2006; Muller et al., 2009), it is yet to be used in emotional prosody analysis in CI users.

The present study aimed to elucidate differences between the effects of the ACE and MP3000 coding strategies on emotional prosody recognition. Based on differences between the two speech coding strategies and the advantage of MP3000 strategy in improving spatial recognition cues, we hypothesized that MP3000 users might perform better compared with the ACE users on prosody identification, which should be reflected in behavioral measures and auditory ERPs. We further hypothesized that induced brain activity obtained with TF analysis should yield complementary effects not observable in the classic ERPs reported above.

In the current study, prosodic information was presented by using neutral, angry and happy tones of voice. Differences were assessed in the behavioral responses, the ERPs as well as TF analysis.

2. Materials and method

2.1. Participants

Forty right-handed native German speakers (22 females, 18 males), aged 25–60 years (mean = 41.5 years, SD = 7) participated in the experiment. The first group of participants (Group I) consisted of 20 CI users (mean = 42.1 years, SD = 7.01) wearing a Nucleus cochlear implant system as depicted in Table 1. Subjects had used their implants continuously for at least 12 months and had at least 20% speech in noise perception scores on the Oldenburg sentence test (Wagener et al., 1999) prior to the study. Furthermore, subjects were divided into two subgroups with the aim of comparing two speech-coding strategies (ACE vs. MP3000). The first group of participants (Group IA) consisted of ten individuals (mean age of 42.1, SD = 8.2) with an ACE strategy as their default while their speech processors were programmed with MP3000 strategy for the experiment. Similarly, the remaining ten participants in (Group IB) were CI users (mean = 41.1 SD = 7.3) with MP3000 as their default and ACE as the experimental strategy.

A control group (Group II) comprised age and gender-matched NH participants (age range: 25–55 years; mean = 41 years, SD = 7.1). All participants had normal intelligence and reported no history of psychological or neurological problems. In order to test for depression, Beck's Depression Inventory (BDI) was used (Beck et al., 1996). None of the subjects had clinically-relevant symptoms of a depressive episode. The study was carried out in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Hannover Medical School. All participants gave written informed consent prior to the recording and received monetary compensation for their participation.

2.2. Stimuli

The stimulus material consisted of 150 semantically-neutral German sentences with neutral, happy and angry prosody (50 each) spoken by a trained female German speaker. Stimuli were recorded with a sampling rate of 44.1-kHz and a 16-bit digitizer (Kotz and Paulmann, 2007; Kotz et al., 2003; Wittfoth et al., 2010). All sentences started with personal pronouns (For example, "*Sie hat die Zeitung gelesen*"; "She has read the newspaper"). The stimulus material was prosodically analyzed using Praat 5.1.19 (Boersma and Weenink, 2005). Table 2, displays the differences in the fundamental frequency (F0), intensity, and duration of the sentences extracted. It was observed that the there were significant pitch differences across emotions but no significant intensity differences seen. These significant differences in pitch contours are illustrated in Fig. 1.

2.3. Procedure

Testing was carried out in a sound-proofed chamber. Subjects were seated in a comfortable armchair facing a computer monitor placed at a distance of one meter. The stimuli were programmed

Table 1
Demographic data of CI users.

Patient	Age	Gender	Duration of implant-use (in years)	Strategy
1	39	M	3	ACE
2	51	F	3	ACE
3	35	F	3	ACE
4	48	M	4	ACE
5	42	F	3	ACE
6	49	M	2	ACE
7	47	F	2	ACE
8	29	F	3	ACE
9	52	M	4	ACE
10	56	M	5	ACE
11	35	F	3	MP3000
12	43	F	4	MP3000
13	22	F	2	MP3000
14	42	M	2	MP3000
15	39	F	3	MP3000
16	46	F	4	MP3000
17	42	F	3	MP3000
18	48	M	3	MP3000
19	40	M	4	MP3000
20	41	F	3	MP3000