Hearing Research 342 (2016) 13–22

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users

Raphael Koning, Jan Wouters*

ExpORL, Dept. Neurosciences, KU Leuven, Herestraat 49 bus 721, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 2 October 2014
Accepted 7 September 2016
Available online 30 September 2016

Keywords:
Cochlear implants
Onset enhancement
Speech signal processing

ABSTRACT

Speech perception by cochlear implant (CI) users can be very good in quiet but their speech intelligibility (SI) performance decreases in noisy environments. Because recent studies have shown that transient parts of the speech envelope are most important for SI in normal-hearing (NH) listeners, the enhanced envelope (EE) strategy was developed to emphasize onset cues of the speech envelope in the CI signal processing chain. The influence of enhancement of the onsets of the speech envelope on SI was investigated with CI users for speech in stationary speech-shaped noise (SSN) and with an interfering talker. All CI users showed an immediate benefit when *a priori* knowledge was used for the onset enhancement. A SI improvement was obtained at signal-to-noise ratios (SNRs) below 6 dB, corresponding to a speech reception threshold (SRT) improvement of 2.1 dB. Furthermore, stop consonant reception was improved with the EE strategy in quiet and in SSN at 6 dB SNR. For speech in speech, the SRT improvements were 2.1 dB and 1 dB when the onsets of the target speaker with *a priori* knowledge of the signal components or of the mixture of the target and the interfering speaker were enhanced, respectively. The latter demonstrates that a small benefit can be obtained without *a priori* knowledge.

1. Introduction

Many CI users achieve very good SI performance in quiet, but in adverse listening conditions (e.g. interfering background sounds and/or reverberation). CI users require a SNR that can be 5–15 dB higher in stationary SSN than NH listeners to achieve the same SI (Dorman et al., 1998; Nelson et al., 2003). In fluctuating background noise, this difference can be even bigger (Nelson et al., 2003; Zeng et al., 2005).

CI users experience difficulties due in part to the reduced frequency resolution that is determined by the number of independent information transmission channels related to the number of electrodes of the electrode array that is implanted along the cochlea. In current clinical CI processing strategies, mainly the temporal envelope information is transmitted. While the envelope information of a few frequency bands is sufficient in quiet for SI (Shannon et al., 1995), there are limits to speech understanding in adverse listening conditions when just the envelope information is presented to the listener. In general, the encoding of additional monaural and binaural cues is very limited in CIs (Wouters et al., 2013). With regard to sound source segregation in CI users, the very poor speech recognition results in adverse listening conditions (in particular with non-stationary interfering background sounds) often rather reflect the limitation of the speech processing strategies than the inability of CI users to segregate sound sources (McLaughlin et al., 2013). Therefore, new speech strategies try to improve SI of CI users by enhancing speech features such as the fundamental frequency F0 in tonal languages (Geurts and Wouters, 2001; Vandali et al., 2005; Milczynski et al., 2012; Laneau et al., 2006; Vandali and van Hoesel, 2011, 2012) or transient sounds (Geurts and Wouters, 1999; Vandali et al., 2000; Bhattacharya et al., 2011; Koning and Wouters, 2012). The improved access to speech cues may lead to better sound segregation and therefore better parsing of the auditory scene.

The idea that enhanced onsets of the speech envelope could potentially increase SI for CI users is motivated by neurophysiological findings and recent studies that investigated which parts of the speech envelope contribute most to SI from an information theory point of view.

By the direct electric stimulation of auditory nerve fibers (ANFs), the CI bypasses the processing up to the ANF in healthy hearing. One effect in NH listeners leading to good encoding of onsets in the

* Corresponding author.
E-mail addresses: raphael.koning@med.kuleuven.be (R. Koning), jan.wouters@med.kuleuven.be (J. Wouters).

http://dx.doi.org/10.1016/j.heares.2016.09.002
0378-5955/

speech envelope is the adaptation effect at the synapse to the ANF (Smith, 1979; Delgutte and Kiang, 1984). At onsets of the speech envelope, there is an increase in the firing rate in the ANF followed by a rapid decay over several milliseconds and a slower decay to a steady level (Smith, 1979; Westerman and Smith, 1984; Lütkenhöner and Smith, 1986). This natural emphasis of the firing rate at onsets of the speech envelope has three main effects (Delgutte, 1997). Firstly, it indicates spectral-temporal regions containing a lot of information. Secondly, the temporal precision of the ANF representation of onsets is increased. Finally, the contrast between successive segments is enhanced. Since the origin of this adaptation effect - the hair-cell-to-neuron synapse - is bypassed by a CI, a signal-processing scheme that includes enhancement of the onsets of the speech envelope seems to be a promising approach to increase SI for CI users in adverse listening conditions. The adaptation effect is not completely caused by the hair-cell-to-neuron synapse and adaptation can be observed in response to electrical stimulation. However, the adaptation characteristics of ANFs differ significantly for electric and acoustic stimulation (Parkins, 1989; Killian et al., 1994; Zhang et al., 2013). The missing rapid adaptation response and the much weaker short-term adaptation response in electric hearing lead to a weaker encoding of onsets of the speech signal. Therefore, the onset speech cue is not as clearly represented in ANF spike patterns in CI users than in NH listeners. We hypothesize that this can have a detrimental effect on SI in CI users.

The fact that all sensorineural systems are sensitive to changes of the input signal (Kluender et al., 2003) led to information-theoretic approaches that investigated which parts of the speech signal are important for SI. In several studies (Stilp and Kluender, 2010; Stilp et al., 2013; Chen and Loizou, 2012) it was shown that the rapidly changing speech cues like onsets, transients and offsets contribute most to SI. Due to the rapid temporal and spectral changes of these speech cues, they contain a lot of information in comparison to the quasi-stationary parts of the speech signal. Entropy-based models have been developed and allowed to quantify SI and which parts contribute most to it.

For all those reasons, envelope enhancement algorithms were developed that focused on the transient parts of the speech signal. An enhancement of the transient parts of the speech signal seems to be even more important in adverse listening conditions, because a lot of transient parts are weak in energy (e.g. consonants) and short in duration. Therefore, they are easily affected by an interfering background sound and are, therefore, very vulnerable parts of the speech envelope. In NH listeners, it was shown that an amplification of the transient parts led to a significant SI improvement in various adverse listening conditions (Kennedy et al., 1998; Lorenzi et al., 1999; Apoux et al., 2004; Yoo et al., 2007; Rasetshwane et al., 2009). All envelope enhancement algorithms for NH listeners required *a priori* knowledge of the incoming speech signal. Furthermore, the signal was enhanced prior to the mixing with the interfering background sound. The use of *a priori* knowledge is necessary because state-of-the-art signal processing techniques do not allow to segregate the target and the interfering background signals from the noisy mixture of the components. Therefore, *a priori* knowledge of the signal components is used to show the potential of those speech enhancement approaches.

None of the algorithms that were developed for NH listeners were evaluated for the application in auditory prostheses. Other algorithms that amplify short-duration transient cues were developed for the application in a CI. The transient emphasis spectral maxima (TESM) strategy (Vandali, 2001) amplifies transient parts of the noisy envelope by applying a gain factor that is obtained by a comparison of the average energy in three successive time windows with each 20 ms length in each frequency band. The algorithm applies maximum gain when the peak of a rapid increase followed by a decrease in energy is reached. Therefore, it focuses more on the transients than on the onsets of the speech envelope. The TESM was shown to be beneficial in multitalker babble noise (Vandali, 2001) and in combination with an additional spectral expansion stage (Bhattacharya et al., 2011). When the speech input level was already high at 75 dB sound pressure level (SPL), no benefit of the TESM strategy was observed (Holden et al., 2005).

The EE algorithm (Koning and Wouters, 2012) was shown to improve SI in CI simulations with NH listeners in stationary SSN by 2.5 dB when the onsets were extracted with knowledge of the target speaker. The improvement was 1.7 dB when an ideal Wiener filter was used in the front-end off the onset extraction stage of the noisy signal. The Wiener filter eliminates with *a priori* knowledge of the target and the interfering components the background noise in speech pauses and scales the input signal according to the SNR when speech and noise are present. It is possible to extract the onsets from the processed signal in the onset extraction stage afterwards. A benefit in SRT of 2.6 dB was obtained when the onsets of the target speaker were enhanced with a competing talker as the background.

The EE algorithm was developed from the enhanced envelope continuous interleaved sampling (EECIS) algorithm. It was based on the idea of mimicking the short-term adaptation effect of the ANF (Geurts and Wouters, 1999). For CI listeners, a significant improvement in SI was obtained in quiet in a vowel-consonant-vowel (VCV) stop consonant recognition task. However, the EECIS algorithm was not tested in adverse listening conditions.

We hypothesized that the enhancement of onset cues in the speech envelope can lead to better SI in adverse listening conditions. Enhanced onset cues could lead to better segregation of the target speaker from the interfering background sounds (Bregman, 1990; Hu and Wang, 2007; Shamma et al., 2011). Rapid adaptation may play a role in stream formation and perceptual grouping for NH listeners. Christiansen et al. (2014) showed that a model consisting of auditory processing (Dau et al., 1997) followed by temporal coherence-analysis (Elhilali et al., 2009) could explain the segregation of temporally overlapping sounds into different auditory streams when the adaptation effect was included in the model.

This study investigated the potential of the EE strategy in CI. SI with the EE strategy was studied using two sentence recognition tasks for speech in stationary SSN and an interfering talker. The improved access to onset cues should help to segregate the target speaker from the interfering background sound independent from the type of interferer because the ability to parse the auditory scene should be enhanced. Furthermore, loudness perception was investigated with a loudness rating (LR) task and the recognition of stop consonants was studied in quiet and SSN to investigate the information transmission capabilities of the EE strategy of the consonant features *voicing* and *place of articulation*. The clinical advanced combination encoder (ACE) strategy served as the reference strategy in all listening tasks.

2. Methods

2.1. Subjects

Six CI listeners participated in the sentence recognition task in stationary SSN, the stop consonant recognition task and a LR. Four of the subjects also participated in the sentence recognition task with the competing talker. The information about the subjects, their etiology and their CI is given in Table 1. All subjects were speaker of Dutch/Flemish language. They were paid for their travel expenses and signed an informed consent form.