Hearing Research 328 (2015) 24–33

Contents lists available at ScienceDirect

Hearing Research

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

Research paper

The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations

Michel R. Benard [a, b, *], Deniz Başkent [b, c]

[a] Pento Audiology Center Zwolle, Oosterlaan 20, 8011 GC Zwolle, The Netherlands
[b] University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, PO Box 30.001, 9700 RB Groningen, The Netherlands
[c] University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form
15 June 2015
Accepted 22 June 2015
Available online 25 June 2015

Keywords:
Speech perception
Top-down restoration
Phonemic restoration
Visual speech cues
Cochlear implants

ABSTRACT

In complex listening situations, cognitive restoration mechanisms are commonly used to enhance perception of degraded speech with inaudible segments. Profoundly hearing-impaired people with a cochlear implant (CI) show less benefit from such mechanisms. However, both normal hearing (NH) listeners and CI users do benefit from visual speech cues in these listening situations. In this study we investigated if an accompanying video of the speaker can enhance the intelligibility of interrupted sentences and the phonemic restoration benefit, measured by an increase in intelligibility when the silent intervals are filled with noise. Similar to previous studies, restoration benefit was observed with interrupted speech without spectral degradations (Experiment 1), but was absent in acoustic simulations of CIs (Experiment 2) and was present again in simulations of electric-acoustic stimulation (Experiment 3). In all experiments, the additional speech information provided by the complementary visual cues lead to overall higher intelligibility, however, these cues did not influence the occurrence or extent of the phonemic restoration benefit of filler noise. Results imply that visual cues do not show a synergistic effect with the filler noise, as adding them equally increased the intelligibility of interrupted sentences with or without the filler noise.

1. Introduction

Normal hearing (NH) listeners benefit from auditory top-down restoration mechanisms in acoustically complex listening situations to enhance speech perception. Warren (1970) showed for the first time that this effect was so strong that listeners believed that they heard a phoneme in a sentence, which was in fact replaced by an extraneous sound. Inspired by this study, the restoration capacity of the perceptual system was later studied with multiple interruptions in speech (e.g. Bashford et al., 1992; Başkent et al., 2009; Jin and Nelson, 2006; Verschuure and Brocaar, 1983).

Interrupting continuous speech distorts the intonation, voice quality and co-articulation patterns of fluent speech (Brennan and Schober, 2001; Mattys et al., 2012). In the phonemic restoration paradigm, filling the gaps between multiple segments of interrupted speech with filler noise bursts not only helps the perceptual system to restore a continuous speech stream (Riecke et al., 2011, 2009; Srinivasan and Wang, 2005), but it also improves the intelligibility (Bashford and Warren, 1979; Benard and Başkent, 2013; Bhargava and Başkent, 2012; Powers and Wilcox, 1977; Verschuure and Brocaar, 1983). Cognitive factors and linguistic skills (Bashford et al., 1992; Bronkhorst et al., 1993; Saija et al., 2013; Sivonen et al., 2006b; Srinivasan and Wang, 2005; Wang and Humes, 2010; Warren, 1983) and especially receptive vocabulary and verbal intelligence (Benard et al., 2014) have been shown to play an important role in the restoration of the audible segments into a meaningful sentence.

Profoundly hearing-impaired people who use a cochlear implant (CI, an implantable auditory prosthesis) experience more

* Corresponding author. Pento Audiology Center Zwolle, Oosterlaan 20, 8011 GC Zwolle, The Netherlands.
E-mail addresses: m.r.benard@umcg.nl (M.R. Benard), d.baskent@umcg.nl (D. Başkent).

http://dx.doi.org/10.1016/j.heares.2015.06.013
0378-5955/

problems than NH listeners in understanding speech in difficult listening situations, e.g. in the presence of background noise (Fu and Nogaki, 2005). One potential factor that is proposed to contribute to this difficulty is that the speech signal transmitted to the auditory nerve via electric stimulation might not contain the necessary speech cues to induce top-down restoration mechanisms (Başkent, 2012; Bhargava et al., 2014). This hypothesis was further supported by Benard and Başkent (2014), who observed no perceptual benefit of filler noise in the silent intervals when NH listeners were trained with interrupted speech further degraded with acoustic simulations of CIs. However, an improved speech perception and restoration benefit was observed with simulations of electric-acoustic stimulation (EAS; the low frequencies are acoustically available and the high frequencies are stimulated via a CI), where the limited additional acoustic low-frequency speech cues seemed to help (Başkent, 2012). Apart from this, Bhargava et al. (2014) observed that the restoration patterns in actual CI users are indeed different from those of NH listeners; a restoration benefit was only observed when the speech segment durations were made longer and the interruptions shorter. These findings combined imply that CI users could perhaps benefit from top-down repair mechanisms, but only if the degraded speech signal contains the necessary speech cues, and perhaps also when supplemented by additional perceptual cues (such as in the case of EAS).

One form of such supplemental cues that can help listeners in complex listening situations, as well as with the auditory top-down restoration mechanisms, are visual speech cues. Visual information strongly influences auditory perception, to the degree that it can induce a different percept than the actual acoustic speech information presented alone (e.g. McGurk, 1976; Wiersinga-Post et al., 2010; Valkenier et al., 2012). Lip-reading increases the intelligibility by aiding in extracting the place of articulation from the visual cues of modulation of the area between the lips (Grant and Seitz, 2000). Seeing the face and the lips of the talker facilitates for example speech segmentation (Mitchel and Weiss, 2013) and it can enhance the capacity of auditory cortex of the listeners to track the temporal speech envelope (Cunillera et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). Furthermore, visual speech cues increase speech intelligibility (in noise) for both NH listeners and hearing-impaired listeners, indicating that they provide cues that would otherwise not be delivered due to (usually high-frequency) hearing loss (Başkent and Bazo, 2011; Gilbert et al., 2012; Grant and Seitz, 2000; Grant et al., 1998; McGrath and Summerfield, 1985; Ross et al., 2007). Previous studies have shown that CI users depend heavily on visual speech cues in complex listening environments (Doucet et al., 2006; Rouger et al., 2007; Song et al., 2014), and greatly benefit from lip-reading (Lyness et al., 2013).

In this study, we investigate if the accompanying video of the speaker, in addition to the auditory stimuli alone, transmits supplementary speech information to the listener in such way that it can enhance phonemic restoration of periodically interrupted sentences, with or without the further degradations of CI simulations. The effects of visual cues on top-down restoration has been investigated by only few studies and it has been mainly limited to using single interruptions, without using any other degradations (Bhat et al., 2014; Shahin and Miller, 2009; Shahin et al., 2012; Trout and Poser, 1990). For example, Trout and Poser (1990) replaced a single critical phoneme in a sentence with white noise and studied the benefit of visual speech cues on the detection of the replaced segment. They found that visual speech cues reduced the bias of reporting replaced phonemes as continuous, but this study did not show an increase in the intelligibility or the top-down restoration of the sentences. On a more optimistic note, Shahin and Miller (2009) investigated the auditory and visual integration of tri-syllabic words with single interruptions. The auditory stimuli used were either interrupted or continuous words, in which the middle fricative/affricate was either replaced by (interrupted word) or superimposed with (continuous word) white noise. Even when the words were interrupted they were identifiable and unambiguous. A video of only the lip movements that were either congruent incongruent or static (no movements) was presented with the auditory stimuli. Participants had to report if the auditory stimulus was continuous (phonemic restoration illusion) or interrupted (illusion failure). In contrast with the findings of Trout and Poser (1990), the results showed that congruent visual speech cues resulted in a stronger illusion of phonemic restoration over longer white noise intervals in single words. Thus, additional visual speech cues (e.g. place of articulation in lip-reading, improved speech segmentation or tracking of the temporal envelope) are shown to increase the speech intelligibility of uninterrupted speech (e.g. Zion Golumbic et al., 2013), but evidence is mixed that these cues might increase the top-down restoration effects on perception of interrupted speech (Shahin and Miller, 2009; Trout and Poser, 1990).

Two main hypotheses are central in the present study. Firstly, based on the overall positive effects of visual cues on intelligibility of distorted speech, we hypothesized that the intelligibility of interrupted sentences with and without filler noise would increase with the addition of visual speech cues. Secondly, we hypothesized that the positive effect of visual speech cues on the restoration of speech with single interruptions would also extend to improved phonemic restoration benefit of filler noise, i.e. stronger restoration effects of filler noise due to visual speech cues for sentences with multiple periodic interruptions. These hypotheses were investigated for interrupted speech without (Experiment 1) and with further spectral degradations as it can happen in CI (Experiment 2) or EAS (Experiment 3) speech transmission. The rational behind the design of these three experiments is that good use of visual speech cues could potentially improve the intelligibility of speech in background noise for CI users. However, it remains largely unknown from the literature what the influence of these cues is on the intelligibility of temporally interrupted sentences in CI or in EAS speech transmission. In the latter, the auditory system can take advantage of the additional strong voicing information cues available, such as voice fundamental frequency (F0) contours (Brown and Bacon, 2009a), low-frequency segmental phonetic cues (Incerti et al., 2013), and low-frequency phonetic cues like the first formant (F1) and formant transition cues (Kong and Carlyon, 2007).

2. Experiment 1: Perception of temporally interrupted speech with visual cues

2.1. Rationale

Shahin and Miller (2009) showed that the integration of auditory and visual stimuli enhances the top-down restoration benefit in words with single interruptions, allowing the auditory system to restore longer noise-filled intervals. In Experiment 1, we explored the effect of visual cues on a different form of top-down restoration, namely, that of sentences with multiple interruptions. The hypotheses were that additional visual cues to the auditory stimuli would increase the overall intelligibility of interrupted speech and that it would enhance the phonemic restoration benefit of filler noise in sentences with multiple interruptions.

2.2. Materials and methods

2.2.1. Participants

Twelve native speakers of Dutch (6 women), aged between 18 and 26 (mean age = 21.7 years, standard deviation (SD) = 2.6