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

years), participated in this experiment. A normal speech and language development was confirmed with a questionnaire. Normal hearing, i.e. thresholds equal or less than 20 dB hearing level for all audiometric speech frequencies for both ears, was confirmed. The participants were unfamiliar with the speech corpus and with the speech manipulations. The Medical Ethical Committee of the University Medical Center of Groningen approved this study. Participants were recruited via posters at public places. They received information about the experiment via e-mail at least two weeks before the experiment, and they provided written informed consent before data collection began. The participants received an hourly payment.

2.2.2. Auditory and visual stimuli

The sentence materials were the same as in the Dutch speech corpus designed by Versfeld et al. (2000), representing daily conversational speech, with each sentence consisting of 4–9 words. The sentences were newly recorded in a photographer studio of the University Medical Center Groningen by means of a Canon 5D Mark II high definition camera with inbuilt microphone, placed at eye height, 2 m from the face of the speaker. The video was digitally recorded in 25 frames per second, with 1920 × 1080 pixels per frame, and was stored on a hard disk. The sound was digitally recorded with a sampling rate of 44.1 kHz.

The speaker was a female master student of the Linguistics Department of the university, of 23 years of age, and with no accent. The face and the top part of the shoulders of the speaker were filmed (Fig. 1). During recording, the written sentences were presented to the speaker via a monitor using a custom-made MATLAB program running on a Macintosh computer. This allowed stabilization of the eye movements of the speaker. The speaker was instructed to speak calmly and clearly while looking in the camera, starting a sentence every 10 s as cued by the program. The words were uttered at a normal speaking rate matching the original recordings of Versfeld et al. (2000) as closely as possible. Each sentence was recorded 3 times and a short break was built in the program after every 10 min of recording time. The recording session lasted approximately 3 h. In total 21 sets of 13 sentences each were recorded (273 in total). After the recording session, individual sentences were extracted from the video stream by selecting the clearest recording of the 3 takes of each sentence. The durations of the video and audio recordings were shortened to 5 s. In addition to editing the recordings into individual sentences, the static background noise was removed by analyzing and subtracting the spectrum of the interval before the speaker started uttering the sentence, by means of the sound editing software Audacity™.

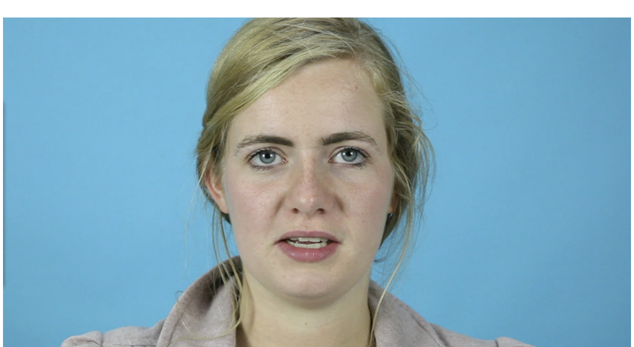

Fig. 1. Example frame of the video of the speaker.

The filler noise, taken from the original corpus of Versfeld et al. (2000), was a speech-shaped noise matching the long-term average speech spectrum of the recorded sentences by Versfeld et al.

2.2.3. Periodic interruptions

The audio recordings of the sentences were interrupted with a periodic square-wave function, with a ramped cosine with on and off transitions of 10 ms, an interruption rate of 1.5 Hz and a duty cycle of 50% (based on Benard and Başkent, 2013). The sentences were presented at 60 dB SPL and the filler noise at 70 dB SPL, calibrated at an approximate position of the participant's ear. This resulted in the first type of speech stimuli with speech segments of 333 ms followed by a silent interval of the same duration. Intrinsic to periodic interruption is that a silent or noise filled interval can make words partially (beginning, middle or end) or even entirely inaudible. The filler noise used to induce top-down repair was produced by multiplying the speech-shaped noise with the inphase inversed square wave. The periodic filler noise was added to the interrupted speech signal, resulting in the second type of speech stimuli with speech segments followed by filler noise segments of 333 ms each, overlapping in the midst of their on and off transitions (Başkent et al., 2009).

2.2.4. Experimental setup

The experimenter and the participant were seated in a sound proof chamber designed for clinical audiometry. A customized MATLAB program (based on Benard and Başkent 2014) was used to run the experiment. The audio stimuli, the processed sentences, were presented via an external soundcard (AudioFire 4, Echo Digital Audio Corporation) to headphones (Sennheiser HD 600). During training sessions the sentence text was also displayed on the computer monitor. The visual stimuli, the video of the speaker, were presented in synchrony with the auditory speech stimuli by making use of the MATLAB Psychophysics Toolbox (version 3). The experimenter listened to the participant's response and scored the entirely correctly repeated words online using the MATLAB program (method based on Benard and Başkent, 2013; Benard et al., 2014; Benard and Başkent 2014). The program calculated percent correct scores, using the ratio of the number of correctly repeated words to the total number of words within the speech set presented.

2.3. Experimental procedure

The participants were trained and tested with interrupted sentences without spectrotemporal degradations. Their task was to listen to one processed sentence stimulus at a time, with or without accompanying video of the speaker, and to repeat all words they heard, even if this led to a nonsense or incomplete sentence. Guessing the missing words to form a sentence was encouraged. Training was given before actual data collection.

During training, feedback was provided after the response of the participant, in the form of playback of the uninterrupted and interrupted sentences sequentially, while the written sentence was presented on the monitor (same feedback procedure as Benard and Başkent, 2013). The participants were trained with interrupted sentences with and without filler noise. One set of 13 sentences was used per condition during the training of this group, resulting in 26 unique sentences in total as training. This training session took around 30 min.

After the training the participants were tested with interrupted speech with and without filler noise in the silent intervals, with or without the video of the speaker presented simultaneously. This resulted in 4 conditions, tested with 2 sets of 13 sentences per condition (104 unique sentences). The order of the conditions was random. The participants were tested in approximately 1.5 h, including one short break.

2.4. Results

Fig. 2 shows the percent correct scores of Experiment 1. The left and right panels show the intelligibility scores without (black symbols) and with (gray symbols) the video of the speaker, respectively. The open symbols represent the percent correct scores of interrupted speech with silent intervals (S) and the filled symbols represent the interrupted speech with filler noise (N).

A two-way repeated measures ANOVA, with the addition of filler noise in the silent intervals and the accompanying video of the speaker as within-subjects factors, was performed. This analysis shows that there was a phonemic restoration benefit from adding filler noise both when the video of the speaker was absent or present. Further, the statistical analysis shows that the restoration benefit of adding filler noise was significant ($F(1,11) = 29.7$, $p < .001$, power = .99) and that the scores with video were significantly higher than without video ($F(1,11) = 57.5$, $p < .001$, power = 1.00). There was no significant interaction effect.

2.5. Discussion Experiment 1

2.5.1. The effect of visual cues on the perception of interrupted speech

We had hypothesized that the overall intelligibility of interrupted sentences with and without filler noise would increase with the addition of visual speech cues (lip-reading). The effects of these additional visual speech cues were analyzed and, confirming our first hypothesis, statistical analysis shows that the perception of interrupted speech improved significantly from 70.1% to 81.4% when the visual cues were added in the S condition

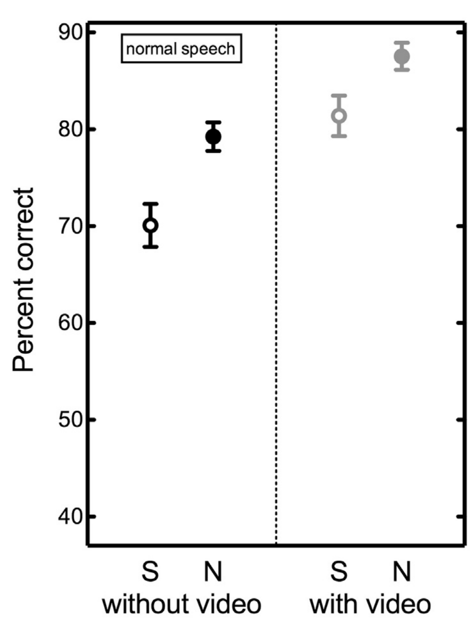

Fig. 2. Intelligibility scores of interrupted speech, with and without filler noise and video of the speaker. The mean percent correct scores of interrupted speech without further spectrotemporal degradations (Experiment 1) are shown. The open and filled symbols represent the intelligibility scores without (S) and with (N) filler noise in the silent intervals, respectively. The black and gray symbols represent the speech scores without and with the accompanying video of the speaker, respectively.

(see Table 1; Experiment 1). For the N condition, we observed a significant increase as well, from 79.2% to 87.5%, once visual speech cues were added. The additive nature of these visual speech cues is confirmed by the statistical analysis, showing no statistically significant interaction effect between providing the video of the speaker and the addition of filler noise. The increase in overall intelligibility, because of adding visual cues, is in agreement with earlier studies on perception of speech without temporal interruptions (Başkent and Bazo, 2011; Gilbert et al., 2012; McGrath and Summerfield, 1985; Ross et al., 2007). For example, Grant and Seitz (2000) showed a small improvement in speech intelligibility when visual and auditory stimuli presented the same highly redundant information. Furthermore, the shape of the mouth and lip movements enhance auditory perception of (Dutch) vowels, especially in background noise (Valkenier et al., 2012), and it aids in extracting the place of articulation from the visual modulation of the area between the lips (Grant and Seitz, 2000). Visual speech cues facilitate speech segmentation (Mitchel and Weiss, 2013) and it can enhance the capacity of auditory cortex of the listeners to track the temporal speech envelopes (Cunillera et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). In interrupted speech, Shahin and Miller (2009) concluded in their study that congruent visual speech cues resulted in a stronger restoration effect over longer white noise intervals compared to a static picture or incongruent video of the speaker. A possible explanation of the mechanism behind increased restoration from visual speech cues is that they might disengage neural processes associated with the interfering filler noise. This might suppress the response to the onsets or offsets of the filler noise, and with strengthened speech cues due to the visual cues, result in a stronger percept of a continuous speech envelope, all contributing to an increase in speech intelligibility (Bhat et al., 2014; Shahin et al., 2012).

The intelligibility of separate words of the presented interrupted sentences may depend on the alignments of the interruptions with the words. Silent intervals are common in natural speech to mark phrase boundaries or sentence endings (Sivonen et al., 2006a). The interval length used in this study was 333 ms, which is too long to fall in a legal silent position in the sentence. Each sentence was interrupted up to 4 times, depending on the length of the sentences (Benard and Başkent, 2014). Apart from this, two sets of 13 sentences were used per test condition, therefore, any possible effect of the alignment of the silent or noise filled intervals with word boundaries would likely be randomized or mitigated. The location of the interruption in the word may affect the continuity illusion. For example, the continuity of a sentence might be distorted more when a silent interruption makes one or more words (partially) inaudible, hence, reducing the intelligibility. On the other hand, in sentence context a noise filled interval masking only the initial phoneme is shown not to affect the phonemic restoration (Sivonen et al., 2006a,b), although it might affect word parsing. A noise filled interval can fall within a word facilitating perception of a continuous temporal envelope and thus improving the intelligibility.

2.5.2. The effect of visual cues on the restoration benefit

In this experiment, we had also hypothesized that additional visual speech cues would enhance the restoration benefit of filler noise for sentences with multiple periodic interruptions. In line with previous studies on phonemic restoration of interrupted sentences, the present data show a significant restoration benefit when the silent intervals are filled with noise and visual cues are absent (Powers and Speaks, 1973; Verschuure and Brocaar, 1983; Bashford et al., 1992; Başkent, 2012; Benard and Başkent, 2013). The mean percent correct scores increased from 70.1% to 79.2%, for

Table 1
Results are shown for the percent correct (PC) scores of the intelligibility measurements of participants tested with temporally interrupted sentences without spectrotemporal degradations (Experiment 1), participants tested with noiseband vocoded interrupted sentences (Experiment 2), and participants tested with simulations of electric-acoustic stimulation of interrupted speech (Experiment 3). The two left and two right columns show for the three groups the PC scores and the phonemic restoration benefit of interrupted speech without and with an accompanying video of the speaker, respectively.

Experiment	PC scores without video (%)		Phonemic restoration benefit (%)	PC scores with video (%)		Phonemic restoration benefit (%)
	S	N	$N_{without} - S_{without}$	S	N	$N_{with} - S_{with}$
1	70.1	79.2	9.2	81.4	87.5	6.2
2	40.3	43.7	3.5	60.0	57.8	−2.2
3	49.8	61.5	11.7	65.8	78.7	12.9

the S and N conditions, respectively. When the video of the speaker accompanied the audio fragments, an increase in intelligibility from 81.4% to 87.5% was observed for the S and N conditions, respectively. However, the second hypothesis was not supported by the presented data; we observed an increased phonemic restoration benefit of the filler noise. This is in contrast to Shahin and Miller (2009), who did observe a stronger restoration of a single interruption in words with addition of visual cues. The difference in findings of the two studies could be due to the study design. An fMRI study of Shahin et al. (2009) suggests that two separate neural mechanisms are active in continuity assessment (as in the study of Shahin and Miller, 2009) and in phonemic restoration (as in the present study). Participants in the study of Shahin and Miller (2009) had to report if the words were perceived continuous (phonemic restoration illusion) or interrupted (illusion failure), and all words were identifiable and unambiguous. The task in present study was to only recall what the participants heard, with no requirement to report whether the spoken sentences were perceived continuous or interrupted. Secondly, as compared to use of words as stimuli by Shahin and Miller (2009), the present study used sentences that were meaningful and provided a rich context, which may have contributed to stronger phonemic restoration observed in this study.

Phonemic restoration of multiple interruptions in sentences seems to first rely on successful separation of speech segments from noise segments (as was suggested by Bhargava et al., 2014), then applying linguistic knowledge, and using linguistic context, to activate the correct lexical decision (Srinivasan and Wang, 2005). Further, while the silent interruptions possibly introduce false speech cues, resulting in wrong lexical possibilities, filling the silent intervals with noise leads to both perception of continuous sentences, as well as increased ambiguity, which likely increases the possibility of potentially accurate lexical candidates. In support of this idea, Mattys et al. (2012) argue in their review that the probability to restore the right word is highest when there is ambiguity in the speech signal; linguistic context will be most effective when the signal is degraded to intermediate intelligibility (Boothroyd and Nittrouer, 1988). Hence, the phonemic restoration conditions of Experiment 1 are rather ideal, and without further spectrotemporal degradations, and even without the visual cues, they likely provided sufficient auditory speech cues that listeners could separate easily from the noise segments and apply linguistic context, for a robust restoration effect (Benard et al., 2014). As a result, visual cues only had an additive effect, which resulted in overall intelligibility increase, but did not necessarily help further with specific restoration mechanisms. We hypothesize that participants would benefit more from visual cues to extract and correctly identify the speech cues, and to apply linguistic context, in spectrotemporally degraded speech as happens in Experiments 2 and 3, where separating speech segments from that of noise will be particularly difficult. Visual cues, then, could be expected to contribute more to separating speech from noise.

3. Experiment 2: Interrupted speech, spectrotemporally degraded with CI simulations

3.1. Rationale

CI users can achieve substantial benefit from combining speech information from auditory and visual modalities relative to the auditory modality alone. This implies that the visual speech cues can effectively supplement with speech cues that are not well transmitted via the spectrotemporally degraded CI signal (Strelnikov et al., 2009). In the case of phonemic restoration, previous research has shown no benefit of the filler noise in acoustic CI simulations of interrupted speech presented to NH listeners (Başkent, 2012; Benard and Başkent, 2014; Bhargava et al., 2014; Chatterjee et al., 2010). Facilitating speech segmentation and temporal envelope tracking could help the listeners to better segregate speech cues from noise segments (Mitchel and Weiss, 2013; Cunillera et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). In Experiment 2, we explored this influence with interrupted noiseband vocoded speech, a degradation that captures some of the CI speech processing (acoustic simulation based on Shannon et al., 1995), with or without the filler noise in the gaps. The aim was to answer the research question whether or not a phonemic restoration benefit could be induced or enhanced when the video of the speaker accompanies the spectrotemporally degraded sentences with interruptions.

3.2. Materials and methods

3.2.1. Participants

Twelve native speakers of Dutch (6 women), aged between 18 and 26 years (mean age = 21.8 years, standard deviation (SD) = 2.5 years), participated in this experiment. The inclusion criteria were the same as Experiment 1.

3.2.2. Stimuli

The same sentences of Experiment 1 were used, except that they were further degraded with an acoustic CI simulation, before the interruptions were applied. The simulation was implemented by means of a noiseband vocoder (Başkent and Chatterjee, 2010; Başkent, 2012; Benard and Başkent, 2014; Bhargava et al., 2014; Chatterjee et al., 2010; Shannon et al., 1995). The signal processing started with dividing the spectrum of the speech material (limited to 150–7000 Hz) into eight channels by means of 6th order Butterworth band-pass filters. The amplitude envelopes were extracted from the bands by means of half-wave rectification, followed by a 3rd order low-pass Butterworth filter with cut-off frequency of 160 Hz. Multiplying the envelopes with white noise bands, produced in a similar manner as the envelopes, results in eight noise carrier bands with center frequencies based on Greenwood's frequency-position function of equally spaced distances of the basilar membrane in the cochlea (Greenwood, 1990).

The eight amplitude modulated noise carrier bands were subsequently added together, forming the acoustic CI simulation.

3.3. Procedure

Benard and Başkent (2014) showed stabilization of perceptual learning of CI simulations of interrupted speech after intensive training. Based on these results, and given that the stimuli of Experiment 2 were more degraded than that of Experiment 1, the participants of Experiment 2 were trained more intensively with feedback than those of Experiment 1. Training was given with six sets of CI simulations of interrupted sentences (78 unique sentences in total) with and without filler noise. This training session took around 1.5 h, including 2 short breaks. Participants were not trained with the video of the speaker and no sentence was presented more than once to the participants.

After the training, the participants were tested with CI simulated interrupted speech with and without filler noise in the silent intervals, with or without the video of the speaker simultaneously presented. The procedure of the rest of the experiment was the same as Experiment 1.

3.4. Results

Fig. 3 shows the percent correct scores of Experiment 2. Similar to Fig. 2, the left and right panels show the intelligibility scores without (black symbols) and with (gray symbols) the video of the speaker, respectively, and the open and closed symbols represent the percent correct scores of S and N conditions.

A two-way repeated measures ANOVA, with the addition of filler noise in the silent intervals and the accompanying video of the speaker as within-subjects factors, shows that the PC scores with video were significantly higher than without video ($F(1,11) = 66.3$, $p < .001$, power = 1.00), but there was no significant phonemic restoration benefit (3.5% without video vs. −2.2% with video; $F(1,11) = .079$, $p = .784$, power = .058). There was a significant interaction effect between the addition of the filler noise and the addition of the video ($F(1,11) = 11.5$, $p = .006$, power = .869).

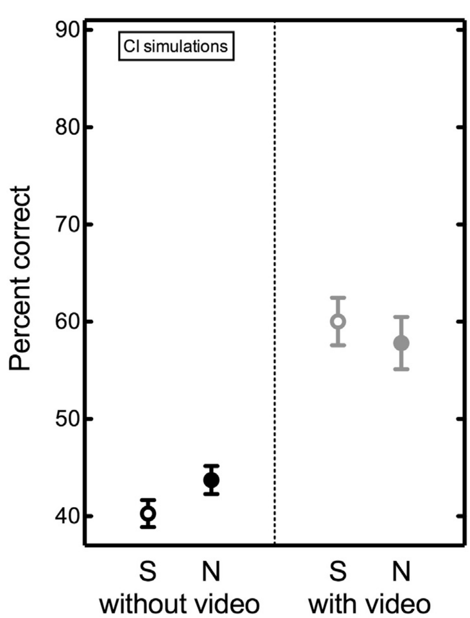

Fig. 3. Intelligibility scores of CI simulations of interrupted speech, with and without filler noise and video of the speaker. Similar to Fig. 2, except the scores shown are the result of CI simulated interrupted speech by means of a eight channel noiseband vocoder (Experiment 2).

3.5. Discussion Experiment 2

3.5.1. The effect of visual cues on the perception of interrupted speech with CI simulations

CI users commonly complain about difficulties in perceiving speech in complex listening situations. A contributing factor might be that the speech signal transmitted to the auditory nerve via electric stimulation does not contain the necessary speech cues to induce top-down restoration mechanisms (Başkent, 2012; Benard and Başkent, 2014). Bhargava et al. (2014) observed a restoration benefit in CI users, but only in specific conditions where speech segments were made longer than the silent/noise-filled interruptions. This implies that CI users could benefit from phonemic restoration if the degraded speech signal contains, or is enriched with, the necessary speech cues, which could also be provided by the additional visual speech cues. Compared to Experiment 1, we observed a drastic decrease in speech intelligibility once the CI simulations were added and the sentences were interrupted afterwards, from 70.1% to 40.3% and from 79.2% to 43.7% for without and with filler noise, respectively (Table 1; Experiment 2). These scores are comparable to the results of Bhargava et al. (2014) and Benard and Başkent (2014).

Similar to the no-simulation case (Experiment 1), the intelligibility improved statistically significantly when the visual cues were added, by 19.8% and 14.1% for without and with filler noise, respectively (Table 1; Experiment 2). CI users benefit from auditory and visual integration, such as lip-reading, in complex listening situations (e.g. Doucet et al., 2006; Lyness et al., 2013; Rouger et al., 2007; Song et al., 2014). From this experiment we can conclude that the accompanying video of the speaker is also beneficial for the perception of spectrotemporally degraded speech. Interrupted speech perception is shown to be difficult to comprehend for CI users (Bhargava et al., 2014; Nelson and Jin, 2004). Therefore, our finding is good news for CI users, since visual cues are often available in daily speech communication and can thus aid in understanding speech with interruptions.

3.5.2. The effect of visual cues on the restoration benefit of interrupted speech with CI simulations

We investigated if a restoration benefit of the filler noise could be induced once visual speech cues are additionally provided with interrupted sentences further degraded spectrotemporally, as it can happen in CI speech transmission. Regarding the restoration benefit, noiseband vocoding has a detrimental effect; commonly no restoration benefit of the filler noise is observed in acoustic CI simulations presented to NH listeners (Başkent, 2012; Benard and Başkent, 2014; Bhargava et al., 2014; Chatterjee et al., 2010). Surprisingly, however, even the accompanying video of the speaker was not able to provide the necessary cues to the perceptual system to distinguish between speech and filler noise in the case of noiseband vocoding; the mean percent correct scores did not increase statistically significantly for both without (from 40.3% to 43.7%) and with the video of the speaker (from 60.0% to 57.8%) for the S and N conditions, respectively. The absence of the restoration benefit of the filler noise implies that, while the manipulated sentences are of sufficiently good acoustic quality to lead to rather high intelligibility scores, they still do not contain the specific speech cues necessary to induce top-down repair mechanisms (Başkent, 2012; Benard and Başkent, 2014; Srinivasan and Wang, 2005). From previous studies we know that listeners are able to extract the temporal speech envelopes from speech-modulated noise fillers in interrupted speech when visual speech cues are absent (Shinn-Cunningham and Wang, 2008). Further, listeners extract the temporal speech envelope cues by observing the movement of face and the lips of the talker, which increases the intelligibility (Cunillera et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). Combining these results, we hypothesize that phonemic restoration benefit of the filler noise is prevented to occur with noiseband vocoded speech stimuli because listeners integrate the filler noise with the noisy speech segments while watching the video of the speaker. A possible explanation for the absence of phonemic restoration in this case is provided by Bhargava et al. (2014); if speech sounds like noise because of the nature of the CI processing (noiseband vocoder), the auditory system might not be able to distinguish between speech and filler noise, hence, perceiving parts of the filler noise wrongfully as speech cues. This could lead to the activation of the incorrect lexical candidates (Bhargava et al., 2014; Srinivasan and Wang, 2005) leading to no additional benefit of the filler noise. Therefore, other types of CI simulations, such as sine-wave vocoding, would probably result in similar findings, because the speech segments are processed similarly as the interrupting noise, and both speech and filler noise would sound tonal.

4. Experiment 3: Interrupted speech, spectrotemporally degraded with EAS simulations

4.1. Rationale

CI users seem to benefit less from restoration mechanisms in complex listening situations, possibly because the speech signal transmitted does not contain the necessary speech cues, like spectral resolution and temporal fine structure, to induce top-down repair mechanisms (Başkent, 2012; Benard and Başkent, 2014; Bhargava et al., 2014). Previously Başkent and Chatterjee (2010) had shown that the overall intelligibility of interrupted sentences that are spectrotemporally degraded was enhanced after low-frequency speech cues were added to the CI simulation (simulation of EAS). Listeners might be able to take advantage of the available additional speech cues, like F0 voicing (Brown and Bacon, 2009a) and low-frequency segmental phonetic cues like F1 and formant transition cues (Incerti et al., 2013; Kong and Carlyon, 2007). The aim of Experiment 3 was, firstly, to explore if the perception of interrupted speech would be further enhanced with the addition of visual cues. We presumed that these cues are complementary to the available (low-frequency) speech cues, thus further enhancement of the intelligibility was expected from the visual speech cues. Secondly, we investigated the influence of the accompanying video of the speaker on the phonemic restoration benefit of the filler noise.

4.2. Materials and methods

4.2.1. Participants

Twelve native speakers of Dutch (8 women) with normal hearing and normal speech and language development participated in this experiment. The participants were aged between 18 and 26 years (mean age = 22.1 years, standard deviation (SD) = 2.6 years). The inclusion criteria were the same as Experiments 1 and 2.

4.2.2. Stimuli

The acoustic simulation of the EAS was implemented based on the CI simulation described in Experiment 2. The two lowest channels of the noiseband vocoder (150–510 Hz) were replaced with low-pass filtered speech (3rd order Butterworth filter, cut-off frequency 500 Hz; based on Başkent, 2012), resulting in a six channel CI simulation combined with additional unprocessed low-frequency speech, producing the simulation of EAS.

4.3. Procedure

The participants of Experiment 3 were trained and tested with the same protocol as Experiment 2, except with interrupted EAS simulations of sentences.

4.4. Results

Fig. 4 shows the percent correct scores of Experiment 3, presented similarly to Fig. 3.

A two-way repeated measures ANOVA, with the addition of filler noise in the silent intervals and the accompanying video of the speaker as within-subjects factors, shows a statistically significant benefit of the filler noise (phonemic restoration benefit of 11.7% without video vs. 12.9% with video; $F(1,11) = 28.8$, $p < .001$, power = .998) and an overall significant increase in intelligibility with the accompanying video of the speaker ($F(1,11) = 47.8$, $p < .001$, power = 1.00). There was no significant interaction effect.

4.5. Discussion Experiment 3

4.5.1. The effect of visual cues on the perception of interrupted speech with EAS simulations

Developments in the design of CI electrodes and improved surgical techniques have resulted in conservation of residual hearing in the implanted ear (Miranda et al., 2014). In general, candidacy requirements for cochlear implantation have gradually loosened (Sampaio et al., 2011), which allows more CI users to have acoustic hearing in the non-implanted ear. Acoustical amplification of the available low frequency hearing, as happens in EAS, can effectively be utilized resulting in better speech intelligibility since listeners benefit from the additional speech cues (Brown and Bacon, 2009a; Kong and Carlyon, 2007). Compared to the simulations of electric stimulation alone (Experiment 2), we observed an

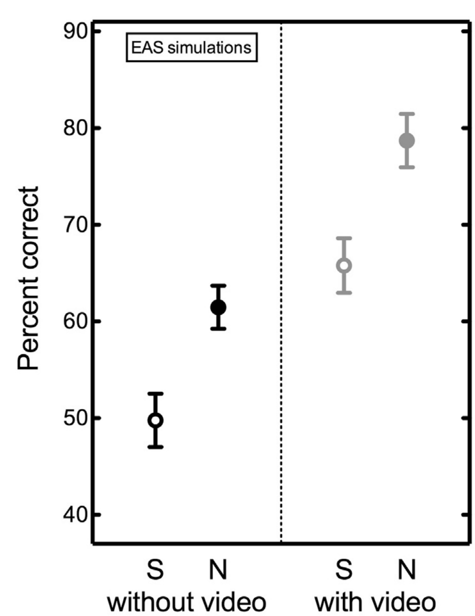

Fig. 4. Intelligibility scores of EAS simulations of interrupted speech, with and without filler noise and video of the speaker. Similar to Fig. 2, except the scores shown are the result of simulations of electric-acoustic stimulation of interrupted speech (Experiment 3).

increase in percent correct scores when low-frequency acoustic speech cues are available, from 40.3 to 49.8% in the S condition, and from 43.7 to 61.5% in the N condition, respectively (Table 1; Experiment 3). This increase has been shown in literature earlier in EAS simulations of uninterrupted (Qin and Oxenham, 2006; Zhang et al., 2012) and interrupted speech (Başkent, 2012). The intelligibility improves in the present data from 60.0 to 65.8% and from 57.8 to 78.7% for the S and N conditions, respectively, when besides the low frequency speech cues the video of the speaker was added. This statistically significant increase by adding visual speech cues is also observed in the case of acoustic CI simulation only (Experiment 2) and is in line with literature (e.g. Doucet et al., 2006; Lyness et al., 2013; Rouger et al., 2007; Song et al., 2014). It can be concluded that the auditory system takes advantage of the additional strong voicing information cues available in the EAS simulations (Incerti et al., 2013; Kong and Carlyon, 2007).

4.5.2. The effect of visual cues on the restoration benefit of interrupted speech with EAS simulations

As opposed to CI simulations without additional low-frequency speech cues, filler noise in the silent intervals of interrupted speech increased the intelligibility statistically significantly in the EAS simulation (Experiment 3). This held true both for the condition without the video of the speaker (percent correct scores of 49.8% and 61.5% for the S and N condition, respectively), and with the video of the speaker (65.8% and 78.7% for the S and the N condition, respectively). The additive nature of the visual cues was confirmed by the statistical analysis showing, in contrast with Experiment 2, no statistically significant interaction effect between providing the video of the speaker and the addition of filler noise. This increase in phonemic restoration benefit for EAS simulations, compared to CI simulations only, can be explained by the additional strong voicing information cues (Brown and Bacon, 2009b; Incerti et al., 2013; Kong and Carlyon, 2007; Zhang et al., 2010).

Bhargava et al. (2014) argued that phonemic restoration is effective if listeners are able to distinguish the speech segments from the filler noise, and that interruptions in speech possibly introduce false speech cues, which in turn can result in wrong lexical possibilities. In contrast to spectrotemporally degraded speech, the voicing information might provide the acoustical cues for EAS simulations to distinguish between speech segments and the filler noise, increasing the probability to guess the right word (Bhargava et al., 2014; Srinivasan and Wang, 2005), inducing a phonemic restoration benefit. Some studies have previously explored the role of F0 on phonemic restoration directly. Clarke et al. (2014) changed in interrupted speech without further spectrotemporally degradations the F0 from one speech segment to another. This resulted in voice discontinuity, reducing the overall intelligibility of the interrupted speech. Disrupting the F0, however, did not have an effect on the phonemic restoration benefit of the sentences presented to NH listeners. They concluded that, in the absence of further degradations on speech cues, listeners can still rely on the linguistic context for effective restoration, and ignore the inconsistent voicing cues. CI users rely heavily on the F0, a part of the voicing information (Fuller et al., 2014). Hence, our results, combined with those of Başkent (2012), indicate that the voicing provided by the low-frequency speech, could be an important cue for restoring interrupted speech that is spectrotemporally degraded, i.e., speech with weak voicing cues.

5. General discussion and conclusions

Adding visual cues in the form of a video of the speaker leads to significantly higher speech intelligibility of interrupted sentences with and without additional filler noise in the silent intervals, confirming our first hypothesis (Experiment 1). This increase in intelligibility from adding the visual cues holds also true for the spectrotemporally-degraded speech as the acoustic simulation of a CI (Experiment 2) or EAS (Experiment 3). A possible explanation of this general intelligibility increase is that, since the acoustic speech information is not available during the segments of silence or noise, the video of the speaker provides additional speech information during these intervals in effect providing the additional speech information needed to increase the intelligibility of the interrupted sentences. The increase in intelligibility was relatively more when the video of the speaker accompanied the acoustic CI and EAS simulations. These results imply that when the spectral resolution and temporal fine structure are reduced, listeners may rely more on the visual speech cues (Boothroyd and Nittrouer, 1988; Mattys et al., 2012). Visual cues are often available in daily speech communication and can thus especially aid CI users to improve speech perception (Lyness et al., 2013; Rouger et al., 2007).

Visual speech cues (lip-reading) enable the detection of the place of articulation (Grant and Seitz, 2000), facilitate speech segmentation (Mitchel and Weiss, 2013) and they can enhance the capacity to track the temporal speech envelopes (Cunillera et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). Therefore, we expected that the positive effect of visual speech cues on the intelligibility of interrupted speech would also extend to stronger restoration effects of filler noise. In general, the absence or presence of phonemic restoration is shown to be not affected by a number of manipulations; changing the F0 per speech segment (Clarke et al., 2014), training intensively with interrupted sentences with feedback without (Benard and Başkent, 2013) or with additional spectrotemporally degradations (Benard and Başkent, 2014) or slowing down the speed of the speech conserving the pitch (Benard et al., 2014; Saija et al., 2013) did not affect the phonemic restoration benefit of filler noise. Overall, these studies imply that listeners tend to separate speech segments from noise segments and rely mostly on linguistic knowledge and context to restore the audible segments into a meaningful sentence, while ignoring inconsistent auditory cues. Furthermore, specifically for this study, even when visual speech cues were added, the phonemic restoration benefit of filler noise was not affected by the additional visual cues. In Experiments 1 and 3, the phonemic restoration benefit of filler noise was present without the visual cues, and the accompanying video did not increase the restoration benefit of filler noise. In Experiment 2, the phonemic restoration benefit of filler noise was absent and remained so with the addition of visual speech cues. All results combined, visual speech cues increase the overall intelligibility of interrupted sentences even after they are further degraded with CI simulations, but they 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.

Acknowledgments

The authors would like to thank Margriet Bekendam, Lisette Stuifzand and Marije Sleurink for help with collecting data, Annemieke ter Harmsel, Marnix Stawicki and Sander Ubbink for recording the audiovisual sentence materials and Pranesh Bhargava for feedback on this manuscript. This study was supported by Pento Audiology Center Zwolle, VIDI grant 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), Heinsius Houbolt Foundation, Rosalind Franklin Fellowship from University Medical Center Groningen, and is part of the research program of the UMCG Otorhinolaryngology Department: Healthy Aging and Communication. The manuscript is one of the

four articles conducted by the first author in fulfillment of requirements for the Ph.D. degree at the University of Groningen, The Netherlands.

References

Bashford, J.A., Riener, K.R., Warren, R.M., 1992. Increasing the intelligibility of speech through multiple phonemic restorations. Percept. Psychophys. 51, 211–217.

Bashford, J.A., Warren, R.M., 1979. Perceptual synthesis of deleted phonemes. J. Acoust. Soc. Am. 65. S112:SS6.

Başkent, D., 2012. Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. J. Assoc. Res. Otolaryngol. 13, 683–692. http://dx.doi.org/10.1007/s10162-012-0334-3.

Başkent, D., Bazo, D., 2011. Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment. Ear Hear. http://dx.doi.org/10.1097/AUD.0b013e31820fca23.

Başkent, D., Chatterjee, M., 2010. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. Hear. Res. 270, 127–133. http://dx.doi.org/10.1016/j.heares.2010.08.011.

Başkent, D., Eiler, C.L., Edwards, B., 2009. Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. J. Acoust. Soc. Am. 125, 3995–4005. http://dx.doi.org/10.1121/1.3125329.

Benard, M.R., Başkent, D., 2013. Perceptual learning of interrupted speech. PLoS One 8, e58149. http://dx.doi.org/10.1371/journal.pone.0058149.

Benard, M.R., Başkent, D., 2014. Perceptual learning of temporally interrupted spectrally degraded speech. J. Acoust. Soc. Am. 136, 1344–1351.

Benard, M.R., Mensink, J.S., Başkent, D., 2014. Individual differences in top-down restoration of interrupted speech: links to linguistic and cognitive abilities. J. Acoust. Soc. Am. 135, EL88–94. http://dx.doi.org/10.1121/1.4862879.

Bhargava, P., Başkent, D., 2012. Effects of low-pass filtering on intelligibility of periodically interrupted speech. J. Acoust. Soc. Am. 131, EL87–EL92. http://dx.doi.org/10.1121/1.3670000.

Bhargava, P., Gaudrain, E., Başkent, D., 2014. Top-down restoration of speech in cochlear-implant users. Hear. Res. 309, 113–123. http://dx.doi.org/10.1016/j.heares.2013.12.003.

Bhat, J., Pitt, M. a, Shahin, A.J., 2014. Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. Front. Neurosci. 8, 173. http://dx.doi.org/10.3389/fnins.2014.00173.

Boothroyd, A., Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. J. Acoust. Soc. Am. 84, 101–114.

Brennan, S., Schober, M., 2001. How listeners compensate for disfluencies in spontaneous speech. J. Mem. Lang. 44, 274–296. http://dx.doi.org/10.1006/jmla.2000.2753.

Bronkhorst, A.W., Bosman, A., Smoorenburg, G.F., 1993. A model for context effects in speech recognition. J. Acoust. Soc. Am. 93, 499–509.

Brown, C.A., Bacon, S.P., 2009a. Low-frequency speech cues and simulated electric-acoustic hearing. J. Acoust. Soc. Am. 125, 1658–1665. http://dx.doi.org/10.1121/1.3068441.

Brown, C.A., Bacon, S.P., 2009b. Achieving electric-acoustic benefit with a modulated tone. Ear Hear 30, 489–493. http://dx.doi.org/10.1097/AUD.0b013e3181ab2b87.

Chatterjee, M., Peredo, F., Nelson, D., Başkent, D., 2010. Recognition of interrupted sentences under conditions of spectral degradation. J. Acoust. Soc. Am. 127, EL37–41. http://dx.doi.org/10.1121/1.3284544.

Clarke, J., Gaudrain, E., Chatterjee, M., Başkent, D., 2014. T'ain't the way you say it, it's what you say – perceptual continuity of voice and top-down restoration of speech. Hear. Res. 315, 80–87. http://dx.doi.org/10.1016/j.heares.2014.07.002.

Cunillera, T., Camara, E., Laine, M., Rodriguez-Fornells, A., 2010. Speech segmentation is facilitated by visual cues. Q. J. Exp. Psychol. (Hove) 63, 260–274. http://dx.doi.org/10.1080/17470210902888809.

Doucet, M.E., Bergeron, F., Lassonde, M., Ferron, P., Lepore, F., 2006. Cross-modal reorganization and speech perception in cochlear implant users. Brain 129, 3376–3383. http://dx.doi.org/10.1093/brain/awl264.

Fu, Q.-J., Nogaki, G., 2005. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. 6, 19–27. http://dx.doi.org/10.1007/s10162-004-5024-3.

Fuller, C., Gaudrain, E., Clarke, J., Galvin, J.J., Fu, Q.-J., Free, R.H., Baskent, D., 2014. Gender categorization is abnormal in cochlear implant users. J. Assoc. Res. Otolaryngol. 15, 1037–1048. http://dx.doi.org/10.1007/s10162-014-0483-7.

Gilbert, J.L., Lansing, C.R., Garnsey, S.M., 2012. Seeing facial motion affects auditory processing in noise. Atten. Percept. Psychophys. 74, 1761–1781. http://dx.doi.org/10.3758/s13414-012-0375-z.

Grant, K.W., Seitz, P.F., 2000. The use of visible speech cues for improving auditory detection. J. Acoust. Soc. Am. 108, 1197–1208.

Grant, K.W., Walden, B.E., Seitz, P.F., 1998. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. J. Acoust. Soc. Am. 103, 2677–2690.

Greenwood, D.D., 1990. A cochlear frequency-position function for several species-29 years later. J. Acoust. Soc. Am. 87, 2592–2605.

Incerti, P.V., Ching, T.Y.C., Cowan, R., 2013. A systematic review of electric-acoustic stimulation: device fitting ranges, outcomes, and clinical fitting practices. Trends Amplif. 17, 3–26. http://dx.doi.org/10.1177/1084713813480857.

Jin, S.H., Nelson, P.B., 2006. Speech perception in gated noise: the effects of temporal resolution. J. Acoust. Soc. Am. 119, 3097–3108. http://dx.doi.org/10.1121/1.2188688.

Kong, Y.-Y., Carlyon, R.P., 2007. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. J. Acoust. Soc. Am. 121, 3717–3727. http://dx.doi.org/10.1121/1.2717408.

Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol. 8, e1000445. http://dx.doi.org/10.1371/journal.pbio.1000445.

Lyness, C.R., Woll, B., Campbell, R., Cardin, V., 2013. How does visual language affect crossmodal plasticity and cochlear implant success? Neurosci. Biobehav. Rev. 37, 2621–2630. http://dx.doi.org/10.1016/j.neubiorev.2013.08.011.

Mattys, S., Davis, M., Bradlow, A., Scott, S.K., 2012. Speech recognition in adverse conditions: a review. Lang. Cogn. Process 27, 953–978. http://dx.doi.org/10.1080/01690965.2012.705006.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748. http://dx.doi.org/10.1038/264746a0.

McGrath, M., Summerfield, A.Q., 1985. Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. J. Acoust. Soc. Am. 77, 678–685.

Miranda, P.C., Sampaio, A.L.L., Lopes, R.A.F., Ramos Venosa, A., de Oliveira, C.A.C.P., 2014. Hearing preservation in cochlear implant surgery. Int. J. Otolaryngol. 2014, 468515. http://dx.doi.org/10.1155/2014/468515.

Mitchel, A., Weiss, D., 2013. Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. Lang. Cogn. Process 29, 771–780. http://dx.doi.org/10.1080/01690965.2013.791703.Visual.

Nelson, P.B., Jin, S.H., 2004. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 115, 2286–2294. http://dx.doi.org/10.1121/1.1703538.

Powers, G.L., Speaks, C., 1973. Intelligibility of temporally interrupted speech. J. Acoust. Soc. Am. 50, 661–667. http://dx.doi.org/10.1121/1.1977590.

Powers, G.L., Wilcox, J.C., 1977. Intelligibility of temporally interrupted speech with and without intervening noise. J. Acoust. Soc. Am. 61, 195–199.

Qin, M.K., Oxenham, A.J., 2006. Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. J. Acoust. Soc. Am. 119, 2417. http://dx.doi.org/10.1121/1.2178719.

Riecke, L., Mendelsohn, D., Schreiner, C.S., Formisano, E., 2009. The continuity illusion adapts to the auditory scene. Hear. Res. 247, 71–77. http://dx.doi.org/10.1016/j.heares.2008.10.006.

Riecke, L., Micheyl, C., Vanbussel, M., Schreiner, C.S., Mendelsohn, D., Formisano, E., 2011. Recalibration of the auditory continuity illusion: sensory and decisional effects. Hear. Res. 277, 152–162. http://dx.doi.org/10.1016/j.heares.2011.01.013.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., Foxe, J.J., 2007. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb. Cortex 17, 1147–1153. http://dx.doi.org/10.1093/cercor/bhl024.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proc. Natl. Acad. Sci. U. S. A. 104, 7295–7300. http://dx.doi.org/10.1073/pnas.0609419104.

Saija, J.D., Akyürek, E.G., Andringa, T.C., Başkent, D., 2013. Perceptual restoration of degraded speech is preserved with advancing age. J. Assoc. Res. Otolaryngol. http://dx.doi.org/10.1007/s10162-013-0422-z.

Sampaio, A.L.L., Araújo, M.F.S., Oliveira, C. a C.P., 2011. New criteria of indication and selection of patients to cochlear implant. Int. J. Otolaryngol. 2011, 573968. http://dx.doi.org/10.1155/2011/573968.

Shahin, A.J., Bishop, C.W., Miller, L.M., 2009. NeuroImage neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44, 1133–1143. http://dx.doi.org/10.1016/j.neuroimage.2008.09.045.

Shahin, A.J., Kerlin, J.R., Bhat, J., Miller, L.M., 2012. Neural restoration of degraded audiovisual speech. Neuroimage 60, 530–538. http://dx.doi.org/10.1016/j.neuroimage.2011.11.097.

Shahin, A.J., Miller, L.M., 2009. Multisensory integration enhances phonemic restoration. J. Acoust. Soc. Am. 125, 1744–1750. http://dx.doi.org/10.1121/1.3075576.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Shinn-Cunningham, B.G., Wang, D., 2008. Influences of auditory object formation on phonemic restoration. J. Acoust. Soc. Am. 123, 295–301. http://dx.doi.org/10.1121/1.2804701.

Sivonen, P., Maess, B., Friederici, A.D., 2006a. Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. Neurosci. Lett. 408, 220–225. http://dx.doi.org/10.1016/j.neulet.2006.09.001.

Sivonen, P., Maess, B., Lattner, S., Friederici, A.D., 2006b. Phonemic restoration in a sentence context: evidence from early and late ERP effects. Brain Res. 1121, 177–189. http://dx.doi.org/10.1016/j.brainres.2006.08.123.

Song, J.-J., Lee, H.-J., Kang, H., Lee, D.S., Chang, S.O., Oh, S.H., 2014. Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. Brain Struct. Funct. http://dx.doi.org/10.1007/s00429-013-0704-6 ePub.

Srinivasan, S., Wang, D., 2005. A schema-based model for phonemic restoration. Speech Commun. 45, 63–87. http://dx.doi.org/10.1016/j.specom.2004.09.002.

Strelnikov, K., Rouger, J., Barone, P., Deguine, O., 2009. Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf

adults with cochlear implants. Scand. J. Psychol. 50, 437–444. http://dx.doi.org/10.1111/j.1467-9450.2009.00741.x.
Trout, J., Poser, W., 1990. auditory and visual influences on phonemic restoration. Lang. Speech 33, 121–135. http://dx.doi.org/10.1177/002383099003300202.
Valkenier, B., Duyne, J., Andringa, T.C., Başkent, D., 2012. Audiovisual perception of congruent and incongruent Dutch front vowels. J. Speech. Lang. Hear. Res. 55, 1788–1801. http://dx.doi.org/10.1044/1092-4388(2012/11-0227)b.
Verschuure, J., Brocaar, M.P., 1983. Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Percept. Psychophys. 33, 232–240.
Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107, 1671–1684.
Wang, X., Humes, L., 2010. Factors influencing recognition of interrupted speech. J. Acoust. Soc. Am. 128, 2100–2111. http://dx.doi.org/10.1121/1.3483733.
Warren, R.M., 1970. Perceptual restoration of missing speech sounds. Science 167, 392–393.
Warren, R.M., 1983. Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. J. Audio Eng. Soc. 31, 623–629.
Wiersinga-Post, E., Tomaskovic, S., Slabu, L., Renken, R., de Smit, F., Duifhuis, H., 2010. Decreased BOLD responses in audiovisual processing. Neuroreport 21, 1146–1151. http://dx.doi.org/10.1097/WNR.0b013e328340cc47.
Zhang, T., Dorman, M.F., Fu, Q.-J., Spahr, A., 2012. Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. Ear Hear 33, e70–79. http://dx.doi.org/10.1097/AUD.0b013e318259e5dd.
Zhang, T., Dorman, M.F., Spahr, A., 2010. Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. Ear Hear 31, 63–69.
Zion Golumbic, E., Cogan, G.B., Schroeder, C.E., Poeppel, D., 2013. Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". J. Neurosci. 33, 1417–1426. http://dx.doi.org/10.1523/JNEUROSCI.3675-12.2013.