Perceptual learning of temporally interrupted spectrally degraded speech

Michel Ruben Benard[a)]
Pento Audiology Center Zwolle, Oosterlaan 20, 8011 GC Zwolle, The Netherlands

Deniz Başkent[b)]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

(Received 13 May 2013; revised 5 May 2014; accepted 11 July 2014)

Normal-hearing (NH) listeners make use of context, speech redundancy and top-down linguistic processes to perceptually restore inaudible or masked portions of speech. Previous research has shown poorer perception and restoration of interrupted speech in CI users and NH listeners tested with acoustic simulations of CIs. Three hypotheses were investigated: (1) training with CI simulations of interrupted sentences can teach listeners to use the high-level restoration mechanisms more effectively, (2) phonemic restoration benefit, an increase in intelligibility of interrupted sentences once its silent gaps are filled with noise, can be induced with training, and (3) perceptual learning of interrupted sentences can be reflected in clinical speech audiometry. To test these hypotheses, NH listeners were trained using periodically interrupted sentences, also spectrally degraded with a noiseband vocoder as CI simulation. Feedback was presented by displaying the sentence text and playing back both the intact and the interrupted CI simulation of the sentence. Training induced no phonemic restoration benefit, and learning was not transferred to speech audiometry measured with words. However, a significant improvement was observed in overall intelligibility of interrupted spectrally degraded sentences, with or without filler noise, suggesting possibly better use of restoration mechanisms as a result of training. © *2014 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4892756]

PACS number(s): 43.71.Es, 43.71.Rt, 43.66.Ts [PBN] Pages: 1344–1351

I. INTRODUCTION

Normal-hearing (NH) listeners use top-down repair for enhancing speech perception in complex listening situations. The benefit of such repair can be demonstrated with perception of temporally interrupted sentences with silent gaps, even when a significant proportion of speech is missing (Başkent and Chatterjee, 2010; Miller and Licklider, 1950), as well as with specific paradigms, such as phonemic restoration (Powers and Speaks, 1973; Warren, 1970). In the latter, intelligibility of interrupted sentences increases once the silent gaps are filled with loud noise bursts even though the noise does not add to the existing speech information (Powers and Wilcox, 1977; Srinivasan and Wang, 2005; Verschuure and Brocaar, 1983). Adding the filler noise creates an ambiguity where the central perceptual nervous system cannot tell if the interrupted speech signal is indeed interrupted, or just continuous and masked by intermittent noise. When faced with such ambiguity, the brain tends toward forming an object, a speech stream, from the speech segments using perceptual grouping mechanisms (Bregman, 1994). Further, such ambiguity increases the number of potential alternatives during lexical activation, increasing the chances for a better fit. Consequently, linguistic rules, prior knowledge, expectations and context are used to restore interrupted speech (Bashford *et al.*, 1992; Bronkhorst *et al.*, 1993; Sivonen *et al.*, 2006; Srinivasan and Wang, 2005), which is a highly cognitive process that involves linguistic skills of the individual (Benard and Başkent, 2013; Benard *et al.*, 2014).

Speech redundancy also plays an important role in restoration of degraded speech. In general, speech signals are highly redundant, so that the human perceptual system can still rely on remaining acoustic cues after distortions, caused in natural listening environments, that make some cues inaudible or inaccessible. Because of this, robust human communication can be achieved (Cooke *et al.*, 2001). The positive effect of such redundancy is shown by robust intelligibility of speech where spectral (Başkent and Shannon, 2006; Lippmann, 1996; Warren *et al.*, 1995) or temporal (Jin and Nelson, 2010; Miller and Licklider, 1950) segments are removed. In the case of periodic temporal interruptions, in fact, intelligibility of speech remains high for a wide range of interruption rates. However, the intelligibility of interrupted speech reduces when speech redundancy is further compromised with additional distortions, for example, as it may happen in hearing loss, where high frequency speech is inaudible (Başkent, 2010; Bhargava and Başkent, 2012). This can also happen with cochlear implants (CIs), implantable auditory prostheses for profoundly hearing impaired people, where spectral resolution is inherently degraded and

[a)]Author to whom correspondence should be addressed. Also at: University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, P.O. Box 30.001, 9700 RB Groningen, The Netherlands. Electronic mail: m.r.benard@umcg.nl
[b)]Also at: University of Groningen, Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), Groningen, The Netherlands.

temporal fine structure is not fully transmitted (Başkent and Chatterjee, 2010; Chatterjee *et al.,* 2010; Nelson and Jin, 2004). Hence, while perceptual restoration of speech is highly cognitive, the changes in the bottom-up speech signals, due to hearing loss, hearing-device processing, or hearing device-auditory nerve interaction, can also affect the top-down restoration (Başkent, 2012; Başkent *et al.,* 2009, 2010; Bhargava *et al.,* 2014). The limited help from high-level restorative processes, because the signal does not contain the necessary or appropriate speech cues to induce top-down repair mechanisms, is perhaps a contributing factor to the challenges that CI users face in understanding speech in complex listening situations with background noise (Fu and Nogaki, 2005; Stickney *et al.,* 2004).

Benard and Başkent (2013) hypothesized that more effective use of high-level perceptual mechanisms can be achieved through training. This was confirmed by the improvement observed in perception of interrupted speech, which was not degraded otherwise, after a short but intensive training was provided. In the present study, we hypothesized that interrupted speech that is additionally degraded with a CI simulation can similarly be trained. If this were the case, it would open the possibility to training actual CI users with a new approach that could possibly increase their speech intelligibility in noise. To date, there have been a number of studies for training CI users and NH listeners with CI simulations. Such auditory training so far has had a focus on training with words or sentences (Davis *et al.,* 2005; Hervais-Adelman *et al.,* 2008) or complex environmental stimuli (Loebach and Pisoni, 2008; Smalt *et al.,* 2011), with computer-based (adaptive) training programs (Fu *et al.,* 2005; Oba *et al.,* 2011; Stacey *et al.,* 2010; Zhang *et al.,* 2012), and sometimes also with additional feedback provided with visual cues (Ingvalson *et al.,* 2013), but they have not been particularly designed to engage the high-level cognitive restoration mechanisms *per se*.

In the present study, we propose a new training paradigm using degraded speech with temporal interruptions, which enforces listeners to rely on top-down repair (similar in design to Benard and Başkent, 2013). While the ultimate aim is to potentially teach CI users to use high-level repair mechanisms better and thus to improve speech perception in complex listening situations, as a first step, we started with an acoustic simulation of CIs. More specifically, we hypothesized that the perception of CI simulations of interrupted sentences can improve, despite the poor baseline intelligibility performance shown by Chatterjee *et al.* (2010) and Başkent (2012). This is a situation CI users specifically have difficulties with (Nelson and Jin, 2004). Previous studies have also shown reduced restoration benefit in acoustic CI simulations presented to NH listeners (Başkent, 2012), and different patterns in restoration in CI users than in NH listeners (Bhargava *et al.,* 2014). Therefore, we secondly hypothesized that more effective use of high-level cognitive mechanisms through training with CI processed and interrupted speech could also teach listeners to derive a restoration benefit. To test these hypotheses we trained NH listeners with interrupted speech that was also spectrally degraded with a noiseband vocoder as an acoustic simulation of CIs (Başkent, 2012; Friesen *et al.,* 2001; Shannon *et al.,* 1995). We thirdly hypothesized that the learning effect could also be reflected in a clinical speech audiometry. To test this hypothesis we used a word identification test typically used in Dutch clinics (Bosman and Smoorenburg, 1995). More specifically, we measured intelligibility of CI simulations of uninterrupted words presented in noise, before and after the training with CI simulations of interrupted sentences.

II. MATERIAL AND METHODS

A. Listeners

Twenty-four native speakers of Dutch between the ages of 18 and 23 years (mean age = 20.9 years, SD = 2.0 years, 10 women) participated in this study. They were not familiar with listening to interrupted speech in general, with the speech materials used, and with CI simulations. Normal hearing was confirmed via a hearing test [at audiometric frequencies of 0.5 kHz up to 4 kHz, hearing thresholds of 20 dB hearing level (HL) or less], and speech and language problems were further ruled out via a questionnaire. The listeners were divided into four experimental groups, matched on age.

The Medical Ethical Committee of the University Medical Center of Groningen approved this study. The listeners were recruited by poster announcements at public places and they received payment for participation. At least two weeks before the experiment they were informed about the details of the study. The written informed consent was collected prior to participation.

B. Training with sentences

1. Stimuli

Speech stimuli used for training were Dutch sentences spoken by a male speaker (Versfeld *et al.,* 2000). These sentences are meaningful, rich in context, semantically neutral, and represent conversational speech. The stimuli were digitally recorded at a sampling rate of 44.1 kHz. The corpus consists of 39 sets. Each set consists of 13 unique sentences, 4 to 9 words per sentence, with a total of 74 to 88 words in each set. The filler noise added to the silent gaps of interrupted sentences to induce restoration was a steady speech-shaped noise. This noise was generated by Versfeld *et al.* (2000), by filtering white noise with a filter that matched the long-term average speech spectrum of the recorded sentences. The sentences were presented to the participants at 60 dB sound pressure level (SPL) and the filler noise at 70 dB SPL, calibrated at the approximate position of the participant's ear (based on procedures by Başkent *et al.,* 2009; Benard and Başkent, 2013).

2. Signal processing

Temporal interruption. The sentence stimuli were periodically interrupted at a rate of 1.5 Hz by a square wave, with a duty cycle of 50% (similar to the method used by Benard and Başkent, 2013). The on and off transitions were ramped with a 10 ms raised cosine-ramp to prevent distortions from abrupt changes in the signal. This resulted in portions of the original signal interspersed by silent interruptions, each with a duration of 333 ms. The average length of the sentences is 2.47 s, which resulted in each sentence being interrupted 3 to 4 times. In this experiment we used two versions of interrupted sentences; one with silent gaps, and one with these gaps filled with the filler noise. The filler noise was produced by applying the same square-wave to the speech-shaped noise with an inverse phase. When the filler noise was added, the portions of speech and filler noise overlapped such that the midst of the speech and filler noise slopes crossed each other at every transition (see Başkent, Eiler, and Edwards 2009 for further details).

Spectral degradation. The interrupted sentences with and without filler noise were spectrally degraded using a noiseband vocoder as an acoustic simulation of CIs (Başkent, 2012; Shannon *et al.,* 1995). The noiseband vocoder was selected as the CI simulation, instead of, for example, a sinewave vocoder, as previous literature relevant to this study also used this kind of acoustic simulation of CIs (Başkent and Chatterjee, 2010; Başkent, 2012; Bhargava *et al.,* 2014). The processing parameters were also selected based on this literature. The bandwidth of the interrupted sentences was first limited to 150–7000 Hz, and then divided into eight channels by means of sixth order Butterworth band-pass filters. The cut-off frequencies of these filters were based on Greenwood's frequency-position function of equally spaced distances of the basilar membrane in the cochlea (Greenwood, 1990). This represented CI electrodes that are equally spaced in the cochlea. The envelope of each of the eight channels was extracted by means of half-wave rectification, followed by a third order low-pass Butterworth filter with the cutoff frequency of 160 Hz. White noise was processed in a similar manner, resulting in eight noise carrier bands with equal frequencies as the analysis filters. These were modulated with the corresponding envelopes in each channel and were subsequently added together to produce the CI simulated speech.

3. Procedure for sentence identification test

The sentence identification test procedure was similar to that used by Benard and Başkent (2013). The sets of sentences were selected at random for each condition and processed online using MATLAB on a Macintosh computer. The processed stimuli were directed from the digital output of an AudioFire 4 external soundcard of Echo Digital Audio Corporation (Santa Barbara, California, USA) to a Tannoy 8D Precision active near-field speaker (Coatbridge, UK) situated in an anechoic chamber. The participants were seated in this chamber, at a distance of approximately 1 m from the speaker, facing the speaker and the monitor. They listened to the audio stimulus presented from the free-field speaker, and repeated what they heard. During training, the visual feedback was presented on the monitor. The responses of the participants were recorded with a digital voice recorder, DR-100 digital by Tascam (Montebello, CA, USA), for a double check of the responses offline. The experimenter was seated outside the anechoic chamber and listened online to the presented stimuli and to the responses of the participants, via a headphone attached to the digital voice recorder. Following the participant's response, the experimenter scored the correctly repeated words using a customized MATLAB graphical user interface. The experimenter then presented the next sentence stimulus after a cue from the participant (by saying the word "next"). Participants were encouraged to guess the missing words as much as they could. The task was, hence, to "repeat all words you have heard, even if this leads to a nonsense sentence. Guess the missing words and try to complete the sentence." The MATLAB program automatically calculated the percentage of correctly identified words (the ratio between the total number of correctly repeated words and the total number of words within the sets presented) and kept log-files of the scoring of the experimenter. Twenty-six unique sentences (two sets, randomly selected) were used in each measurement before and after training, and in each training session. As a result, participants were exposed to 234 unique sentences in total.

C. Speech audiometry with words

1. Stimuli

For speech audiometry, we used a word identification test in noise (Bosman and Smoorenburg, 1995), which is typically used in Dutch clinics to assess the speech intelligibility performance. The only modification was the application of the CI simulation, but stimuli were not processed otherwise, and no interruption was applied. The words in the database were Dutch consonant-vowel-consonant (CVC) words spoken by a female speaker, and digitally recorded at a sampling rate of 44.1 kHz. The corpus consists of 16 lists. Each of the lists consists of 12 common and meaningful Dutch words with 3 phonemes each (36 phonemes in total). The lists of words were constructed by selecting an initial consonant, a vowel in the middle, and a final consonant, from three different sets of phonemes. All phonemes of a set were used only once per list and the sets were of nearly the same perceptual difficulty (Bosman, 1989). The background noise, a steady noise shaped in accordance with the long-term average spectrum of the female speaker, was provided with the original database for the purpose of performing speech-in-noise tests in speech audiometry. The words were presented to the participants at 60 dB SPL and the background noise at 60, 55, and 50 dB SPL [with signal-to-noise ratios (SNRs) of 0, 5, and 10 dB, respectively].

2. Procedure for word identification test

All groups were tested with CI simulated words-in-noise before and after the measurements with CI simulated interrupted sentences. The experimental setup and procedure was similar to the testing with sentences. The order of the different SNRs was randomly selected and presented by 24 unique words (two lists, randomly selected) per condition.

D. Overall procedure

All four experimental groups were tested with CI simulations of interrupted sentences with or without filler noise

TABLE I. Experimental procedure, shown for the silence group (SG), the noise group (NG), the silence group without feedback (SGnoF), and the noise group without feedback (NGnoF) during training sessions. "Num" represents the number of participants. "B1 & B2" and "A1 & A2" denote the measurements before and after the 5 training sessions (T1–T5), respectively. "N" and "S" denote testing with CI simulations of interrupted sentences with and without the filler noise, respectively. The SGnoF and NGnoF were tested without feedback during the training sessions, while the SG and NG were tested with feedback. The testing with CI simulations of the word identification in noise was performed before and after the measurements with interrupted sentences at multiple SNRs.

Speech stimuli	Words in noise	Interrupted sentences			Words in noise
Groups	Before (SNR in dB)	Before training B1 & B2	Training T1–T5	After training A1 & A2	After (SNR in dB)
SG, Num = 8	0, 5, 10	S & N	S, feedback	N & S	0, 5, 10
NG, Num = 8	0, 5, 10	N & S	N, feedback	S & N	0, 5, 10
SGnoF, Num = 4	0, 5, 10	S & N	S, no feedback	N & S	0, 5, 10
NGnoF, Num = 4	0, 5, 10	N & S	N, no feedback	S & N	0, 5, 10

and with CI simulated words-in-noise, before (B1 and B2) and after (A1 and A2) the five training sessions (T1–T5). The testing procedures for the four listener groups differed only during the training sessions, see Table I for more details on the experimental procedure.

The silence group (SG) was tested with interrupted sentences with silent gaps without the filler noise. The noise group (NG) was tested with interrupted sentences with the filler noise. These two groups received feedback after their verbal response during the training sessions. This feedback consisted of playing back once the uninterrupted and once the interrupted CI simulation of the sentence, while the text of the sentence was simultaneously displayed on the monitor (based on Benard and Başkent, 2013; Davis *et al.*, 2005; Hervais-Adelman *et al.*, 2008). The other two groups, the noise group no feedback (NGnoF) and the silence group no feedback (SGnoF), received no feedback during the training sessions and they served to study the effectiveness of the feedback provided. Another group to complete the design could have been a group tested with CI simulated but not interrupted sentences. However, this was practically not possible, as the baseline scores in this condition were already close to ceiling [see Bhargava *et al.* (2014) for a similar condition], leaving no room for improvement from perceptual learning.

The entire experiment was completed within one week, including the screening of the participant, the word and sentence identification measurements before and after training, and the training sessions spread over three days.

III. RESULTS

A. Training effect on intelligibility of temporally interrupted spectrally degraded sentences and phonemic restoration

Figure 1 shows the average percent correct scores for intelligibility of CI simulated and interrupted sentences as a function of testing and training sessions. In each panel, the scores of the measurements B1, B2, for before, and A1, A2, for after training, are represented in the left and right segments, respectively, and the scores of the training sessions T1–T5 in the middle segments, see Table I for more details on the experimental procedure. Either the first (B1) or the second (B2) baseline measurement before training could be the initial silence (S) or noise (N) condition, depending on the training group (see Table II).

The effect of training on intelligibility of CI simulated and interrupted sentences (hypothesis 1) was first analyzed by comparing the performance before training (B1, B2) with after training (A1, A2). This comparison shows that training increased the performance of all groups (Table II), by 10.8 to 21.6 percentage points for the groups tested with feedback, and 5.5 to 10.9 percentage points for groups tested without feedback. This effect was significant, shown by a three-factor repeated measures analysis of variance (ANOVA) run on the data presented in the measurements B1, B2, A1, and A2 (with before and after training and the addition of the filler noise as two within-subjects factors, and the addition of feedback as the between-subjects factor). The average intelligibility for the four listener groups (groups tested with interrupted speech with or without filler noise and with feedback, and groups tested with or without filler

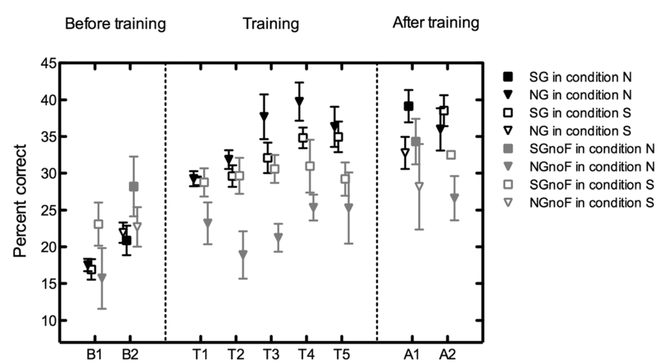

FIG. 1. Intelligibility of CI simulations of interrupted speech with and without filler noise. The mean percent correct scores are shown for the four groups of participants: the silence group with (SG) and the silence group without feedback (SGnoF) during the training sessions, the noise group with (NG) and the noise group without feedback (NGnoF) during the training sessions. The "S" (silence) and "N" (noise) conditions denote the conditions with interrupted sentences with silent intervals or combined with filler noise, respectively. The square symbols represent the groups tested with interrupted speech without filler noise (SG and SGnoF), the triangle symbols represent the groups tested with interrupted speech with filler noise (NG and NGnoF) during the training sessions. The filled symbols represent the scores of the interrupted sentences with filler noise, the open symbols without the filler noise. The black symbols with black error bars and gray symbols with gray error bars represent the groups tested with and without the feedback in the training sessions, respectively. The horizontal axis shows the measurements before (B1, B2) and after (A1, A2) the five training sessions (T1–T5). The error bars denote the standard error of the mean.

TABLE II. The mean percent correct scores before and after training. The left and middle columns show the percent correct scores before (B1, B2) and after (A1, A2) training, for the groups tested with feedback (SG, NG) and tested without feedback (SGnoF, NGnoF) during the training sessions. “N” and “S” refer to testing conditions with and without filler noise in the sentences interruptions, respectively. The right column shows the improvement in the S and in the N conditions.

	Percent correct scores before training (%)		Percent correct scores after training (%)		Improvement (%)	
Groups	B1	B2	A1	A2	S	N
SG ($n=8$)	S, 16.9	N, 20.7	N, 39.1	S, 38.5	21.6	18.3
NG ($n=8$)	N, 17.5	S, 21.9	S, 32.8	N, 36.0	10.8	18.5
SGnoF ($n=4$)	S, 23.1	N, 28.2	N, 34.3	S, 32.5	9.4	6.1
NGnoF ($n=4$)	N, 15.7	S, 22.7	S, 28.2	N, 26.6	5.5	10.9

noise and without feedback) improved significantly after the training sessions (main effect of factor before and after training, F(1,22) = 79.8, $p < 0.001$, $\eta^2 = 0.784$, power = 1). The effect of training on phonemic restoration benefit (hypothesis 2), defined as the increase in intelligibility of interrupted sentences after the silent intervals are filled with noise (see, e.g., Powers and Wilcox, 1977), was investigated by comparing the performance of S and N conditions. At the measurements before and after training (B1, B2, and A1, A2, respectively), no restoration benefit was observed, as the addition of filler noise in the silent intervals did not increase the percent correct scores significantly [main effect of addition of the filler noise, F(1,22) = 0.029, $p = 0.866$, $\eta^2 = 0.001$, power = 0.053]. The groups tested with the feedback (SG and NG) performed overall better than the groups tested without feedback (SGnoF and NGnoF; see this improvement in the right column in Table II). There was no significant difference between the groups tested with or without feedback [main effect of addition of the feedback, F(1,22) = 0.840, $p = 0.369$, $\eta^2 = 0.037$, power = 0.142]. However, there was a significant interaction between the factors before and after training and the addition of feedback during training [F(1,22) = 10.884, $p = 0.003$, power = 0.833].

The effect of training (hypotheses 1 and 2) was also analyzed by comparing performance across the training sessions T1–T5 (the middle segment of Fig. 1). The groups tested with feedback (SG and NG) performed overall better than the groups tested without feedback (SGnoF and NGnoF). A three-factor repeated measures ANOVA of these data (with the five training sessions as within-subjects factors, and the addition of the filler noise and the addition of feedback during training as the between-subjects factor) shows a significant effect of training in general on intelligibility [main effect of factor training sessions, F(4,17) = 4.772, $p = 0.009$, $\eta^2 = 0.209$, power = 0.878]. The groups tested without feedback scored significantly less than the groups tested with feedback [main effect of feedback F(1,20) = 19.8, $p < 0.001$, $\eta^2 = 0.496$, power = 0.988]. The addition of filler noise did not influence the performance [main effect of factor the addition of the filler noise, F(1,20) = 1.636, $p = 0.215$, $\eta^2 = 0.076$, power = 0.230]. There were no significant interactions between factors.

B. Speech audiometry

Figure 2 shows the results of speech audiometry. The mean percent correct scores are shown for the identification of CI simulated words presented in noise before and after the measurements with CI simulated interrupted sentences (left and right panels, respectively), as a function of SNRs and for all groups tested. The performance improved significantly with a more favorable SNR for all groups. This was confirmed with a four-factor repeated measures ANOVA, with before and after training and SNR as within-subjects factors, and the addition of filler noise during training (SG, SGnoF vs NG, NGnoF) and addition of feedback during training (SG, NG vs SGnoF, NGnoF) as the between-subjects factors [main effect of SNR, F(2,19) = 28.9, $p < 0.001$, $\eta^2 = 0.680$, power = 1]. The training with interrupted sentences did not increase the intelligibility of words in noise, as there was no main effect of training [main effect of factor before and after training, F(1,20) = 0.373, $p = 0.548$, $\eta^2 = 0.018$, power = 0.090]. The word identification by the two noise groups, NG, NGnoF, did not differ significantly from the two silence groups, SG and SGnoF [main effect of factor the addition of filler noise during training, F(1,20) = 0.332, $p = 0.571$, $\eta^2 = 0.016$, power = 0.085], and so did the word identification between the two groups tested with feedback, SG, NG, and the two groups tested without feedback, SGnoF, NGnoF [main effect of the addition of feedback, F(1,20) = 0.002,

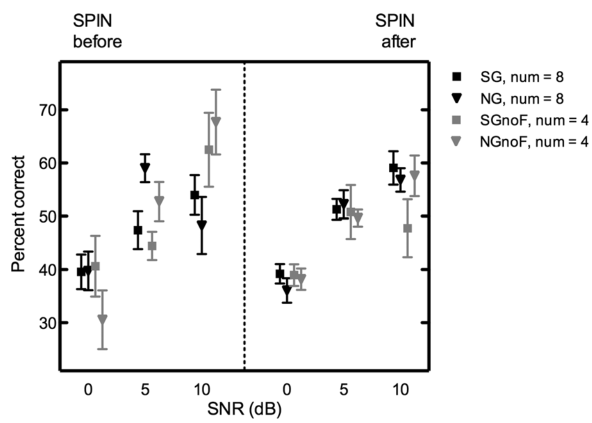

FIG. 2. The mean percent correct scores for the word-in-noise identification measurements. The mean percent correct scores shown for the word-in-noise identification measurements of the CI simulated consonant-vowel-consonant words presented in noise, for the four different listeners groups and at different SNRs. The dark and the light gray squares represent the results from the silence group with (SG) and the silence group without feedback (SGnoF) during the training sessions, respectively. The dark and the light triangles represent the results from the noise group with (NG) and noise group without feedback (NGnoF) during the training sessions, respectively. The left and right panels show the results before and after the measurements and training sessions, respectively. The error bars denote the standard error of the mean.

$p = 0.961$, $\eta^2 < 0.001$, power = 0.050]. There were no significant interactions between factors.

IV. DISCUSSION

Benard and Başkent (2013) have previously shown that intelligibility of interrupted speech can improve with intensive training, indicating more effective use of top-down repair mechanisms after training than before. This effect was shown in a young NH population and with no further degradations applied to the interrupted speech. In this study, we hypothesized that such improvement would also be observed with simulations of CI speech, which could in return potentially lead to new training programs for CI users. Such an approach, namely using distorted speech as training materials to specifically induce more active involvement from top-down mechanisms of speech perception, has not been directly used before in training programs developed for CI users (Davis *et al.*, 2005; Fu *et al.*, 2005; Hervais-Adelman *et al.*, 2008; Ingvalson *et al.*, 2013; Loebach and Pisoni, 2008; Nogaki *et al.*, 2007; Oba *et al.*, 2011; Smalt *et al.*, 2011; Stacey *et al.*, 2010; Zhang *et al.*, 2012). Given that CI users greatly suffer from difficulties understanding speech degraded in any form, either due to background noise (Fu and Nogaki, 2005; Stickney *et al.*, 2004), or due to interruptions (Bhargava *et al.*, 2014; Nelson and Jin, 2004), such training could potentially be of great help to this patient population.

Confirming our main hypothesis, NH listeners tested with CI simulations increased their performance significantly as a result of training. In the beginning, intelligibility of temporally interrupted and CI simulated sentences was very low, almost at floor level, in line with studies with interrupted speech with CI simulations or actual CI users where listeners were acutely tested (Başkent and Chatterjee, 2010; Başkent, 2012; Chatterjee *et al.*, 2010). Training increased the performance significantly, even doubling the initial baseline performance in some conditions, and perhaps more importantly, pulling the performance level further away from floor. In a real life situation, such a learning effect could have substantially positive consequences for speech communication.

In comparison to the previous study by Benard and Başkent (2013), we observed that the learning rate of the present study was somewhat faster. This can be explained on account of the unfamiliarity with the degradations imposed by both CI simulations and temporal interruptions (in contrast to interruptions alone of the previous study). The fast learning of CI simulations of interrupted speech in NH listeners might imply that CI users may potentially be taught to use the top-down cognitive and linguistic mechanisms more efficiently to enhance intelligibility of interrupted speech, for example, due to fluctuating background noise. Further research with CI users is needed to confirm these potential benefits more confidently.

Benard and Başkent (2013) had observed that repeated testing with feedback produced stronger and faster learning than repeated testing without feedback. In the present study, no such strong effect of feedback was observed when the groups tested with feedback (SG, NG) and without feedback (SGnoF, NGnoF) were directly compared; however, the statistical analysis for this comparison was also underpowered (power = 0.142). On the other hand, there was a significant interaction between the factors of speech score before and after training, addition of filler noise, and feedback (with power = 0.833), suggesting that the interpretation of these isolated effects alone might be incomplete. Given that there was no restoration benefit (i.e., no effect of adding the filler noise *per se*), this interaction may still indicate a small effect of providing feedback. In support of this idea, Loebach *et al.* (2010) previously showed that a combined visual and auditory feedback that allowed the participant to read the sentence (visual) while the degraded sentence was played back (auditory) was more efficient than presenting spectrally intact auditory feedback alone. These observations are good news for CI users, as an un-degraded auditory feedback would not be possible in their case, but a more realistic feedback with visual text display as well as playback of uninterrupted speech materials could still be useful.

A second hypothesis was that, even though it did not exist in the initial baseline conditions, a restoration benefit could appear during training. More specifically, training with the two types of interrupted speech stimuli would teach listeners to use high-level cognitive mechanisms more effectively to learn to derive a restoration benefit of filler noise (Başkent *et al.*, 2009; Benard and Başkent, 2013; Repp, 1992; Srinivasan and Wang, 2005; Verschuure and Brocaar, 1983). The initial lack of restoration benefit was in line with previous studies that used a similar configuration with CI simulations of interrupted speech presented to NH listeners (Başkent, 2012; Chatterjee *et al.*, 2010) or with actual CI users [Bhargava *et al.* (2014); only at longer speech segments some restoration benefit was observed]. The repeated measures three-factor ANOVA used (with before and after training and the addition of the filler noise as two within-subjects factors, and the addition of feedback as the between-subjects factor), showed that the application of a CI simulation to interrupted speech prevented participants to benefit from filler noise also during and after training, even though overall intelligibility performance increased. Benard and Başkent (2013) showed in their study with comparable design, but without spectrally degraded stimuli, statistically significant restoration benefits of the filler noise (effect size $f = 0.97$). Based on their study, a significant restoration benefit of the filler noise after training was *a priori* expected with the present sample size of 24 participants (expected effect size $f = 0.60$). However, a *post hoc* power analysis of the three-factor repeated measures ANOVA showed a very low effect size ($f = 0.04$), which means that only an unrealistically large number of participants (> 4000) would make a significant difference between the silence and the noise conditions after training. This suggests that the lack of an observed restoration benefit is a real and valid finding and not a result of the relatively low number of participants. As discussed by Başkent (2012) and Bhargava *et al.* (2014), a weak or non-existent restoration benefit observed in CI listeners or CI simulations implies that the top-down repair mechanisms can fail to be helpful depending on the type of degradations that occur in the speech signals. In the present study, such degradations were caused by the noiseband vocoder, and the processed speech signals were noisy due to the nature of the specific simulation method used. It is likely that in this situation, the brain attributes (parts of) the filler noise to the speech, perceiving them wrongfully as speech cues. This, in turn, could lead to the activation of the incorrect lexical candidates (Bhargava *et al.*, 2014; Srinivasan and Wang, 2005). This observation with CI simulations implies that how well a CI user can take advantage of high-level restoration mechanisms is likely highly dependent on the characteristics of the speech signal that is transmitted by their device. In support of this idea, Bhargava *et al.* (2014) observed that the CI users who performed better with their CI device for speech intelligibility in general were also more likely to show a restoration benefit.

Overall, the intensive training increased intelligibility of interrupted speech with and without filler noise, implying that listeners indeed made better use of the speech cues in the audible speech portions. However, it did not revive the restoration benefit, hinting that when the combination of degraded speech with filler noise creates misleading speech cues, these can perhaps not be overcome with training. Further research is needed to test the potential explanation that the brain interprets the filler noise as erroneously speech in noiseband vocoded simulations, and that perhaps less noisy CI simulations would yield a restoration benefit. This can be achieved, for example, by using different methods of CI simulation, such as simulating electric-acoustic stimulation (EAS) or sinewave vocoding. Sinewave vocoding can provide, for example, stronger pitch cues compared to noiseband vocoder, which can make a significant difference in pitch-related tasks such as gender identification or categorization of the speaker (Fuller *et al.*, 2014; Gonzalez and Oliver, 2005). It is not yet clear if and what effect it would have on phonemic restoration.

A third hypothesis was about speech audiometry. This test was conducted to investigate if improvements with the specific paradigm of using interrupted sentences could also be captured with standard clinical tests that use much simpler speech materials, such as words. The participants performed better with word-in-noise identification with CI simulation, as expected, in more favorable SNRs. However, the performance did not significantly increase after the intensive training with the interrupted and spectrally degraded sentences. This finding is not entirely unexpected, as, while a widely used test in clinics, the word identification in steady background noise would not be the most appropriate test to specifically explore the potential benefits from training with interrupted sentences. Teaching listeners to make use of audible speech segments would perhaps be more relevant to situations with fluctuating background noise, instead of a steady one, as in this situation listeners would have access to audible segments of speech when the noise level is low (Cooke, 2006). These could then be used to construct the message, a task similar to the one used in understanding interrupted speech. Further, for the increase in performance of the sentences the participants perhaps have learned to make better use of sentence context, which is not available in isolated words (Bronkhorst *et al.*, 1993; Grossberg and Kazerounian, 2011; Sivonen *et al.*, 2006; Verschuure and Brocaar, 1983). Because the word test was selected to represent typical speech audiometry, the results imply that if a CI training program were implemented based on the present study, a clinical word-in-noise test very likely would not reflect the learning effects from such training. A sentence identification test with a fluctuating background noise could be a more appropriate choice.

To conclude, even though restoration benefit was not revived and the learning effect was not transferred to speech audiometry, training still provided a robust and large increase (10.8 to 21.6 percentage points with feedback and 5.5 to 10.9 percentage points without) in overall intelligibility of interrupted speech with or without filler noise. Based on this strong learning effect with CI simulations in NH listeners and the fact that the CI users have to deal with interrupted speech in daily life due to non-optimal listening conditions, we propose that the perceptual learning of interrupted speech could potentially be a useful direction for further research in developing training programs for CI users. To date, previous training studies with CI users or simulations of CIs have not been designed to particularly make use of high-level restoration mechanisms (Davis *et al.*, 2005; Fu *et al.*, 2005; Hervais-Adelman *et al.*, 2008; Ingvalson *et al.*, 2013; Loebach and Pisoni, 2008; Nogaki *et al.*, 2007; Oba *et al.*, 2011; Smalt *et al.*, 2011; Stacey *et al.*, 2010; Zhang *et al.*, 2012). The significant improvement in intelligibility observed in the present study suggests that training with interrupted speech likely does that, and enforces listeners to rely on top-down repair to fill in for the inaudible speech parts to enhance intelligibility. Such a skill could be useful in the real life listening conditions where speech is commonly interrupted by dynamic background maskers and its message needs to be reconstructed for robust communication. The reduced spectral resolution and temporal fine structure in CI sound transmission can make it more difficult for CI users to use top-down mechanisms in enhancing intelligibility of interrupted speech (Başkent, 2012; Bhargava *et al.*, 2014). Therefore, improving the sound transmission in CI devices combined with effective training programs could help CI users to better understand speech in noise (Başkent, 2012; Fu and Galvin, 2003, 2008; Stacey and Summerfield, 2008).

ACKNOWLEDGMENTS

The authors would like to thank Annemieke ter Harmsel for help with collecting data, and Age Hoekstra, Peter van Hengel, and Jefta Saija for their valuable feedback on earlier versions of this manuscript. This study was supported by Pento Audiology Center Zwolle, VIDI Grant No. 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 our department: Healthy Aging and Communication.

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

Başkent, D. (**2010**). "Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores," J. Acoust. Soc. Am. **128**, EL169–EL174.

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.

Başkent, D., and Chatterjee, M. (**2010**). "Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech," Hear. Res. **270**, 127–133.

Başkent, D., Eiler, C. L., and Edwards, B. (**2009**). "Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech," J. Acoust. Soc. Am. **125**, 3995–4005.

Başkent, D., Eiler, C. L., and Edwards, B. (**2010**). "Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss," Hear. Res. **260**, 54–62.

Başkent, D., and Shannon, R. V (**2006**). "Frequency transposition around dead regions simulated with a noiseband vocoder," J. Acoust. Soc. Am. **119**, 1156–1163.

Benard, M. R., and Başkent, D. (**2013**). "Perceptual learning of interrupted speech," PLoS One **8**, e58149.

Benard, M. R., Mensink, J. S., and Başkent, D. (**2014**). "Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities," J. Acoust. Soc. Am. **135**, EL88–EL94.

Bhargava, P., and Başkent, D. (**2012**). "Effects of low-pass filtering on intelligibility of periodically interrupted speech," J. Acoust. Soc. Am. **131**, EL87–EL92.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top-down restoration of speech in cochlear-implant users," Hear. Res. **309**, 113–123.

Bosman, A. (**1989**). "Speech perception by the hearing impaired," Ph.D. thesis, University of Utrecht, The Netherlands, pp. 101–132.

Bosman, A., and Smoorenburg, G. F. (**1995**). "Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment," Int. J. Audiol. **34**, 260–284.

Bregman, A. S. (**1994**). *Auditory Scene Analysis—The Perceptual Organization of Sound* (MIT press, Cambridge, MA), pp. 345–934.

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

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). "Recognition of interrupted sentences under conditions of spectral degradation," J. Acoust. Soc. Am. **127**, EL37–EL41.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooke, M., Green, P., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Commun. **34**, 267–285.

Davis, M. H., Johnsrude, I. S., Hervais-adelman, A., Taylor, K. J., and Mcgettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Galvin, J. J. (**2003**). "The effects of short-term training for spectrally mismatched noise-band speech," J. Acoust. Soc. Am. **113**, 1065–1072.

Fu, Q.-J., and Galvin, J. J. (**2008**). "Maximizing cochlear implant patients' performance with advanced speech training procedures," Hear. Res. **242**, 198–208.

Fu, Q.-J., Galvin, J. J., Wang, X., and Nogaki, G. (**2005**). "Moderate auditory training can improve speech performance of adult cochlear implant patients," Acoust. Res. Lett. Online **6**, 106–111.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fuller, C. D., Galvin, J. J., Free, R. H., and Başkent, D. (**2014**). "Musician effect in cochlear implant simulated gender categorization," J. Acoust. Soc. Am. **135**, EL159–EL165.

Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

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

Grossberg, S., and Kazerounian, S. (**2011**). "Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise," J. Acoust. Soc. Am. **130**, 440–460.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., and Carlyon, R. P. (**2008**). "Perceptual learning of noise vocoded words: effects of feedback and lexicality," J. Exp. Psychol. Hum. Percept. Perform. **34**, 460–474.

Ingvalson, E. M., Lee, B., Fiebig, P., and Wong, P. C. M. (**2013**). "The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients," J. Speech, Lang. Hear. Res. **56**, 81–89.

Jin, S. H., and Nelson, P. B. (**2010**). "Interrupted speech perception: the effects of hearing sensitivity and frequency resolution," J. Acoust. Soc. Am. **128**, 881–889.

Lippmann, R. (**1996**). "Accurate consonant perception without mid-frequency speech energy," IEEE Trans. Speech Audio Process. **4**, 66–69.

Loebach, J. L., and Pisoni, D. B. (**2008**). "Perceptual learning of spectrally degraded speech and environmental sounds," J. Acoust. Soc. Am. **123**, 1126–1139.

Loebach, J. L., Pisoni, D. B., and Svirsky, M. A. (**2010**). "Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant," J. Exp. Psychol. Hum. Percept. Perform. **36**, 224–234.

Miller, G. A., and Licklider, J. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Nelson, P. B., and 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.

Nogaki, G., Fu, Q.-J., and Galvin, J. J. (**2007**). "Effect of training rate on recognition of spectrally shifted speech," Ear Hear. **28**, 132–140.

Oba, S. I., Fu, Q.-J., and Galvin, J. J. (**2011**). "Digit training in noise can improve cochlear implant users' speech understanding in noise," Ear Hear. **32**, 573–581.

Powers, G. L., and Speaks, C. (**1973**). "Intelligibility of temporally interrupted speech," J. Acoust. Soc. Am. **54**, 661–667.

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

Repp, B. H. (**1992**). "Perceptual restoration of a 'missing' speech sound: Auditory induction or illusion?," Percept. Psychophys. **51**, 14–32.

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

Sivonen, P., Maess, B., Lattner, S., and Friederici, A. D. (**2006**). "Phonemic restoration in a sentence context: Evidence from early and late ERP effects," Brain Res. **1121**, 177–189.

Smalt, C. J., Talavage, T. M., Pisoni, D. B., and Svirsky, M. A. (**2011**). "Neural adaptation and perceptual learning using a portable real-time cochlear implant simulator in natural environments," *IEEE EMBS 33rd Annual International Conference*, pp. 1145–1148.

Srinivasan, S., and Wang, D. (**2005**). "A schema-based model for phonemic restoration," Speech Commun. **45**, 63–87.

Stacey, P. C., Raine, C. H., O'Donoghue, G. M., Tapper, L., Twomey, T., and Summerfield, A. Q. (**2010**). "Effectiveness of computer-based auditory training for adult users of cochlear implants," Int. J. Audiol. **49**, 347–356.

Stacey, P. C., and Summerfield, A. Q. (**2008**). "Phoneme-based training strategies in improving the perception of spectrally distorted speech," Hear. Res. **51**, 526–538.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Verschuure, J., and 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., and 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.

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science. **167**, 392–393.

Warren, R. M., Riener, K. R., Bashford, J. A., and Brubaker, B. S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophys. **57**, 175–182.

Zhang, T., Dorman, M. F., Fu, Q.-J., and Spahr, A. J. (**2012**). "Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation," Ear Hear. **33**, e70–e79.