AUDITORY AND COGNITIVE MECHANISMS OF TOP-DOWN RESTORATION OF DEGRADED SPEECH

IMPLICATIONS FOR COCHLEAR IMPLANT USERS

M.R. BENARD

Auditory and cognitive mechanisms of top-down restoration of degraded speech

Implications for cochlear implant users

M.R. Benard

Cover design:	Tara Kinneging, Persoonlijk Proefschrift
Layout:	Tara Kinneging, Persoonlijk Proefschrift
Illustration page 9:	Tara Kinneging, Persoonlijk Proefschrift
Printed by:	Ipskamp Drukkers, Enschede

ISBN/EAN: 978-94-6259-792-1

rijksuniversiteit
groningen

Auditory and cognitive mechanisms of top-down restoration of degraded speech

Implications for cochlear implant users

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 28 september 2015 om 11.00 uur

door

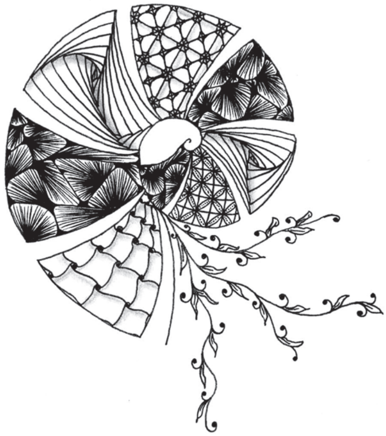
Michel Ruben Benard

geboren op 28 augustus 1982
te Krimpen aan den IJssel

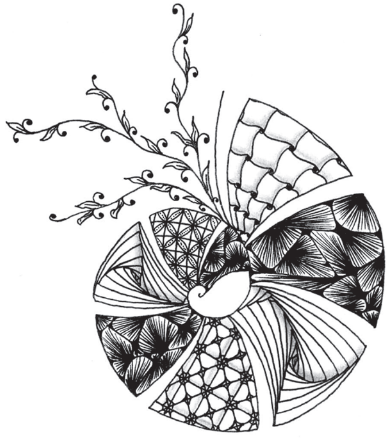
Promotores
Prof. dr. D. Başkent
Prof. dr. P. van Dijk

Beoordelingscommissie
Prof. dr. A.W. Bronkhorst
Prof. dr. P. Hendriks
Prof. dr. ir. W.A. Dreschler

CONTENTS



To my wife Lian
and to our daughters Silke and Lotte



CHAPTER 1

General introduction

INTRODUCTION

In daily life, spoken speech is one the most important ways of communication between humans. However, speech is often interrupted by other sounds that draw our attention. Imagine arriving home after a day of work; your family awaits you, food is being prepared and children are playing. Your house will be filled with natural sounds: the laughter of children, the sounds of the cooking, notifications on your smartphone, while the wind is softly blowing through an open window, bringing the sounds of birds and a car passing in the street. In this auditory scene you are having a conversation about your day's experiences. Your perceptual system is presented with a lot of acoustic streams that could be followed. There is a lot of competition and also masking of information. To retrieve the right message may be sometimes demanding, but as a normal-hearing listener you will be able to recognize the sounds of interest relatively well (Riecke et al. 2007). To follow a conversation in this complex acoustic environment your perceptual system makes use of a number of automatic restoration mechanisms.

Nevertheless, having a conversation in a noisy room is often reported to be difficult. The listener has to focus on the speaker without getting distracted by other conversations or by background noise. These challenging acoustic environments are even more problematic for hearing impaired listeners. The number of hearing impaired people is considerable (Stevens et al. 2013); an averaged hearing threshold across the frequencies 0.5, 1, 2 and 4 kHz of 35 dB HL or more in the better ear, has an estimated global prevalence of around 11% for above 15 years of age. Hearing impairment leads to a degraded auditory input to the auditory system for which has to be compensated. Hearing-impaired listeners often try to relieve the consequences of their hearing loss by relying on non-auditory speech cues, such as lip-reading. Even these compensation mechanisms and strategies do not always solve the problem hearing-impaired listeners face in everyday speech communication. As a result, hearing-impaired listeners report often that they tend to avoid social situations with background noise.

The **hope** that is driving the research presented in this thesis is that hearing-impaired listeners could learn to communicate better in complex acoustic environments, if we could provide an effective and adequate training for them.

The main **aim** is to understand more about the auditory and cognitive factors that are influencing, or are involved in, perceptually restoring degraded speech in such a way that the perceptual system can form it into an intelligible speech stream. This knowledge can lead to new training programs for the hearing impaired.

The reason why normal-hearing listeners are able to have a conversation in background noise is that the perceptual system uses multiple high-level cognitive mechanisms, in which many factors are involved. For example, as soon as the listeners get familiar with a speaker they benefit from the pitch and timbre of his/her voice, from the known direction of the sound and from the speech rate. Using these acoustic properties, *perceptual grouping* is used to find the sounds patterns pertaining to the target speech produced by the speaker of interest, which are then used to form a coherent speech stream. *Segmentation* is used to recognize individual words to form meaningful lexical units (Bregman 1994; Davis and Johnsrude 2007). The perceptual system makes use of highly automated restoration mechanisms in this complex acoustic environment so that the conversation can take place without much effort and can continue seamlessly, at a fast pace. These restoration mechanisms play an important role in speech perception of degraded speech, as will be shown in this thesis.

Cognitive mechanisms are involved in restoring degraded speech streams in such a way that the brain perceives it as an intelligible speech stream. Aging has been shown to affect the perception, as well as restoration, of degraded speech, even without age-related hearing loss, implying that the cognitive system is involved (Saija et al. 2013). If indeed cognitive factors play a role in the restoration of degraded speech, training hearing-impaired listeners to make better use of these cognitive mechanisms might possibly lead to better speech perception in real-life noisy environments. Therefore, the main aim of this research was to better understand how these cognitive speech restoration mechanisms work in general. The *ultimate* goal of the present study is to contribute to a reliable, valid, and effective training method to improve speech intelligibility in complex listening situations for hearing-impaired listeners, fitted with a cochlear implant (CI), an implantable auditory prosthesis. Even though there are a number of auditory training programs available, both in research settings and commercially, these have not been specifically designed to employ cognitive restoration mechanisms in training. For these purposes, we designed and performed a number of studies, of which the main results obtained are presented in this thesis.

PHONEMIC RESTORATION

By studying the restoration capacities of the auditory system of normal hearing listeners we gained more knowledge of how auditory perception of interrupted speech works in general. These studies are interesting per se for understanding more about the auditory perception of hearing-impaired listeners. In order to find (partial) answers to the main aim of this thesis, a well-controlled auditory scene, or rather, a well controlled and well-understood auditory stimulus is employed. This stimulus is periodically interrupted sentences with and without filler noise in the silent intervals, and induces restoration. Figure 1.1 shows this stimulus schematically; silent intervals interrupt the spoken sentence in such way that only fragments of the original sentence are audible.

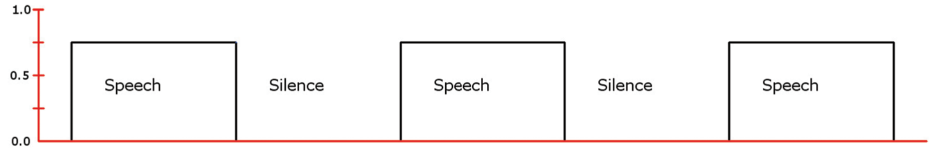


Figure 1.1: Schematic representation of interrupted speech. The original sentence is periodically interrupted with silent intervals.

Research on interrupted speech goes back to the fifties. Miller & Licklider (1950) described that an interrupted word can be perceived as continuous through noise. In their paper they described that if speech is interrupted, it is similar to seeing a landscape through a picket fence.

"It is much like seeing a landscape through a picket fence - the pickets interrupt the view at regular intervals, but the landscape is perceived as continuing behind the pickets" Miller & Licklider (1950)

This reflects the ability of the listener to perceive "glimpses" of speech between the silent interruptions or between the loud noise fillers of the manipulated speech signals. Miller and Licklider showed that words interrupted by silent intervals were less intelligible than uninterrupted speech and that interrupting with non-stationary maskers lead to better perception. They used a wide range of interruption rates in their study. Warren (1970) was the first to show that listeners believed they heard in a sentence a phoneme, which was only masked by an extraneous sound, while in reality the phoneme did not exist at all. The stimulus was a recording of the following sentence,

*"The state governors met with their respective legi*latures convening in the capital city".*

The first "s" in the word "legislatures" and the phonemes adjoining were replaced by a recorded cough. Participants were asked to encircle the exact position of the cough on the sentence written on paper. No participant identified the exact position of the cough and half of the participants circled beyond the word "legislatures". This special type of auditory induction is called phonemic restoration, and is evoked when a portion of speech is perceived as present, while it is missing in reality, and replaced by noise (Warren and Sherman 1974). Phonemic restoration was also studied with periodic temporal interruptions filled with silence or with filler noise. For example, Powers & Wilcox (1977) examined interrupted speech with several interruption rates, and found that interrupted sentences were significantly more intelligible when interruptions were filled with noise (Fig.1.2) than when left as silent (Fig.1.1).

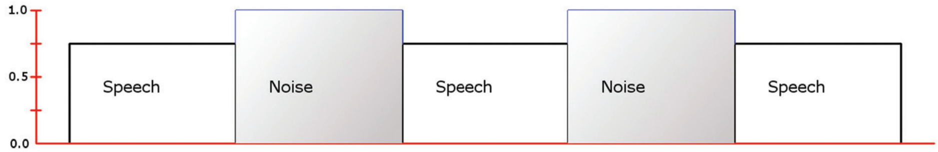


Figure 1.2: Schematic representation of interrupted speech with filler noise. The speech stream is periodically interrupted and the silent intervals are filled with bursts of speech shaped noise.

Subsequent studies have shown the positive influence of context and linguistic rules. Restoration seems to be easier with speech segments that are linked by context and linguistic rules (Sivonen et al. 2006b). Context influences possibly the prior knowledge needed to restore the interrupted speech (Shahin and Miller 2009). Linguistic skills, vocabulary, and verbal comprehension were indirectly suggested to be of importance on the perception of interrupted speech (Bashford and Warren 1987; Bashford et al. 1992; Bronkhorst et al. 1993).

Another factor influencing the benefit from restoration mechanisms is *hearing impairment*. Hearing-impairment changes the quality of the bottom-up speech signals, which makes it more difficult for the highly cognitive restoration mechanisms to do their work. Başkent, Eiler, & Edwards (2010) demonstrated that mildly hearing impaired listeners experience a similar restoration benefit from filler noise as NH listeners. However, negligible or no restoration benefit was observed in moderate hearing impairment. Front-end signal processing of hearing aids may have a negative affect on the sound signals presented to the hearing-impaired listener as well (Başkent et al. 2009). Nelson & Jin (2004) reported that severely hearing-impaired listeners using a CI experience even more difficulties in understanding interrupted speech. The spectral resolution is degraded in CI speech processing and the temporal fine structure is not fully transmitted (Nelson and Jin 2004; Chatterjee et al. 2010). CI users often face challenges in speech perception in complex acoustic listening situations, possibly due to limited help from the benefit of high-level restoration mechanisms. A possible cause is that the signal does not contain the necessary speech cues to induce cognitive restoration mechanisms.

Normal-hearing (NH) listeners presented with acoustic simulations of CI's face similar difficulties, suggesting that the spectrotemporally reduced speech signal, as it can occur in a real CI, may be contributing to the lack of beneficial phonemic restoration. CI's divide the sound into frequency channels and eventually stimulate the auditory nerve fibers via electrodes on the array implanted in the cochlea. The number of electrodes is limited; up to 24 electrodes are available, depending on the manufacturer of the CI, however, the actual functional spectral channels have been shown to be even smaller, possibly due to channel interactions (Friesen et al. 2001). This small number of independent stimulation channels stimulated simultaneously leads to spectral degradation of sound. This degradation has a detrimental effect on the restoration benefit of filler noise in interrupted speech (Başkent 2012), when tested with acoustic simulations of CIs. However, a recent study by Bhargava, Gaudrain, & Başkent (2014) found more promising results in CI users; a restoration benefit was observed when the speech segments were made longer than the interruptions. Further, a restoration benefit was observed with a few CI users, who were also star participants. This suggests that there is a possibility that CI users can achieve benefit from restoration; however, it currently does not work for them as effectively as it does for normal-hearing individuals.

Definitions

Phonemic restoration: the brain fills in for missing parts of speech thus forming a continuous percept of the speech stream

Phonemic restoration benefit of filler noise: the increase in intelligibility of periodically interrupted speech when the silent intervals are filled with noise bursts

OUTLINE OF THE THESIS

Chapter 1 '*Introduction*' is a general introduction to restoration mechanisms of degraded speech. A short overview is given, describing phonemic restoration and the influences of hearing impairment on speech communication and on the effectiveness of restoration mechanisms.

PART I

In the first part of the thesis we report studies with NH listeners who were tested with interrupted speech without further spectral degradations. Before the studies presented in this thesis were performed it was under debate if interrupted speech without filler noise was an ecologically valid stimulus, since it does not occur in natural circumstances. Thus, it was unknown if the phonemic restoration benefit of filler noise was a real perceptual phenomenon, or simply a side effect of the unfamiliarity with this stimulus. If the effect is real, it would be persistent after intensive training with interrupted speech without filler noise, which would remove the unfamiliarity factor. Furthermore, it was implied in literature that linguistic and cognitive factors would influence the restoration capacities of people, however, it was unknown exactly which of these factors would be important for speech restoration.

In Chapter 2 '*Perceptual Learning of Interrupted Speech*' we explored the effect of training on the perception of interrupted speech with silent intervals and with filler noise. Furthermore, we studied how the restoration benefit of the filler noise would evolve during training with feedback (auditory feedback/sentences presented on monitor). An indication of the listening-effort involved is given.

In Chapter 3 '*Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities*' we investigated whether linguistic skills, such as receptive vocabulary, and cognitive skills, such as overall intelligence, might account for the individual differences in the use of top-down restoration mechanisms of trained participants.

PART II

In the second part of the thesis we took a step towards the ultimate goal, namely, to be able to teach CI users to make better use of high-level repair mechanisms and thus to improve speech perception in complex listening situations. We tested and trained NH listeners with interrupted speech, that was degraded as it may happen in CI speech processing (spectrotemporally degraded sentences). Further, auditory and visual factors influencing the phonemic restoration benefit of filler noise, with or without such degradations on the speech signal, are studied.

In Chapter 4 '*Perceptual learning of temporally interrupted spectrally degraded speech*' we studied whether the perception of CI simulations of interrupted sentences can improve despite poor baseline intelligibility performances. Then, we studied if listeners can derive a restoration benefit of the filler noise through training with feedback.

In Chapter 5 '*The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations*' the influence of visual speech cues, in the form of an accompanying video of the speaker, on the enhancement of the intelligibility and phonemic restoration benefit of filler noise is investigated. This is performed with interrupted speech with and without spectrotemporal degradations, with and without additional visual cues.

In Chapter 6 '*General conclusions*' is a discussion of the main outcomes of the studies in this thesis and of their implications and future perspectives for hearing impaired listeners and cochlear implant users.

PART I

TEMPORALLY DEGRADED SPEECH

CHAPTER 2

Perceptual learning of interrupted Speech

M.R. Benard
D. Başkent

This chapter is published in:
PLoS ONE
2013 | Volume 8 | Issue 3 | e58149

ABSTRACT

The intelligibility of periodically interrupted speech improves once the silent gaps are filled with noise bursts. This improvement has been attributed to phonemic restoration, a top-down repair mechanism that helps intelligibility of degraded speech in daily life. Two hypotheses were investigated using perceptual learning of interrupted speech. If different cognitive processes play a role in restoring interrupted speech with and without filler noise, these two forms of interrupted speech would be learned at different rates and with different perceived mental effort. If the restoration benefit is an artificial outcome of using the ecologically invalid stimulus of speech with silent gaps, this benefit would diminish with training. Two groups of normal-hearing listeners were trained, one with interrupted sentences with the filler noise, and the other without. Feedback was provided with the auditory playback of the unprocessed and processed sentences, as well as the visual display of the sentence text. Training increased the overall performance significantly, however restoration benefit did not diminish. The increase in intelligibility and the decrease in perceived mental effort were relatively similar between the groups, implying similar cognitive mechanisms for the restoration of the two types of interruptions. Training effects were generalizable, as both groups improved their performance also with the other form of speech than that they were trained with, and retainable. Due to null results and relatively small number of participants (10 per group), further research is needed to more securely confirm that a general top-down mechanism is responsible for the restoration of interrupted speech with or without the filler noise, as the present results suggest. Nevertheless, training with interrupted speech seems to be effective, stimulating participants to more actively and efficiently use the top-down restoration, thereby implying the potential of this training approach as a rehabilitative tool for hearing-impaired/elderly populations.

INTRODUCTION

Normal-hearing listeners use several top-down mechanisms that help speech perception in difficult listening environments. They may, for example, perceptually restore inaudible or masked portions of temporally interrupted speech, taking advantage of the context and redundancy in speech signals, as well as using linguistic rules, prior knowledge, and expectations (Warren 1970; Bregman 1994; Srinivasan and Wang 2005; Kashino 2006). In the special case of phonemic restoration, the restoration benefit is commonly shown by the increase in intelligibility of periodically interrupted speech when the silent intervals are filled with noise bursts that would be capable of masking the speech (Powers and Wilcox 1977; Bashford and Warren 1979; Verschuure and Brocaar 1983; Bashford et al. 1992; Bregman 1994; Srinivasan and Wang 2005; Kashino 2006; Başkent et al. 2009). Increase in intelligibility as a result of adding noise to speech signals is somewhat counterintuitive. However, in the case of restoration, the filler noise adds ambiguity for the perceptual system, where the system then tends towards forming a full object, rather than perceiving the individual pieces per se, referred to as the Gestalt principles of closure (Bregman 1994; Srinivasan and Wang 2005). These closure mechanisms, then, presumably help with the speech restoration.

A number of hypotheses have been proposed to explain the underlying mechanisms of restoration that produce the improvement in intelligibility with the filler noise. Huggins (1964) noted that the filler noise masks the distortions that occur due to the sudden onsets and offsets in interrupted speech, and thus suggested that the bottom-up processes of the auditory system are entirely responsible for this benefit. Others, on the contrary, pointed to the involvement of the high-level cognitive processes, based on the influence that the context and the type of speech materials used had on the perception of interrupted speech (Miller and Licklider 1950; Warren and Sherman 1974; Bashford and Warren 1979; Warren 1983; Bashford et al. 1992; Sivonen et al. 2006b; Grossberg and Kazerounian 2011). Recent studies that showed a deficit in restoration benefit with (real or simulated) hearing impairment implied that the restoration may actually be governed by a combination of the bottom-up peripheral and top-down cognitive processes (Nelson and Jin 2004; Başkent et al. 2010; Chatterjee et al. 2010; Başkent 2010; Başkent and Chatterjee 2010; Başkent 2012; Bhargava and Başkent 2012), in agreement with general high-level speech and sound perception mechanisms in complex listening environments (Bronkhorst et al. 1993; Samuel 1997; Alain et al. 2001; Winkler et al. 2005; Davis and Johnsrude 2007; Janse and Ernestus 2011). Hence, the consensus from recent studies is that cognitive processes are involved in the phonemic restoration mechanism, but up to what degree is still not clear.

Based on their observations, Verschuure and Brocaar (1983) suggested that the degree of involvement of cognitive processes might differ in the perception of interrupted speech with silent intervals from the perception of interrupted speech combined with filler noise. For example, without the filler noise, the listeners were anecdotally reported to be aware of the silent intervals in the signal, and seemed to be forced to guess consciously what could have been presented to them. With the noise, the listeners unconsciously filled in the missing speech information. One could expect different effects of training on tasks that require different cognitive resources and that differ in how automatic and effortless they are (Shiffrin and Schneider 1977). Therefore, exploring learning effects with interrupted speech with or without the filler noise could be used to show if there is such a difference.

Other than indicating potentially varying cognitive processes, perceptual learning effects could reveal other factors relevant to phonemic restoration. Speech interrupted with silent intervals is a less ecologically valid signal than interrupted speech with filler noise, because in real life speech is more often obliterated by noise than by silence. A poorer intelligibility with interrupted speech with silent intervals than interrupted speech with filler noise may then be observed, not due to the restoration benefit per se, but due to the participants being less used to hearing such artificial manipulations. If the restoration benefit of adding filler noise were not a real effect but a consequence of such an artifact, then it would be expected to diminish or disappear after listeners are exposed to and trained with these artificial speech stimuli.

The present study explored the effects of perceptual learning, i.e. the improvement in performance after systematic long-term training (Goldstone 1998; Fahle 2004), on the perception of interrupted speech. The purpose was to explore the hypotheses that the cognitive involvement could differ for understanding interrupted speech with or without the filler noise, and that the restoration benefit could be an artifact of using interrupted speech, an ecologically not valid signal produced by an artificial manipulation. The performance on many auditory skills improves with training (Goldstone 1998; Wright and Zhang 2009), commonly given in the form of an explicit training (Davis et al. 2005), although improvement due to unattended exposure is also possible (Seitz et al. 2011). While humans adapt relatively automatically to rather simple stimuli (Eisner and McQueen 2006), more complex stimuli, like speech manipulated with time compression (Adank and Janse 2009), spectral reduction (Hervais-Adelman et al. 2008; Loebach et al. 2009), or interruption by masker (Gnansia et al. 2010) may need more effort to adapt to. Participants of the present study were systematically trained with speech manipulated with two kinds of interruptions, with silent intervals or with silent intervals combined with the filler noise, and speech intelligibility and perceived mental effort were measured before, during, and after training. Based on the studies listed above and due to the complex nature of the stimuli, an intensive training with feedback was preferred. If the cognitive involvement varied between the two kinds of speech signals, the effort requirement of the two tasks and the effects of learning on intelligibility and perceived effort would be expected to differ. If the restoration benefit was due to an artifact of using interrupted speech with silent intervals, it would be expected to diminish or disappear at the end of training.

MATERIALS AND METHODS

Listeners

Thirty normal-hearing listeners, ages between 18 and 28 years (Mage = 21.3 years, SD = 2.4 years, 21 women), participated in the study. During the initial screening, normal hearing via a hearing test (at test frequencies of 0.5 kHz up to 4 kHz, hearing thresholds of 20 dB HL or less) and normal development of speech and language via a questionnaire were confirmed. The listeners, all native speakers of Dutch, were divided into three groups, matched on age and gender. The baseline performances were measured before and after the training sessions. Two groups received training with feedback between the baseline measurements. The noise group (NG) was trained with interrupted speech with the filler noise and the silence group (SG) without. The third group did not receive any training. They were only tested with baseline conditions applied at two different days, and thus served as the control group (CG). From the SG and the NG, 7 and 6 listeners, respectively, participated in a follow-up testing at a later time to observe the retainability of the learning effects.

Ethics statement

The study was approved by the Medical Ethical Committee of the University Medical Center of Groningen. The listeners were recruited by poster announcements at public places and participation was compensated financially. Information about the experiment was provided and written informed consent was collected prior to participation.

Stimuli

The speech stimuli were Dutch sentences digitally recorded at 44.1 kHz sampling rate and spoken by a male speaker (Versfeld et al. 2000). The sentences are semantically neutral and represent conversational speech. The database consists of 39 sets. Each set contains 13 sentences, with 4 to 9 words per sentence, and 74 to 88 words in total. The sentences were interrupted by a cosine-ramped (ramp duration of 10 ms) periodic square wave with 1.5 Hz interruption rate and 50% duty cycle. This resulted in speech portions followed by interruptions of 333 ms of duration each. Former studies (Miller and Licklider 1950; Powers and Wilcox 1977; Başkent and Chatterjee 2010; Başkent 2012; Bhargava and Başkent 2012) and our pilot study have shown that these parameters led to a low baseline intelligibility performance with interrupted speech, leaving room for potential improvement in intelligibility after adding the filler noise and providing training to the listeners. The noise used to fill the silent intervals in the sentences was steady speech-shaped noise, 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 filler noise bursts were produced by applying the same periodic square wave, except with inverted phase, to the speech-shaped noise. The interrupted speech and the noise bursts were combined in a way such that there was sufficient but minimal overlap between the two (see Başkent et al. (2009) for details).

The root mean square intensity was normalized to the same fixed value for all sentences. The presentation levels of the speech and filler noise were calibrated to 60 and 70 dB SPL (based on Başkent, 2012), respectively, when measured at an approximate position of the participant's head.

Experimental procedure

The participants were seated in an anechoic chamber, facing the free-field loudspeaker and the monitor that presented the visual feedback at a distance of 1 m. The digitized processed stimuli were directed from an external AudioFire 4 soundcard of Echo Digital Audio Corporation to a Tannoy 8D Precision active near-field speaker. The experimenter was seated outside the anechoic chamber and listened to participants' responses via a headphone connected to the digital voice recorder, DR-100 digital by Tascam, of the anechoic room. As the stimuli were presented in free field, the experimenter inadvertently also heard the stimuli. Any potential bias that may have been caused by this single-blind design must have been negligible, as the restoration effect observed in the present study was comparable to the restoration effects observed in double-blind studies conducted within our research group (Başkent 2012).

The experimental procedure consisted of initial and final baseline measurements of intelligibility and perceived effort, with training sessions in-between (Table 1; details below). The difference in the baseline scores before and after training thus showed the improvement in performance due to perceptual learning as a result of training. The interval between the initial and final baseline measurements varied slightly, between 2 and 3 days, depending on the availability of the participants. The training was spread over three days and the entire experiment, including the participant screening and initial and final baseline tests, was completed in less than one week. At a later time, 6 to 18 weeks after the training was completed, a follow-up baseline test was conducted to observe how the training effects were retained. In the entire study, a MATLAB program was used to process the stimuli online and to present the processed stimuli and audio and visual feedback to the participants via a graphical user interface.

In the baseline measurements before and after training, all three groups were tested on speech intelligibility and perceived mental effort with interrupted speech, with and without the filler noise. In each condition of the baseline measurements, 2 sets (26 sentences) were randomly selected from the 39 sets, so participants were exposed to 52 unique sentences before and 52 unique sentences after training. In baseline measurements and during training, no sentence was heard for more than once. In the intelligibility tests, the participants listened to one sentence at a time, and they were instructed to repeat all of the words they heard, even if this led to nonsense sentences. Guessing the missing words was encouraged as the purpose of the test was to assess the reconstructed perception of the sentence, rather than what is heard per se. The participants were instructed to tell the experimenter when they were ready for the next sentence (by saying Next). Scoring of correctly repeated words was first performed in real time by the experimenter, and was later checked once more by offline listening to digital recordings of participants' responses. All words were included in the scoring. The percentage of correctly identified words was calculated as the ratio between the total number of correctly repeated words and the total number of words within the sets. The participants heard a set of sentences only once, and they were not familiar either with the speech material used or with listening to interrupted speech in general before their participation in this study.

In the training sessions, as shown in Table 1, the NG and SG were trained with different stimuli. The CG received no training and did not attend the training sessions. They participated only in two sessions for the baseline measurements. The duration between these baseline measurements of the CG represented the typical duration, of 2 to 3 days, between the two baseline measurements (before and after) of the trained groups. In each of the 5 training sessions, 26 sentences were used, so that the SG and NG were trained with 130 sentences. In total, the CG was exposed to 104 unique sentences and the SG and NG to 234 unique sentences. The difference in the training sessions compared to the baseline measurement sessions was that during the training feedback was provided. After receiving the participant's response, first the unprocessed then the processed sentence were played back (based on Davis et al., 2005), while the text of the sentence was simultaneously displayed on the computer screen.

Groups	Baseline measurement before	Training	Baseline measurement after	Follow-up baseline measurement
SG (n=10)	Silence and noise	Five silence training sessions	Silence and noise	(n=7), silence and noise
NG (n=10)	Silence and noise	Five noise training sessions	Silence and noise	(n=6), silence and noise
CG (n=10)	Silence and noise	No training or testing session	Silence and noise	

Table 1: Experimental procedure, shown for the noise (NG), silence (SG) and control (CG) groups, along with the number of participants (n). "Silence" denotes testing with interrupted sentences with silent intervals, and "noise" denotes testing with interrupted sentences that are combined with filler noise bursts. The CG did not receive training; they were only tested with the baseline measurements at two different times, with an in-between time comparable to that of the training duration.

At the end of each session (baseline or training) and for all participants, the perceived mental effort was measured using the Visual Analogue Scale (VAS), a subjective measure shown to be sensitive to small differences in mental effort (Maxwell 1978; Paas et al. 1994). This method, while not evaluated objectively in previous studies, was selected due to the ease of use. The participants were instructed to rate the effort of the comprehension for the entire session by a mark on a 10 cm long scale, varying from "effortless" (0 on VAS-scale) to "effortful" (10 on VAS-scale) on paper. Listening to a known poem in quiet (effortless) and having a conversation in loud noise (effortful) were given as examples to the participants to interpret the full range of the VAS scale.

RESULTS

Speech intelligibility

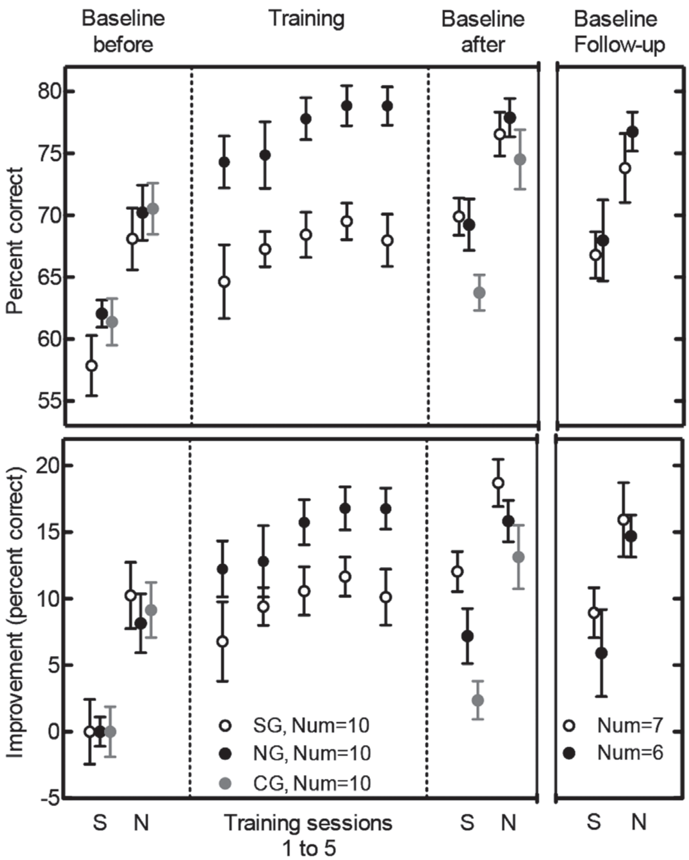


Figure 1: Intelligibility of interrupted speech with and without filler noise. The absolute mean percent correct scores from all listener groups are shown for baseline and training sessions in the top panel. The relative mean percent correct improvement, calculated by normalizing the absolute scores with respect to the 'S' condition before training, is shown in the bottom panel, The 'S' (Silence) and 'N' (Noise) on the horizontal axes denote the conditions with interrupted sentences with silent intervals and with filler noise in the interruptions, respectively. The open, filled, and gray data points represent the results from the silence (SG), noise (NG), and control (CG) groups, respectively. The panels from left to right show the results of baseline measurements before training, measurements made right after each training session during the training, baseline measurements after training, and the follow-up baseline measurements conducted at a later time (also see Table 1). The CG received no training and were only tested with the baseline measurements. Error bars denote one standard error of the mean.

The top panel of Fig. 1 shows the mean percent correct scores for all sessions (baseline and training), as well as the follow-up baseline measurement; the bottom panel shows the increase in percent correct for all sessions, relative to the silence (S) condition of the baseline measurement before training. The purpose of the normalization in the lower panel was to better visualize the change in intelligibility due to training, as well as due to addition of the filler noise. The baseline speech intelligibility scores measured before and after the training are shown in the first and third segments of Fig. 1, respectively, in both top and bottom panels (also summarized in Table 2). These data show that there was a restoration benefit before training with each listener group, and even though the training increased the scores in both S and noise (N) conditions, a similar restoration benefit could still be observed after the training. In the initial baseline measurement, on average, there was a restoration benefit of 9.2%, as shown by the increase in scores with the addition of the filler noise ('N' column compared to the 'S' column in "before training" scores in Table 2). After the training, a similar restoration benefit was observed with, on average, 8.7% (middle column of Table 2). Repeated measures ANOVAs were performed with both forms of the percent correct scores, the absolute percent correct scores in Fig. 1 top panel and the relative percent correct scores in Fig. 1 bottom panel, with addition of filler noise and training as within-subjects factors and participant group as the between-subjects factor. The ANOVAs showed that this restoration benefit was significant ($F(1,27) = 106.4$, $p < 0.001$, partial $\eta^2 = 0.798$, power = 1). The improvement after the training sessions is shown in the increase of scores in S and N conditions from before to after baseline measurements in Fig. 1, and also in the rightmost columns of Table 2. The training produced significant overall improvement ($F(1,27) = 28.3$, $p < 0.001$, partial $\eta^2 = 0.512$, power = 1), varying from 7.2 to 12%, for both training groups and for both testing conditions of S and N. Although the CG improved in performance as well, their improvement was smaller, 2.4% to 4%. The analysis performed with the absolute percent correct scores (top panel) showed no significant difference between the three groups ($F(2,27) = 1.2$, $p = 0.307$, partial $\eta^2 = 0.084$, power = 0.25) and no significant interaction effect. The analysis performed with the relative percent correct scores (bottom panel), however, showed a significant difference between the three groups ($F(2,27) = 3.6$, $p = 0.041$, partial $\eta^2 = 0.211$, power = 0.62) and no significant interaction effect. Note that the SG started from a lower baseline performance level than the NG and CG (Fig.1, top panel). Hence, the training effect was highest for the SG (Fig.1, lower panel).

The middle section of the top panel of Fig. 1 shows the absolute percent correct scores measured during the training sessions where the feedback was provided; the bottom panel shows the same, except that the scores are normalized relative to the S condition of the baseline measurement before training. These data show that training increased the scores with interrupted speech with or without the filler noise, but the intelligibility of interrupted sentences combined with filler noise was always higher than the interrupted sentences with silent intervals. This means that the restoration benefit observed in the baseline measurement before the training was retained throughout the training. Repeated measures ANOVAs were performed with both absolute and relative percent correct scores, with the training sessions as within-subjects factor and the addition of filler noise as between-subject factor. These showed that the improvement in both absolute and relative scores between the five training sessions was not significant ($F(4,15) =$

2.0, $p = 0.145$, partial $\eta^2 = 0.106$, power = 0.47; same for both absolute and relative scores as the normalization did not change this effect), but the restoration benefit due to added noise was, for both absolute ($F(1,18) = 35.9$, $p < 0.001$, partial $\eta^2 = 0.667$, power = 1) and relative percent correct scores ($F(1,18) = 10.9$, $p = 0.004$, partial $\eta^2 = 0.377$, power = 0.88). There was no significant interaction effect.

Groups	Absolute PC scores baseline before (%)		Absolute PC scores baseline after (%)		Improvement (%)	
	S	N	S	N	S	N
SG (n=10)	57.9	68.1	69.9	76.6	12.0	8.5
NG (n=10)	62.1	70.2	69.2	77.9	7.2	7.7
CG (n=10)	61.4	70.5	63.8	74.5	2.4	4.0
	Relative PC scores baseline before (%)		**Relative PC scores baseline after (%)**		**Improvement (%)**	
	S	**N**	**S**	**N**	**S**	**N**
SG (n=10)	0.0	10.2	12.0	18.7	12.0	8.5
NG (n=10)	0.0	8.1	7.2	15.8	7.2	7.7
CG (n=10)	0.0	9.1	2.4	13.1	2.4	4.0

Table 2: The absolute (top rows) and relative (bottom rows) mean percent correct (PC) scores of the baseline measurements before and after training of the SG and NG (left and middle columns), and overall improvement taken from Fig. 1 (right column). The CG received no training and were only tested with the baseline measurements, to see the potential learning effects due to the exposure to testing paradigm only, in the lack of a targeted training. 'S' and 'N' refer to testing conditions with interrupted sentences with silent intervals or with filler noise, respectively.

The right segments of Fig.1 show the intelligibility of the follow-up testing, performed with 7 participants from the SG and 6 participants from the NG, at 42 to 127 days (M = 92 days, SD = 27 days) after the second baseline measurement. These data show that the restoration benefit was still significant and overall, the scores were more similar to the trained-level scores than the initial un-trained level scores. Repeated measures ANOVAs (conducted on this subset of listeners only; with the within-subjects factors testing after training or follow-up and the addition of filler noise and the between-subjects factor the participant group) showed no significant difference between the baseline measurement after the training and the follow-up testing ($F(1,11) = 1.4$, $p = 0.265$, partial $\eta^2 = 0.111$, power = 0.19), but a significant effect of the restoration benefit ($F(1,11) = 23.6$, $p = 0.001$, partial $\eta^2 = 0.682$, power = 0.99) for both representations of data (as the statistics were independent than the normalization). There was no significant interaction effect. The analysis performed with the absolute percent correct scores ($F(1,11) = 0.04$, $p = 0.853$, partial $\eta^2 = 0.003$, power = 0.05, top panel) and with the relative percent correct scores ($F(1,11) = 3.2$, $p = 0.101$, partial $\eta^2 = 0.226$, power = 0.37, bottom panel) showed no significant difference between the two groups.

Perceived mental effort

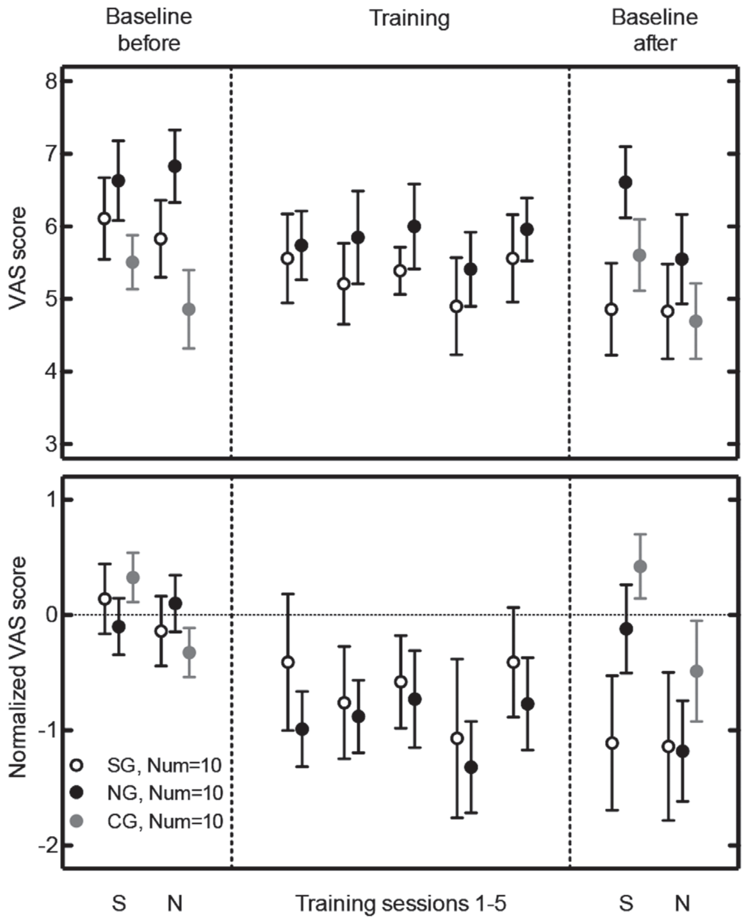


Figure 2: Perceived mental effort. The absolute and normalized mean mental effort scores are shown in the top and bottom panels, respectively. These scores are measured by means of a visual-analogue scale (VAS), varying from "effortless" (0 on VAS-scale) to "effortful" (10 on VAS-scale). The first and the third panels show the scores measured before and after the training, respectively. The middle panel shows the scores during the training sessions. Error bars denote one standard error of the mean.

The top panel of Fig. 2 shows the mean perceived mental effort scores for all testing sessions, and the bottom panel shows the mean perceived mental effort scores normalized over the average of the S and N conditions before training. The purpose of this normalization, different than Fig. 1, was to minimize the variability in the utilization of the VAS-scale between the participants. Therefore, the scores were not normalized with respect to 'S', but instead, with respect to participants' own baseline ratings. The first and third segments of both panels of Fig. 2 represent the perceived mental effort of the baseline measurements before and after the training sessions, respectively. These data show that while there was a tendency for the N condition to be perceived less effortful compared to the S condition, during and after the training sessions, there were also some exceptions, such as the S condition after the training. In the initial and final baseline measurements, there was on average a significant decrease in perceived mental effort with the addition of the filler noise ('N' column compared to the 'S' column in "before training" and "after training" scores in Table 3; $F(1,27) = 7.0$, $p = 0.014$, partial $\eta^2 = 0.205$, power = 0.72, for both absolute and normalized VAS-scores). The training significantly reduced the perceived mental effort, shown by the decrease in VAS between "before" and "after" baseline measurements in Fig. 2, and also in the rightmost columns of Table 3 ($F(1,27) = 8.0$, $p = 0.009$, partial $\eta^2 = 0.228$, power = 0.78, for both representations of the VAS-scores). There was no significant difference between the three groups both when represented in absolute ($F(2,27) = 2.35$, $p = 0.114$, partial $\eta^2 = 0.148$, power = 0.43) and normalized VAS-scores ($F(2,27) = 2.20$, $p = 0.130$, partial $\eta^2 = 0.140$, power = 0.41) There was no significant interaction effect.

Groups	VAS scores before		VAS scores after		Improvement	
	S	N	S	N	S	N
SG (n=10)	6.1	5.8	4.9	4.8	1.3	1.0
NG (n=10)	6.6	6.8	6.6	5.6	0.0	1.3
CG (n=10)	5.5	4.9	5.6	4.7	-0.1	0.2
	Normalized VAS scores before		**Normalized VAS scores after**		**Improvement**	
	S	N	S	N	S	N
SG (n=10)	0.14	-0.14	-1.11	-1.14	1.25	1.00
NG (n=10)	-0.10	0.10	-0.12	-1.18	0.02	1.28
CG (n=10)	0.33	-0.33	0.42	-0.49	0.09	0.16

Table 3: Similar to Table 2, except the scores shown are the absolute mean perceived mental effort scores (top rows) and the normalized mean perceived mental effort scores with respect to the baseline measurements before training (bottom rows), measured by means of a visual-analogue scale (VAS).

The middle sections of Fig. 2 show the VAS-scores measured during the training, after each training session. These data show that there was no significant change in VAS scores during the training ($F(4,15) = 1.4$, $p = 0.269$, partial $\eta^2 = 0.049$, power = 0.34, for both absolute and normalized VAS-scores). The absolute VAS-scores show that the SG rated the perceived effort lower than the NG, but the difference was not significant ($F(1,18) = 0.54$, $p = 0.472$, partial $\eta^2 = 0.029$, power=0.11). This difference is also not significant for the normalized VAS-scores ($F(1,18) = 0.37$, $p = 0.550$, partial $\eta^2 = 0.020$, power=0.09). There was no significant interaction effect.

DISCUSSION

Before training, a baseline intelligibility of interrupted sentences, with and without the filler noise, was measured. These pre-training results were comparable to previous studies on intelligibility of interrupted speech (Powers and Speaks 1973; Nelson and Jin 2004; Gilbert et al. 2007; Bhargava and Başkent 2012), and on restoration benefit observed with additional filler noise in silent intervals (Powers and Wilcox 1977; Bashford and Warren 1979; Verschuure and Brocaar 1983; Başkent et al. 2009).

The first interest of the present study was to observe the effect of training on the perception of interrupted speech with silent intervals and with the filler noise, as a way of exploring the similarity in the underlying cognitive mechanisms involved in the perception of the two types of stimuli (with or without the filler noise). Verschuure and Brocaar (1983) hypothesized, based on their observations during their study, that the perception of speech interrupted by silence involves other cognitive processes than the perception of interrupted speech combined with filler noise. We further hypothesized that if the cognitive involvement varied in the perception of the two kinds of speech signals, they would be learned at different rates with training. The speech intelligibility results showed that the percent correct scores increased during the training sessions similarly for both training groups. In other words, speech with both forms of experimental manipulations could both be learned, and in a similar manner too. Hence, the results imply that speech perception with both forms of interruptions (with silence or with filler noise) involves similar cognitive mechanisms.

In addition to speech intelligibility, perceived mental effort was also measured. Processes requiring cognitive awareness are suggested to be more effortful than unconscious processes (Cohen et al. 1990). Based on the observations by Verschuure and Brocaar (1983), therefore, we had hypothesized that if cognitive mechanisms differed between the perception of interrupted speech with or without the filler noise, we would see a difference in the perceived effort scores with the two forms of speech. In fact, the analysis of the perceived mental effort showed on average a small, though significant, decrease in VAS-score with the addition of the filler noise, both before and after training. However, because of irregularities in the scoring between the groups, such as the high score of the NG for the S condition after training, we reckon that these differences in VAS-scores, although normalized, stem from the individual preferences of different groups, rather than a direct result of the experimental manipulation. There was no systematic change in effort scores during training. However, when the scores were compared for before and after training, there was a decrease in the rating of the perceived mental effort, and in similar values for the two forms of interruptions. These results on perceived effort, hence, only partially support the hypothesis.

The second hypothesis of the study was that if the restoration benefit was due to an artifact of using interrupted speech with silent intervals, an unusual and less ecologically valid form of speech, it would be reduced or entirely disappear at the end of the training. This idea was also suggested by Verschuure and Brocaar (1983), who reported that participants did not benefit from adding noise in the silent intervals when they were familiar (i.e. trained) with this type of interrupted speech. The suggestion was only anecdotal, as their data were limited due to the ceiling effects and there was no systematic investigation of learning effects. The results from the present study are in contradiction to the observations by Verschuure and Brocaar (1983), because training did not bring the intelligibility of the two forms of interrupted stimuli to the same level. By training the participants, we observed a relatively similarly increasing curve in the overall percent correct scores of the SG and NG, and the restoration benefit persevered. The baseline measurements after the training showed that the restoration benefit of adding noise was still present after training, indicating that the restoration benefit is not an effect due to the artificiality of interrupted speech with silent intervals.

During the training, a plateau was observed in the scores, in a similar manner between the two training groups. We interpreted this as that the groups reached the limit of learning with these stimuli and that sufficient training was given. We made our conclusions based on this interpretation. However, there were perhaps some additional factors that affected the results. For example, we cannot exclude the possibility that a part of the increase in performance can be explained by the familiarity of the participants with the talker's voice (Nygaard and Pisoni 1998; Magnuson et al. 2003), as we used sentences spoken by one talker only. Because the SG and NG were trained with different stimuli, stimuli-specific effects might also have additionally (but perhaps only slightly) influenced the shape of the increasing curves of these groups. Further, null results combined with relatively small number of subjects indicate that the paradigm used in the present study was perhaps not sufficient to fully validate the conclusions, and further research with more statistical power would be needed to more confidently make such conclusions.

The training results also imply a potential benefit of the specific training paradigm used in the study. Note that the amount of speech information provided and the distortions inherent in the signals due to interruptions were the same across training sessions, and yet intelligibility of interrupted speech, with or without the filler noise, increased significantly as a result of the training. This outcome not only suggests that the restorative mechanisms for understanding interrupted speech are probably highly cognitive, but also that our training paradigm seems to train the listeners effectively to make better use of the top-down repair mechanisms. The training was generalizable; participants showed an increase in performance also for the other speech manipulation than the one they were trained with. Additionally, the training effects were retained in line with earlier perceptual-learning studies (Schwab et al. 1985; Francis et al. 2000; Francis et al. 2007; Francis and Nusbaum 2009); several weeks after the training the percent correct scores were not significantly different from the baseline measurements taken immediately after the training. Perceptual learning, a relatively permanent change of perception as a result of training, was hence achieved. These observations point to the potential benefits of the type of training used in the present study as a tool for speech perception rehabilitation, for example, for the (elderly) users of hearing aids and cochlear implants (Montgomery et al. 1984; Fu and Galvin 2003; Sweetow and Palmer 2005; Stacey and Summerfield 2008; Fu and Galvin 2008).

ACKNOWLEDGEMENTS

The authors would like to thank Annemieke ter Harmsel for her help with collecting data, and Carina Pals, Pranesh Bhargava, and Etienne Gaudrain for their valuable feedback on the study design and on earlier versions of this manuscript. The study is part of the research program of our department: Healthy Aging and Communication.

CHAPTER 3

Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities

M.R. Benard
J.S. Mensink
D. Başkent

Published in modified form:
The Journal of the Acoustical Society of America
2014 | Volume 135 | Issue 2 | EL88-94

ABSTRACT

Top-down restoration mechanisms can enhance perception of degraded speech. Even in normal hearing, however, a large variability has been observed in how effectively individuals can benefit from these mechanisms. To investigate if this variability is partially caused by individuals' linguistic and cognitive skills, normal-hearing participants with a wide range of ages were assessed for receptive vocabulary (Peabody Picture Vocabulary Test; PPVT-III-NL), for full-scale intelligence (Wechsler Adult Intelligence Scale; WAIS-IV-NL), and for top-down restoration of interrupted speech (with silent or noise-filled gaps). Receptive vocabulary was significantly correlated with the other measures, suggesting linguistic skills to be highly involved in restoration of degraded speech.

INTRODUCTION

Listeners use top-down restoration mechanisms in complex acoustical environments to enhance perception of degraded speech (Warren 1983; Sivonen et al. 2006b). The brain is able to use audible segments of interrupted speech for its perceptual restoration. However, while this effect is robust and highly repeatable on average, even in well-controlled young normal-hearing populations, a large variability has been observed in how well individuals can make use of these mechanisms (Başkent 2010).

The influences of linguistic and cognitive skills on the perception of interrupted speech has been a topic of debate (Sivonen et al. 2006b; Başkent 2012; Benard and Başkent 2013). Previous research has indirectly implied that, in such top-down restoration, linguistic and cognitive factors seem to play an important role. For example, restoration seems to be easier with speech segments that are linked by context and linguistic rules (Verschuure and Brocaar 1983; Sivonen et al. 2006b; Wang and Humes 2010). Linguistic skills, vocabulary and verbal comprehension were indirectly suggested to be of importance in the perception of interrupted speech (Bashford et al. 1992; Saija et al. 2013). Training improves the intelligibility of interrupted speech, and aging can have an effect on it, both implying an involvement of the cognitive system (Benard and Başkent 2013; Saija et al. 2013). Despite these suggestions, however, to date, no systematic study has been performed to directly and fully establish if and to what degree such association exists between linguistic and cognitive factors and the top-down restoration of interrupted speech.

In this study, we hypothesized that 1) linguistic skills, such as receptive vocabulary, and 2) cognitive skills, such as overall intelligence, would account for the individual differences in the use of top-down restoration mechanisms. To test the two hypotheses, standardized and validated tests were used to characterize linguistic and cognitive abilities of normal-hearing individuals with varying ages, and the correlations of these measures with the ability to perceptually restore interrupted speech were investigated.

MATERIALS AND METHODS

Listeners

Twelve native speakers of Dutch, aged between 21 and 63 years (mean age = 40.4 years, standard deviation (SD) = 15.3 years, 9 women), participated in the study. Normal hearing was confirmed with hearing thresholds less than or equal to 20 dB hearing level (HL) when averaged across 500, 1000, 2000 and 4000 Hz, applied to the better hearing ear. Speech and language problems were ruled out via a questionnaire.

Speech stimuli

Dutch high context sentences, representing conversational speech, spoken by a male speaker and digitally recorded at a sampling rate of 44.1 kHz, were used in the study (Versfeld et al. 2000). This corpus consists of 39 sets, with 13 unique sentences per set. Each sentence consists of 4 to 9 words and each set contains 74 to 88 words in total.

Perceptual restoration of interrupted speech was measured with various speech manipulations so that individual variability could be fully exploited. Some of these manipulations involved interrupting speech stimuli with periodic silent intervals, where the listener had to reconstruct the speech stream from the remaining samples (Chatterjee et al. 2010), and some others involved filling the silent intervals with loud noise bursts, which induced continuity illusion and phonemic restoration, quantified by an increase in intelligibility of interrupted speech (Benard and Başkent 2013). Further, to capture cognitive effects fully, such as potential effects of cognitive slowing down due to aging, slow speech rates were also included. As a result, a number of settings were used to measure top-down restoration, with two interruption rates (1.25, 2.50 Hz) at two speech rates (slow, normal), presented with or without filler noise, producing eight testing conditions.

During signal processing, first the speech rate was changed. For slowing down, sentences were lengthened to twice the original duration, preserving the original pitch using PRAAT software (Saija et al. 2013). The normal and slow-rate sentences were interrupted with the two interruption rates, with a duty cycle of 50%, and with 10-ms cosine-ramping on and off transitions. The filler noise bursts were produced with the same gating function, but with an inverse phase, from the steady speech-shaped noise matching the long-term average speech spectra of the sentences in the corpus (Versfeld et al. 2000). In total, 104 unique sentences (8 conditions x 13 sentences per sentence set) were used for training and 104 unique sentences were used for testing. The training and the test procedures took one hour each, including a break. All manipulations were processed online using Matlab on a Macintosh computer.

The sentences were calibrated to a presentation level of 60 dB sound pressure level (SPL) at the approximate position of the participant's ear, and the filler noise to 70 dB SPL.

Peabody Picture Vocabulary Test

We used the Peabody Picture Vocabulary Test (PPVT-III-NL) to measure receptive Dutch vocabulary and verbal intelligence (Bell et al. 2001). The experimenter presented the test verbally, at the same time showing a series of four numbered pictures to the participant. The participants' task was to say the number of the picture that described the word stated by the experimenter. Each series of pictures was part of a set of 12 series. The first presented set depended on the age of the participant, after more than 3 incorrect responses in this set the experimenter presented an easier set. The starting set was the set in which less than 4 incorrect responses were given. After completing one set the experimenter presented the next and more difficult set, ending with the set in which 9 or more incorrect responses were given. The test procedure took around 15 minutes.

The raw scores of this test are age-independent, and can be translated into an age-dependent receptive vocabulary quotient. A score of 100 points is defined as the median of the norming sample; the standard deviation is 15 points. The Dutch standardization sample contains 1746 children (age 2:3 up to 15:11 year) and 1164 adults (age 17:0 up to 90:0 year), controlled for gender, age, education, social status and geographic spread (Schlichting 2005).

Wechsler Adult Intelligence Scale

We used the Dutch version of Wechsler Adult Intelligence Scale (WAIS-IV-NL) to measure the Full Scale Intelligence Quotient, FSIQ (Wechsler 2012), a composite score composed of 10 subtests that measure major components of intelligence, namely verbal comprehension, perceptual reasoning, working memory and processing speed. The experimenter verbally instructed the participants prior to each subtest. The participants had to arrange blocks according to a pattern, describe similarities between two concepts, repeat digit spans, solve nonverbal abstract problems, describe meaning of words (vocabulary), perform arithmetic tasks, search or substitute symbols on a response form, arrange visual puzzles and perform a general knowledge subtest (Benson et al. 2010). The experimenter scored each response and proceeded to a subsequent subtest when the stop criteria, based on incorrect responses or time limits, were met. The test procedure took around two hours, including a short break.

Different than the earlier version WAIS-III-NL, the WAIS-IV-NL does not provide separate verbal and performance intelligence quotient scores, but quotient scores for verbal comprehension, perceptual reasoning, working memory and processing speed. A score of 100 points is defined as the median of the norming sample; the standard deviation is 10 points. The norms are divided into 10 age groups, from 16:0 to 84:11 year. The sample of standardization is based on 1,000 participants in The Netherlands and 500 Dutch-speaking participants in Belgium.

Experimental procedure

The participants were assessed in random order with the PPVT-III-NL and the WAIS-IV-NL, and they were trained and tested with the normal and slow-rate interrupted sentences with or without filler noise (based on Benard and Başkent, 2013). A clinical psychologist conducted the linguistic and cognitive tests according to the standard clinical procedures, as explained above. A clinical audiologist measured the intelligibility of interrupted speech. The speech stimuli were directed to an AudioFire 4 external soundcard of Echo Digital Audio Corporation (California, USA). The participants were seated together with the experimenter in a sound proof booth and the participant listened to the audio stimulus via a Sennheiser HD 600 headphone (Connecticut, USA). The experimenter was listening to the participant's response and scored online using a customized Matlab graphical user interface (GUI; method based on Benard and Başkent, 2013) with the sentence text presented on the screen. All words of the sentences were used for scoring; the Matlab program calculated the ratio between the total number of correctly repeated words and the total number of words within the set presented.

RESULTS

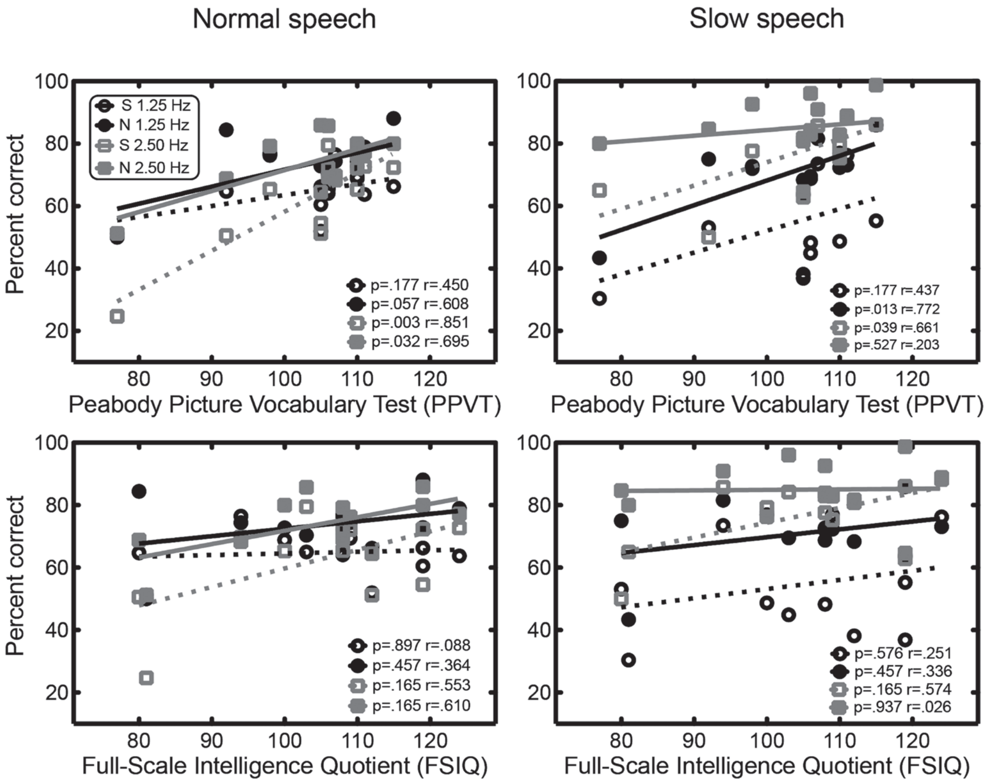


Figure 1: The percent correct scores for intelligibility of interrupted speech shown as a function of the quotient scores of the Peabody Picture Vocabulary Test (PPVT-III-NL; upper panels) and the full-scale intelligence quotient scores of the Wechsler Adult Intelligence Scale (WAIS-IV-NL; lower panels). The columns show results with normal and slow-rate sentences. The open and filled symbols represent the scores without (S conditions) and with filler noise (N conditions), respectively. The black and gray symbols represent the scores with the slow (1.25 Hz) and fast (2.50 Hz) interruptions. The black and gray trend lines represent the best linear regression, by means of the linear least squares method, through the black and gray data points (continuous for N conditions, and dashed for S conditions), respectively. The correlation analyses are indicated in each panel for individual regression lines.

Figure 1 shows the percent correct scores for intelligibility of interrupted speech plotted against the quotient scores of the PPVT-III-NL (upper panels) and the WAIS-IV-NL (lower panels). The left and right panels show the results of the normal and slow speech, respectively. The individual scores are overlaid with fitted linear regression lines and shown separately for different interruption conditions. The statistical analysis, by means of linear regression, shows that half of the speech conditions (4 out of 8) correlate significantly ($p < 0.05$) with the PPVT-III-NL after applying the post-hoc correction False Discovery Rate (FDR) in R statistics. No conditions correlate significantly with the FSIQ scores of the WAIS-IV-NL.

Linear regression was conducted to explore the correlation between the two tests, as the WAIS-IV-NL has subcomponents on linguistic skills (Bell et al. 2001; Benson et al. 2010). This analysis showed that the PPVT-III-NL is indeed highly correlated with the FSIQ ($F(1,10) = 12.93$, $p = .005$, $r = .751$).

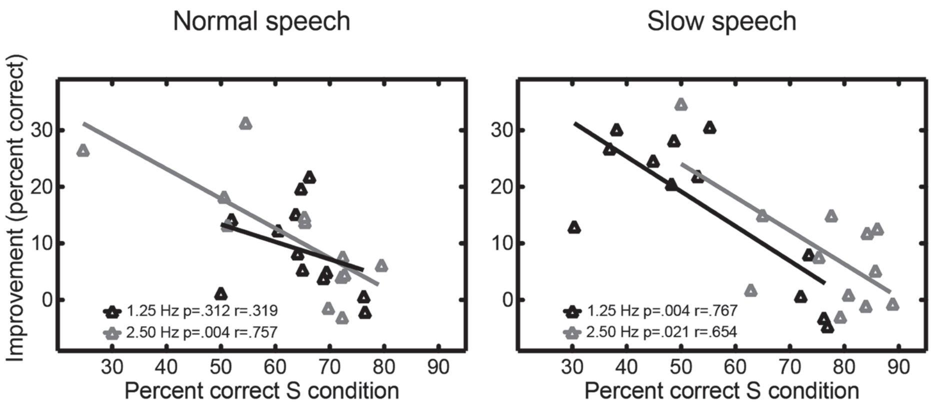


Figure 2: Phonemic restoration benefit shown as function of the intelligibility scores in the silent conditions (S) for the normal speech (left panel) and slow speech (right panel). The black and gray triangles show the results for the slow (1.25 Hz) and fast (2.50 Hz) interruptions, respectively, and the regression lines are shown with corresponding colors.

To compare the present findings with previous literature on phonemic restoration benefit, i.e., the increase in performance as a result of the addition of the filler noise (N), the data were also inspected for this aspect. Figure 2 shows the increase in intelligibility when the silent intervals are filled with noise (N scores minus S scores), plotted against the scores of the interrupted sentences without filler noise (S scores) for normal (left panel) and slow speech (right panel), along with best-fit regression lines. Figure 1 and 2 together confirm the benefit from adding filler noise (better N scores than corresponding S scores in Figure 1 and statistical significant improvement directly shown in Figure 2). A three-factor repeated measure analysis of variance (ANOVA) with the within-subject factors the speed of the sentences (normal and slow), the interruption rate (1.25 and 2.50 Hz) and the addition of filler noise, showed significant differences between the S and N scores ($F(1,11) = 32.86$, $p < .001$, power = .999), and between the slow and fast interruption rates ($F(1,11) = 16.04$, $p = .002$, power = .957). No significant difference was observed between the normal and slow speech conditions ($F(1,11) = 2.74$, $p = .126$, power = .328). There was a significant interaction between the factors speed of the sentences and the interruption rates ($F(1,11) = 32.55$, $p < .001$, power = .999). Further, similar to previous studies (Verschuure and Brocaar 1983; Başkent 2010), the improvement depends on the baseline scores (S conditions, see Figure 2); more benefit is observed at lower S scores.

DISCUSSION

The intelligibility of interrupted sentences, interrupted at two different rates and played at normal and slow speed, was measured with and without filler noise in the silent intervals. Similar to previous studies, phonemic restoration benefit was observed, in a magnitude and trend as was expected from literature (Verschuure and Brocaar 1983; Bashford et al. 1992; Başkent 2010; Başkent 2012; Benard and Başkent 2013). Hence, despite the relatively small number of participants in this study, we can assume that the intelligibility scores of the present study were a good representation of top-down restoration of interrupted speech.

In the present study, using the individual differences for understanding interrupted speech, we explored the involvement of linguistic and cognitive factors on top-down restoration of speech. Our first hypothesis was about linguistic skills, and to test this hypothesis, we assessed the receptive vocabulary and verbal intelligence, by means of the PPVT-III-NL (see Figure 1). Confirming our first hypothesis, the PPVT-III-NL scores showed in general significant correlations with the intelligibility scores of normal and slow speech, with or without filler noise in the silent intervals. Given these correlations, we conclude that the restoration mechanisms of interrupted speech make use of knowledge of receptive vocabulary and verbal intelligence.

A second hypothesis was on cognitive skills, and we used a measure for the full-scale intelligence, namely the WAIS-IV-NL. This hypothesis was not supported by our data. This may mean that linguistic factors are more important for top-down restoration. An alternative explanation is that a full-scale intelligence score is not sufficiently sensitive to capture effects from specific cognitive components, such as working memory (Akeroyd 2008), because of the composite nature of the WAIS-IV-NL. The full-scale intelligence quotient covers verbal comprehension and working memory, assessed with mostly verbal subtests, perhaps the cause of high correlation of the WAIS scores with PPVT scores. However, besides this, perceptual reasoning and processing speed are assessed with mostly non-verbal or visual subtests (Benson et al. 2010). Hence, future research with more statistical power would be needed to determine which major components of intelligence are involved in the comprehension of interrupted speech.

All results combined, our data showed significant correlations between receptive vocabulary and verbal intelligence (PPVT-III-NL) and perception of interrupted speech, possibly implying an involvement of these factors in top-down restoration mechanisms. This knowledge can be useful in developing future training paradigms for older or hearing-impaired people. These can, for example, focus on linguistically relevant training methods (such as crossword puzzles), and may be easier to implement compared to cognitive training methods.

ACKNOWLEDGEMENTS

The authors would like to thank Grietje Bos for valuable feedback on the study design, Jefta Saija for help with the PRAAT speech manipulations, Wydineke Boels for help with collecting data and Anke van der Meijde, Age Hoekstra, Peter van Hengel and Bas Franck for their valuable feedback on earlier versions of this manuscript. This study was supported by Pento Speech and Hearing 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 our department: Healthy Aging and Communication.

PART II

SPECTROTEMPORALLY DEGRADED SPEECH

CHAPTER 4

Perceptual learning of temporally interrupted spectrally degraded speech

M.R. Benard
D. Başkent

Published in modified form:
The Journal of the Acoustical Society of America
2014 | Volume 136 | Issue 3 | 1344-51

ABSTRACT

Normal-hearing (NH) listeners make use of context and 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: i) training with CI simulations of interrupted sentences can teach listeners to use the high-level restoration mechanisms more effectively, ii) phonemic restoration benefit, an increase in intelligibility of interrupted sentences once its silent gaps are filled with noise, can be induced with training, and iii) 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 by 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.

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 (Miller and Licklider 1950; Başkent and Chatterjee 2010), as well as with specific paradigms, such as phonemic restoration (Warren 1970; Powers and Speaks 1973). 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; Verschuure and Brocaar 1983; Srinivasan and Wang 2005). 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; Srinivasan and Wang 2005; Sivonen et al. 2006b), 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 (Warren et al. 1995; Lippmann 1996; Başkent and Shannon 2006) or temporal (Miller and Licklider 1950; Jin and Nelson 2010) 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 temporal fine structure is not fully transmitted (Nelson and Jin 2004; Chatterjee et al. 2010; Başkent and Chatterjee 2010). 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 et al. 2009; Başkent et al. 2010; Başkent 2012; 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 (Stickney et al. 2004; Fu and Nogaki 2005).

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 by a CI vocoder simulator 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; Stacey et al. 2010; Oba et al. 2011; 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 program 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 have specifically 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 (Shannon et al. 1995; Friesen et al. 2001; Başkent 2012). We thirdly hypothesized that the learning effect could be also 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.

MATERIALS AND METHODS

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.

Training with sentences

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).

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 seconds, 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 (Shannon et al. 1995; Başkent 2012). 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 6th 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 3th 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.

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 one meter 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, California, 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.

Speech audiometry with words

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 with 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 a signal-to-noise ratio of 0, 5 and 10 dB, respectively).

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 twenty-four unique words (two lists, randomly selected) per condition.

Overall procedure

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

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

Table 1: 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. B1 and A2 were of the same stimulus type (S or N) as the participants were trained with. 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 signal-to-noise ratios (SNRs).

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.

RESULTS

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 to 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).

Groups	Percent correct scores before training (%)		Percent correct scores after training (%)		Improvement (%)	
	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

Table 2: 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.

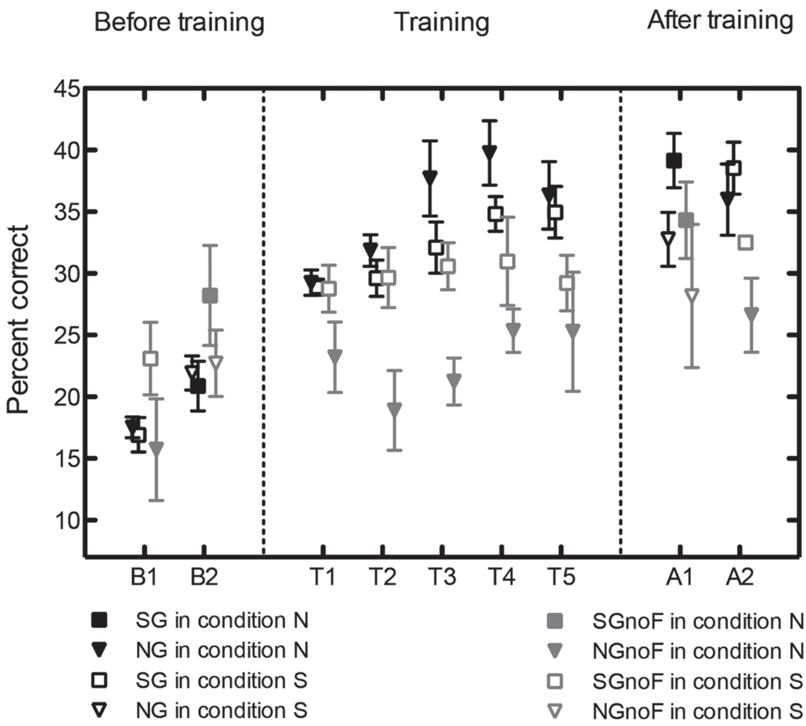


Figure 1: Intelligibility of CI simulations of interrupted speech with and without filler noise. The mean percent correct scores are shown for the four groups of participants: the silence group with (SG) and the silence group without feedback (SGnoF) during the training sessions, the noise group with (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.

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 noise and without feedback) improved significantly after the training sessions (main effect of factor before and after training, $F(1,22) = 79.8$, $p < .001$, $\eta^2 = 0.784$, 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, $F(1,22) = .029$, $p = .866$, $\eta^2 = 0.001$, power = .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, $F(1,22) = .840$, $p = .369$, $\eta^2 = 0.037$, power = .142). However, there was a significant interaction between the factors before and after training and the addition of feedback during training ($F(1,22) = 10.884$, $p = .003$, power = .833).

The effect of training (Hypotheses 1 and 2) was also analyzed by comparing performance across the training sessions T1 to 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, $F(4,17) = 4.772$, $p = .009$, $\eta^2 = 0.209$, power = .878). The groups tested without feedback scored significantly less than the groups tested with feedback (main effect of feedback $F(1,20) = 19.8$, $p < .001$, $\eta^2 = 0.496$, power = .988). The addition of filler noise did not influence the performance (main effect of factor the addition of the filler noise, $F(1,20) = 1.636$, $p = .215$, $\eta^2 = 0.076$, power = .230). There were no significant interactions between factors.

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, $F(2,19) = 28.9$, $p < .001$, $\eta^2 = 0.680$, 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, $F(1,20) = .373$, $p = .548$, $\eta^2 = 0.018$, power = .090). The word identification by the two noise

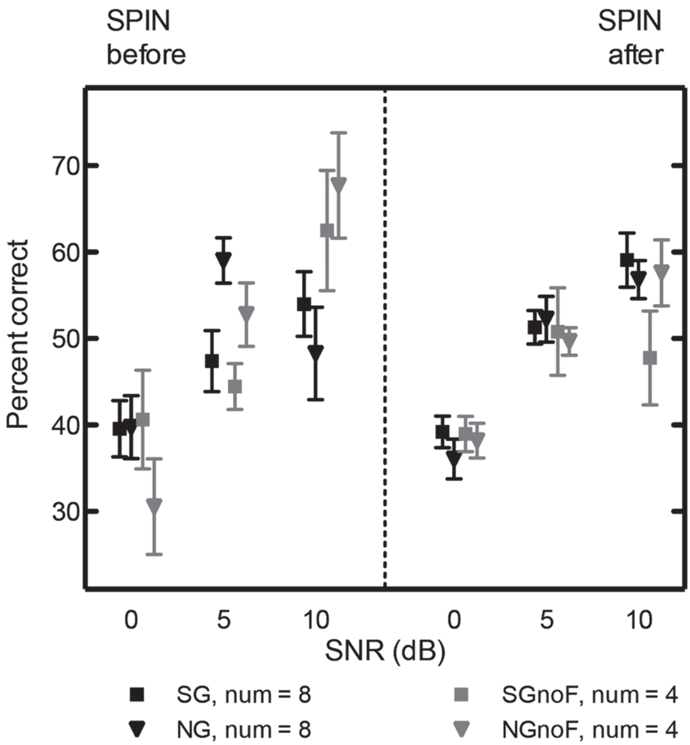


Figure 2: The mean percent correct scores shown for the word-in-noise identification measurements of the CI simulated consonant-vowel-consonant words presented in noise, for the four different listeners groups and at different signal-to-noise ratios (SNRs). The dark and the light gray squares represent 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.

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, $F(1,20) = .332$, $p = .571$, $\eta^2 = 0.016$, power = .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, $F(1,20) = .002$, $p = .961$, $\eta^2 < 0.001$, power = .050). There were no significant interactions between factors.

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; Nogaki et al. 2007; Loebach and Pisoni 2008; Hervais-Adelman et al. 2008; Stacey et al. 2010; Oba et al. 2011; Smalt et al. 2011; Zhang et al. 2012; Ingvalson et al. 2013). Given that CI users greatly suffer from difficulties understanding speech degraded in any form, either due to background noise (Stickney et al. 2004; Fu and Nogaki 2005), or due to interruptions (Nelson and Jin 2004; Bhargava et al. 2014), 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 (Chatterjee et al. 2010; Başkent and Chatterjee 2010; Başkent 2012). 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 = .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 = .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 (Verschuure and Brocaar 1983; Repp 1992; Srinivasan and Wang 2005; Başkent et al. 2009; Benard and Başkent 2013). 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 (Chatterjee et al. 2010; Başkent 2012) 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 $f = .97$). 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 $f = .60$). However, a post-hoc power analysis of the three-factor repeated measures ANOVA showed a very low effect size ($f = .04$), 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 (Srinivasan and Wang 2005; Bhargava et al. 2014). 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 (Gonzalez and Oliver 2005; Fuller et al. 2014a). 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 (Verschuure and Brocaar 1983; Bronkhorst et al. 1993; Sivonen et al. 2006b; Grossberg and Kazerounian 2011). This observation could also indicate that linguistic factors beyond word level, such as syntactic and semantic patterns, may be of even more relevance for the top-down restoration mechanisms. 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; Nogaki et al. 2007; Loebach and Pisoni 2008; Hervais-Adelman et al. 2008; Stacey et al. 2010; Oba et al. 2011; Smalt et al. 2011; Zhang et al. 2012; Ingvalson et al. 2013). 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 (Fu and Galvin 2003; Stacey and Summerfield 2008; Fu and Galvin 2008; Başkent 2012).

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 Speech and Hearing 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.

CHAPTER 5

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

M.R. Benard
D. Başkent

This chapter is published in:
Hearing Research
2015 | Volume 328 | 24-33

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.

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 & Nelson, 2006; Verschuure & 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 (Srinivasan and Wang 2005; Riecke et al. 2009; Riecke et al. 2011) but it also improves the intelligibility (Powers and Wilcox 1977; Bashford and Warren 1979; Verschuure and Brocaar 1983; Bhargava and Başkent 2012; Benard and Başkent 2013). Cognitive factors and linguistic skills (Warren 1983; Bashford et al. 1992; Bronkhorst et al. 1993; Srinivasan and Wang 2005; Sivonen et al. 2006b; Wang and Humes 2010; Saija et al. 2013) 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 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 & MacDonald, 1976; Valkenier, Duyne, Andringa, & Başkent, 2012; Wiersinga-Post et al., 2010). 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 (Luo et al. 2010; Cunillera 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 (McGrath and Summerfield 1985; Grant et al. 1998; Grant and Seitz 2000; Ross et al. 2007; Başkent and Bazo 2011; Gilbert et al. 2012). 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 (Trout and Poser 1990; Shahin and Miller 2009; Shahin et al. 2012; Bhat et al. 2014). 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 in interrupted speech (Trout and Poser 1990; Shahin and Miller 2009).

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).

EXPERIMENT 1: PERCEPTION OF TEMPORALLY INTERRUPTED SPEECH WITH VISUAL CUES

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.

MATERIALS AND METHODS

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.

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 to 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.

Figure 1: Example frame of the video of the speaker.

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 (Figure 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 seconds 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 minutes of recording time. The recording session lasted approximately 3 hours. 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 seconds. 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™.

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. (2000).

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 in-phase 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).

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, 2014; Benard et al., 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.

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 minutes.

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 hour, including one short break.

RESULTS

Figure 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).

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

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.

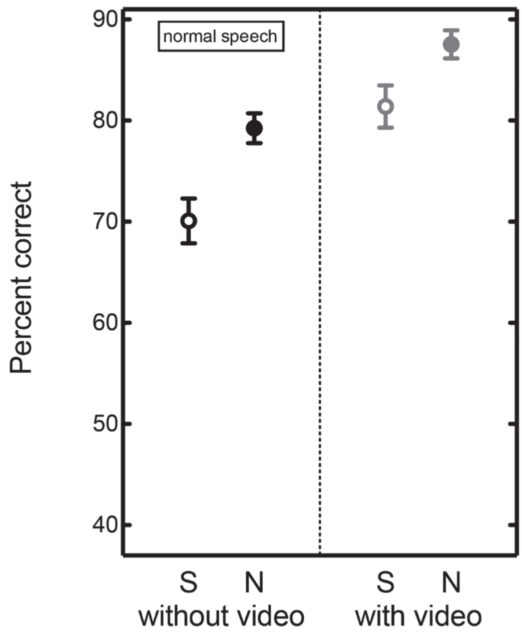


Figure 2: Intelligibility scores of interrupted speech, with and without filler noise and video of the speaker. The mean percent correct scores of interrupted speech without further spectrotemporal degradations (Experiment 1) are shown. The open and filled symbols represent the intelligibility scores 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.

A two factor 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.

DISCUSSION EXPERIMENT 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 (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 (McGrath and Summerfield 1985; Ross et al. 2007; Başkent and Bazo 2011; Gilbert et al. 2012). 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 (Luo et al. 2010; Cunillera 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 (Shahin et al. 2012; Bhat et al. 2014).

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; Sivonen et al. 2006b) 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.

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 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 did not 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, Bishop and Miller (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.

EXPERIMENT 2: INTERRUPTED SPEECH, SPECTROTEMPORALLY DEGRADED WITH CI SIMULATIONS

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 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 (Chatterjee et al. 2010; Başkent 2012; Benard and Başkent 2014; Bhargava et al. 2014). Facilitating speech segmentation and temporal envelope tracking could help the listeners to better segregate speech cues from noise segments (Luo et al. 2010; Cunillera et al. 2010; Mitchel and Weiss 2013; 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 spectrally degraded sentences with interruptions.

MATERIALS AND METHODS

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.

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 (Shannon et al. 1995; Chatterjee et al. 2010; Başkent and Chatterjee 2010; Başkent 2012; Benard and Başkent 2014; Bhargava et al. 2014). 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.

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 hours, 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.

RESULTS

Figure 3 shows the percent correct scores of Experiment 2. Similar to Figure 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 the N condition.

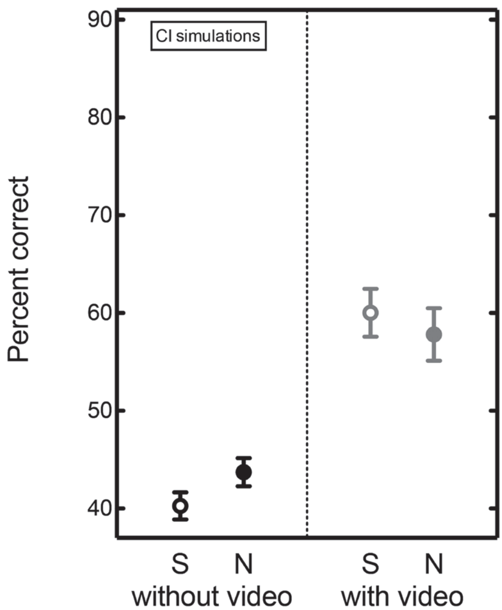


Figure 3: Intelligibility scores of CI simulations of interrupted speech, with and without filler noise and video of the speaker. Similar to Fig. 2, except the scores shown are the result of CI simulated interrupted speech by means of an eight channel noiseband vocoder (Experiment 2).

A two factor 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; $F(1,11) = 0.079$, $p = .784$, 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).

DISCUSSION EXPERIMENT 2

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 & 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 (Nelson and Jin 2004; Bhargava et al. 2014). 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.

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 (Chatterjee et al. 2010; Başkent 2012; Benard and Başkent 2014; Bhargava et al. 2014). 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 (Srinivasan and Wang 2005; Başkent 2012; Benard and Başkent 2014). 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 (Luo et al. 2010; Cunillera 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 (Srinivasan and Wang 2005; Bhargava et al. 2014) 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.

EXPERIMENT 3: INTERRUPTED SPEECH, SPECTROTEMPORALLY DEGRADED WITH EAS SIMULATIONS

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 (Kong and Carlyon 2007; Incerti et al. 2013) 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.

MATERIALS AND METHODS

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.

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.

PROCEDURE

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

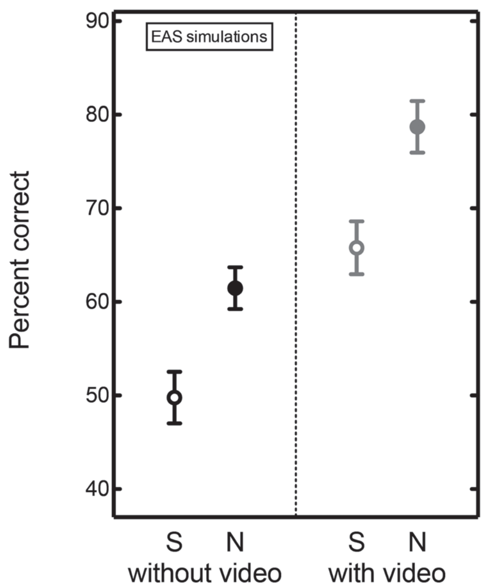


Figure 4: Intelligibility scores of EAS simulations of interrupted speech, with and without filler noise and video of the speaker. Similar to Fig. 2, except the scores shown are the result of simulations of electric-acoustic stimulation of interrupted speech (Experiment 3).

RESULTS

Figure 4 shows the percent correct scores of Experiment 3, presented similar to Figure 3.

A two factor 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.

DISCUSSION EXPERIMENT 3

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 (Kong and Carlyon 2007; Brown and Bacon 2009a). Compared to the simulations of electric stimulation alone (Experiment 2), we observed an 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 (Kong and Carlyon 2007; Incerti et al. 2013).

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 conditions, 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 (Kong and Carlyon 2007; Brown and Bacon 2009b; Zhang et al. 2010; Incerti et al. 2013).

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 (Srinivasan and Wang 2005; Bhargava et al. 2014), inducing a phonemic restoration benefit. Some studies have previously explored the role of F0 on phonemic restoration directly. Clarke, Gaudrain, Chatterjee and Başkent (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. 2014b). 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.

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 (Rouger et al. 2007; Lyness et al. 2013).

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 (Luo et al. 2010; Cunillera 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 (Saija et al. 2013; Benard et al. 2014) 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.

ACKNOWLEDGEMENTS

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 Speech and Hearing 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.

CHAPTER 6

General conclusions

INTRODUCTION

In total one-hundred-and-two participants were trained and tested to answer our research questions and to explore our hypotheses. This research was carried out to investigate how cognitive restoration mechanisms are involved in the perception of degraded speech and what the implications are for improving speech perception of CI users. As defined in the introduction, the hope that drives the research presented in this thesis is that hearing-impaired listeners could learn with training to better apply cognitive restoration mechanisms to better understand speech in complex acoustic environments. To achieve this hope, the main aim of the research presented in this thesis was to understand more of the cognitive factors that influence, or are involved in restoring degraded speech in such a way that the brain can form it into an intelligible speech stream. The four studies presented in this thesis have several factors and observations in common which I would like to discuss.

PERCEPTUAL LEARNING OF INTERRUPTED SPEECH

First of all, the overall results showed that understanding interrupted speech can be learned. Especially chapters 2 and 4 are dedicated to this observation. Learning was observed even when further spectral degradations, by means of a noise band vocoder, were applied to interrupted speech. In chapter 2 the interest in learning to better understand interrupted speech stems from the hypothesis that if two different cognitive mechanisms were involved in the restoration of interrupted speech with and without filler noise, this would lead to different trends in learning curves. Training the participants with these kinds of stimuli led to the observation that even after intensive training phonemic restoration benefit of filler noise does not disappear. The learning curves follow a similar trend for interrupted speech with and without filler noise; there is a stabilization of the intelligibility after roughly the third training session. The similarity in learning trends indicates that similar cognitive restoration mechanisms are active in the restoration of interrupted speech, with or without the filler noise. In chapter 4 spectrotemporally degraded speech is used as stimulus, with and without spectrally degraded filler noise in the silent intervals. Here, the hypothesis was that training with CI simulations of interrupted sentences might teach listeners to use the high-level restoration mechanisms more effectively with less speech information available. This study shows similar learning curves for both spectrally degraded stimuli, from which it can be concluded that understanding spectrally degraded speech can be learned as well as understanding interrupted speech without spectral degradation. However, due to the unfamiliarity of the NH participants with spectrally degraded speech, there is a steeper learning curve, compared to learning without spectral degradation. Training with combined written and auditory feedback adds to the perceptual learning of participants to understand spectrotemporally-degraded speech. Perceptual learning to understand interrupted speech was not a main topic of research in chapters 3 and 5 per se. However, it was demonstrated that understanding slow speed interrupted speech could be learned, as can understanding simulations of Electric Acoustic Stimulation, respectively. *In conclusion*, understanding interrupted speech can be learned to some extent regardless of additional spectral degradations, manipulation of the speed of the sentences, or addition of filler noise in the silent intervals. The trends of the learning curves do not depend on the addition of filler noise. Providing feedback seems to lead to effective training with interrupted speech.

ECOLOGICAL VALIDITY

The basic functioning of the peripheral auditory system is commonly tested by using simple sounds that are often not ecologically valid. For example, pure tones are – except from electric equipment – not commonly present in the natural environments of humans. Testing hearing thresholds or measuring acoustic reflexes can be performed under the right conditions with pure tones, as these tones are easy to control for in terms of their acoustic properties. However, to gain more insight into cognitive auditory processing, ecologically valid stimuli should be used since top-down features are often involved in real-life listening. For example, auditory attention and cognitive and linguistic skills play an important role in speech perception in general, as well as in the restoration of degraded speech in particular. I argue that speech streams interrupted by (periodic) noise are ecologically valid since similar interruptions occur in daily life as well. Closing doors, moving chairs and other non-speech signals frequently interrupt everyday conversations. However, in contrast with interruptions by these auditory streams, the silent intervals do not form a correlated auditory stream. It has been suggested that interrupted speech with silent intervals has no ecological validity, and that phonemic restoration was simply a consequence of this. More specifically, the benefit of filler noise in the silent intervals was suggested not be due to the additive nature of filler noise, but that it is rather the unfamiliarity with interrupted speech with silent intervals that results in poor performance. Chapter 2 provides the evidence contrary to this suggestion. Participants were trained intensively with interrupted speech with silent intervals and the intelligibility increased. After reaching maximum performance the addition of filler noise still improved the performance as much as before training. Furthermore, the trends of the learning curves are similar for both forms of interruptions and do not depend on the addition of filler noise or on additional spectral degradations, as described in chapter 4 with CI simulations of interrupted speech with and without filler noise. *In conclusion,* the benefit of the filler noise does not seem to be a side effect of participants being tested with the less ecologically valid interrupted speech with silent intervals.

PERSISTENCY OF THE BENEFIT OF FILLER NOISE

In the studies presented in this thesis the benefit of filler noise is shown to be persistent. Chapter 2 shows that the addition of filler noise improves intelligibility not only before, but also during and after intensive training. Results from follow-up measurements from (on average) 3 months after the last test session show that this benefit is persistent. Chapter 3 shows that the benefit of the filler noise depends on the baseline scores. It has been observed that in situations with low intelligibility scores of interrupted speech without filler noise there is more benefit when the filler noise is added, but this is not very surprising by itself as there is more room for improvement. Slowing down the speed of the sentences with conserved pitch does not diminish the benefit of the filler noise. Chapter 5 shows that adding an accompanying video of the speaker does not suppress a restoration benefit of the filler noise. *In conclusion,* the benefit of filler noise is persistent: long-term training, playback-rate manipulation and the addition of an accompanying video of the speaker does not influence the benefit of filler noise in interrupted speech, in the lack of any further spectral degradations. The phonemic restoration benefit can therefore be seen as hardwired in the auditory system.

However, different behavior of intelligibility scores is observed when these speech materials are spectrotemporally degraded as can happen in CI speech processing. Chapter 4 shows the results of intensive training of the participants with CI simulations by means of noiseband vocoding. No benefit of the filler noise is observed before training, and it also did not appear during and after intensive training with active feedback. Indeed using acoustic simulations of CIs, Başkent (2012) suggested 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. In Chapter 5, this finding is confirmed. The addition of the video of the speaker did increase the intelligibility scores to the level at which a benefit of the filler noise is normally observed in interrupted speech without spectral degradations. Even with such high intelligibility scores a restoration benefit of the filler noise in CI simulations was not observed. An improved speech perception and restoration benefit was observed with simulations of EAS, where the additional acoustic low-frequency speech cues seemed to help in the restoration of interrupted speech. This benefit of the filler noise is observed with and without additional visual cues. *In conclusion,* if a restoration benefit of the filler noise is not present in spectrotemporally degraded speech, a restoration benefit cannot be induced with training or with additional speech cues such as a video of the speaker. However, the overall perception of degraded speech, even in multiple degradations of temporal interruptions and spectrotemporal degradation of CI simulation, seems to be improvable with training.

TRAINING PROGRAMS WITH INTERRUPTED SPEECH

The main question of this thesis was: what happens in the process of restoring interrupted speech? In other words, what are the mechanisms behind phonemic restoration? The ultimate aim of performing the present studies is to produce knowledge that can eventually contribute to increasing speech intelligibility performance of CI users and this thesis contains a few 'good news' massages for CI users. While speech perception in quiet is usually good, understanding speech obliterated by noise or in competing background sounds is still challenging for CI users. From the signal processing point of view, significant effort has been put into research to improve the intelligibility of speech in complex listening situations by improving the sound processors, the electrode design, and the surgical techniques of CIs. From the user point of view, there is still a need for effective training programs to increase the perception of speech in these situations, ideally towards the level of normal-hearing listeners. As described in Chapter 4, there are few studies on training CI users with speech in noise. Chapter 3 discusses the significant correlations between receptive vocabulary and verbal intelligence on one side and the perception of interrupted speech on the other side. This finding implies an important involvement of linguistic skills in top-down restoration mechanisms, more so than the involvement of cognitive skills. This is good news for the development of future training paradigms since linguistic training methods might be easier to implement compared to cognitive training methods; these can perhaps even be achieved with self-teaching, for example, by reading many books to increase one's lexicon. Chapter 4 describes that the feedback used for training with the CI simulations was more efficient than presenting intact auditory feedback alone. The feedback used contained the playback of the degraded signal accompanied with the written sentence on monitor. The good news of this method of providing feedback is that it applies well to CI users since there is no possibility to present them with spectrally intact auditory feedback. Maybe future gadgets such as Google Glass can make it possible to present the spoken sentence to CI users visually, giving even more meaning to the feedback and training methods presented in this thesis. Chapter 5 shows that participants benefit from additional visual speech cues even for the interrupted sentences that are also spectrally degraded. This is good news for CI users, since visual cues are often available in daily speech communication in the form of lip-reading.

CONCLUSION: MECHANISMS OF TOP-DOWN RESTORATION OF DEGRADED SPEECH

All results combined, I conclude that both top-down and bottom-up processes play an important role in the restoration of interrupted speech. Especially high-level linguistic mechanisms seem to have a large influence on the restoration of interrupted speech. Receptive vocabulary and verbal intelligence are shown to be significant predictors of successful restoration of interrupted sentences without spectral degradations. These *top-down* restoration mechanisms are shown to be less effective if the *bottom-up* auditory signal is of insufficient quality (as occurs in CI speech processing).

Implications for CI users

Our overall results suggest that better perception of interrupted speech can indeed be achieved via training, even with spectrotemporal degradations of CI speech transmission. Since linguistic skills play an important role in the restoration of spectrally degraded interrupted speech, CI users can possibly be trained to improve their linguistic skills by reading books or solving crossword puzzles. Furthermore, providing relatively simple feedback, even the text of the sentence, seems to be an effective feedback to lead to successful learning. Finally, lip-reading aids in speech perception of interrupted speech and is often available in daily speech communication for CI users.

REFERENCES

Adank P, Janse E (2009) Perceptual learning of time-compressed and natural fast speech. J Acoust Soc Am 126:2649–59. doi: 10.1121/1.3216914

Akeroyd MA (2008) Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int J Audiol 47 Suppl 2:S53–71. doi: 10.1080/14992020802301142

Alain C, Arnott SR, Picton TW (2001) Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence From Event-Related Brain Potentials. J Exp Psychol Hum Percept Perform 27:1072–1089. doi: 10.1037//TO96-1523.27

Bashford JA, Riener KR, Warren RM (1992) Increasing the intelligibility of speech through multiple phonemic restorations. Percept Psychophys 51:211–7.

Bashford JA, Warren RM (1987) Effects of spectral alternation on the intelligibility of words and sentences. Percept Psychophys 42:431–8.

Bashford JA, Warren RM (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. doi: 10.1007/s10162-012-0334-3

Başkent D (2010) Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores. J Acoust Soc Am 128:EL169–74. doi: 10.1121/1.3475794

Başkent D, Bazo D (2011) Audiovisual Asynchrony Detection and Speech Intelligibility in Noise With Moderate to Severe Sensorineural Hearing Impairment. Ear Hear. doi: 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 270:127–33. doi: 10.1016/j.heares.2010.08.011

Başkent D, Eiler CL, Edwards B (2010) Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. Hear Res 260:54–62. doi: 10.1016/j.heares.2009.11.007

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

Başkent D, Shannon RV (2006) Frequency transposition around dead regions simulated with a noiseband vocoder. J Acoust Soc Am 119:1156–63. doi: 10.1121/1.2151825

Bell NL, Lassiter KS, Matthews TD, Hutchinson MB (2001) Comparison of the Peabody Picture Vocabulary Test-Third Edition and Wechsler Adult Intelligence Scale-Third Edition with university students. J Clin Psychol 57:417–22.

Benard MR, Başkent D (2013) Perceptual Learning of Interrupted Speech. PLoS One 8:e58149. doi: 10.1371/journal.pone.0058149

Benard MR, Başkent D (2014) Perceptual learning of temporally interrupted spectrally degraded speech. J Acoust Soc Am 136:1344–51.

Benard MR, Mensink JS, 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. doi: 10.1121/1.4862879

Benson N, Hulac DM, Kranzler JH (2010) Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): what does the WAIS-IV measure? Psychol Assess 22:121–30. doi: 10.1037/a0017767

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

Bhargava P, Gaudrain E, Başkent D (2014) Top-down restoration of speech in cochlear-implant users. Hear Res 309:113–23. doi: 10.1016/j.heares.2013.12.003

Bhat J, Pitt MA, Shahin AJ (2014) Visual context due to speech-reading suppresses the auditory response to acoustic interruptions in speech. Front Neurosci 8:173. doi: 10.3389/fnins.2014.00173

Boothroyd A, Nittrouer S (1988) Mathematical treatment of context effects in phoneme and word recognition. J Acoust Soc Am 84:101–14.

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

Bosman A, Smoorenburg GF (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 AS (1994) Auditory Scene Analysis - The Perceptual Organization of Sound. The MIT press, Cambridge, Massachusetts

Brennan S, Schober M (2001) How Listeners Compensate for Disfluencies in Spontaneous Speech. J Mem Lang 44:274–296. doi: 10.1006/jmla.2000.2753

Bronkhorst AW, Bosman A, Smoorenburg GF (1993) A model for context effects in speech recognition. J Acoust Soc Am 93:499–509.

Brown CA, Bacon SP (2009a) Low-frequency speech cues and simulated electric-acoustic hearing. J Acoust Soc Am 125:1658–65. doi: 10.1121/1.3068441

Brown CA, Bacon SP (2009b) Achieving electric-acoustic benefit with a modulated tone. Ear Hear 30:489–93. doi: 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. doi: 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 315:80–7. doi: 10.1016/j.heares.2014.07.002

Cohen J, Dunbar K, McClelland J (1990) On the control of automatic processes: A parallel distributed processing account of the Stroop effect. Psychol Rev 97:332–61.

Cooke M (2006) A glimpsing model of speech perception in noise. J Acoust Soc Am 119:1562–73. doi: 10.1121/1.2166600

Cooke M, Green P, Josifovski L, Vizinho A (2001) Robust automatic speech recognition with missing and unreliable acoustic data. Speech Commun 34:267–285. doi: 10.1016/S0167-6393(00)00034-0

Cunillera T, Camara E, Laine M, Rodriguez-Fornells A (2010) Speech segmentation is facilitated by visual cues. Q J Exp Psychol (Hove) 63:260–74. doi: 10.1080/17470210902888809

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132–47. doi: 10.1016/j.heares.2007.01.014

Davis MH, Johnsrude IS, Hervais-adelman A, et al. (2005) Lexical Information Drives Perceptual Learning of Distorted Speech : Evidence From the Comprehension of Noise-Vocoded Sentences. J Exp Psychol Gen 134:222–241. doi: 10.1037/0096-3445.134.2.222

Doucet ME, Bergeron F, Lassonde M, et al. (2006) Cross-modal reorganization and speech perception in cochlear implant users. Brain 129:3376–83. doi: 10.1093/brain/awl264

Eisner F, McQueen JM (2006) Perceptual learning in speech: Stability over time. J Acoust Soc Am 119:1950–1953. doi: 10.1121/1.2178721

Fahle M (2004) Perceptual learning: A case for early selection. J Vis 4:879–890. doi: 10.1167/4.10.4

Francis AL, Baldwin K, Nusbaum HC (2000) Effects of training on attention to acoustic cues. Percept Psychophys 62:1668–80.

Francis AL, Nusbaum HC (2009) Effects of intelligibility on working memory demand for speech perception. Atten Percept Psychophys 71:1360–74. doi: 10.3758/APP.71.6.1360

Francis AL, Nusbaum HC, Fenn K (2007) Effects of training on the acoustic phonetic representation of synthetic speech. J Speech Lang Hear Res 50:1445–65. doi: 10.1044/1092-4388(2007/100)

Friesen LM, Shannon RV, Başkent D, 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 110:1150–63. doi: 10.1121/1.1381538

Fu Q-J, Galvin JJ (2003) The effects of short-term training for spectrally mismatched noise-band speech. J Acoust Soc Am 113:1065–72. doi: 10.1121/1.1537708

Fu Q-J, Galvin JJ (2008) Maximizing cochlear implant patients' performance with advanced speech training procedures. Hear Res 242:198–208. doi: 10.1016/j.heares.2007.11.010

Fu Q-J, Galvin JJ, Wang X, Nogaki G (2005) Moderate auditory training can improve speech performance of adult cochlear implant patients. Acoust Res Lett Online 6:106–111. doi: 10.1121/1.1898345

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. doi: 10.1007/s10162-004-5024-3

Fuller C, Galvin JJ, Free RH, Başkent D (2014a) Musician effect in cochlear implant simulated gender categorization. J Acoust Soc Am 135:EL159–65. doi: 10.1121/1.4865263

Fuller C, Gaudrain E, Clarke J, et al. (2014b) Gender Categorization Is Abnormal in Cochlear Implant Users. J Assoc Res Otolaryngol 15:1037–1048. doi: 10.1007/s10162-014-0483-7

Gilbert G, Bergeras I, Voillery D, Lorenzi C (2007) Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. J Acoust Soc Am 122:1336–1339. doi: 10.1121/1.2756161

Gilbert JL, Lansing CR, Garnsey SM (2012) Seeing facial motion affects auditory processing in noise. Atten Percept Psychophys 74:1761–81. doi: 10.3758/s13414-012-0375-z

Gnansia D, Pressnitzer D, Péan V, et al. (2010) Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. Hear Res 265:46–53. doi: 10.1016/j.heares.2010.02.012

Goldstone RL (1998) Perceptual learning. Annu Rev Psychol 49:585–612. doi: 10.1146/annurev.psych.49.1.585

Gonzalez J, Oliver JC (2005) Gender and speaker identification as a function of the number of channels in spectrally reduced speech. J Acoust Soc Am 118:461–70. doi: 10.1121/1.1928892

Grant KW, Seitz PF (2000) The use of visible speech cues for improving auditory detection. J Acoust Soc Am 108:1197–1208.

Grant KW, Walden BE, Seitz PF (1998) Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. J Acoust Soc Am 103:2677–90.

Greenwood DD (1990) A cochlear frequency-position function for several species-29 years later. J Acoust Soc Am 87:2592–605.

Grossberg S, Kazerounian S (2011) Laminar cortical dynamics of conscious speech perception: neural model of phonemic restoration using subsequent context in noise. J Acoust Soc Am 130:440–60. doi: 10.1121/1.3589258

Hervais-Adelman A, Davis MH, Johnsrude IS, Carlyon RP (2008) Perceptual learning of noise vocoded words: effects of feedback and lexicality. J Exp Psychol Hum Percept Perform 34:460–74. doi: 10.1037/0096-1523.34.2.460

Huggins AWF (1964) Distortion of the temporal pattern of speech: interruptions and alternation. J Acoust Soc Am 36:1055–1064.

Incerti PV, Ching TYC, Cowan R (2013) A systematic review of electric-acoustic stimulation: device fitting ranges, outcomes, and clinical fitting practices. Trends Amplif 17:3–26. doi: 10.1177/1084713813480857

Ingvalson EM, Lee B, Fiebig P, Wong PCM (2013) The Effects of Short-Term Computerized Speech-in-Noise Training on Postlingually Deafened Adult Cochlear Implant Recipients. 56:81–89. doi: 10.1044/1092-4388(2012/11-0291)b

Janse E, Ernestus M (2011) The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. J Phon 39:330–343. doi: 10.1016/j.wocn.2011.03.005

Jin SH, Nelson PB (2010) Interrupted speech perception: the effects of hearing sensitivity and frequency resolution. J Acoust Soc Am 128:881–9. doi: 10.1121/1.3458851

Jin SH, Nelson PB (2006) Speech perception in gated noise: The effects of temporal resolution. J Acoust Soc Am 119:3097–3108. doi: 10.1121/1.2188688

Kashino M (2006) Phonemic restoration: The brain creates missing speech sounds. Acoust Sci Technol 27:318–321. doi: 10.1250/ast.27.318

Kong Y-Y, Carlyon RP (2007) Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. J Acoust Soc Am 121:3717–27. doi: 10.1121/1.2717408

Lippmann R (1996) Accurate consonant perception without mid-frequency speech energy - Speech and Audio Processing, IEEE Transactions on. 4:66–69.

Loebach JL, Pisoni DB (2008) Perceptual learning of spectrally degraded speech and environmental sounds. J Acoust Soc Am 123:1126–39. doi: 10.1121/1.2823453

Loebach JL, Pisoni DB, Svirsky MA (2009) Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: implications for cochlear implants. Ear Hear 30:662–74. doi: 10.1097/AUD.0b013e3181b9c92d

Loebach JL, Pisoni DB, Svirsky MA (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 36:224–34. doi: 10.1037/a0017609

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. doi: 10.1371/journal.pbio.1000445

Lyness CR, Woll B, Campbell R, Cardin V (2013) How does visual language affect crossmodal plasticity and cochlear implant success? Neurosci Biobehav Rev 37:2621–30. doi: 10.1016/j.neubiorev.2013.08.011

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D (2003) The time course of spoken word learning and recognition: Studies with artificial lexicons. J Exp Psychol Gen 132:202–227. doi: 10.1037/0096-3445.132.2.202

Mattys S, Davis M, Bradlow A, Scott SK (2012) Speech recognition in adverse conditions: A review. Lang Cogn Process 27:953–78. doi: 10.1080/01690965.2012.705006

Maxwell C (1978) Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. Br J Clin Pharmacol 6:15–24.

McGrath M, Summerfield AQ (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. J Acoust Soc Am 77:678–85.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–48.

Miller GA, Licklider J (1950) The intelligibility of interrupted speech. J Acoust Soc Am 22:167–173.

Miranda PC, Sampaio ALL, Lopes RAF, et al. (2014) Hearing preservation in cochlear implant surgery. Int J Otolaryngol 2014:468515. doi: 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. doi: 10.1080/01690965.2013.791703.Visual

Montgomery AA, Walden BE, Schwartz DM, Prosek RA (1984) Training auditory-visual speech reception in adults with moderate sensorineural hearing loss. Ear Hear 5:30–36.

Nelson PB, Jin SH (2004) Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. J Acoust Soc Am 115:2286–2294. doi: 10.1121/1.1703538

Nogaki G, Fu Q-J, Galvin JJ (2007) Effect of training rate on recognition of spectrally shifted speech. Ear Hear 28:132–40. doi: 10.1097/AUD.0b013e3180312669

Nygaard LC, Pisoni DB (1998) Talker-specific learning in speech perception. Percept Psychophys 60:355–76.

Oba SI, Fu Q-J, Galvin JJ (2011) Digit training in noise can improve cochlear implant users' speech understanding in noise. Ear Hear 32:573–81. doi: 10.1097/AUD.0b013e31820fc821

Paas FG, Van Merriënboer J, Adam JJ (1994) Measurement of cognitive load in instructional research. Percept Mot Skills 79:419–430.

Powers GL, Speaks C (1973) Intelligibility of Temporally Interrupted Speech. J Acoust Soc Am 50:661–667. doi: 10.1121/1.1977590

Powers GL, Wilcox JC (1977) Intelligibility of temporally interrupted speech with and without intervening noise. J Acoust Soc Am 61:195–199.

Qin MK, Oxenham AJ (2006) Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. J Acoust Soc Am 119:2417. doi: 10.1121/1.2178719

Repp BH (1992) Perceptual restoration of a "missing" speech sound: auditory induction or illusion? Percept Psychophys 51:14–32.

Riecke L, Mendelsohn D, Schreiner CS, Formisano E (2009) The continuity illusion adapts to the auditory scene. Hear Res 247:71–7. doi: 10.1016/j.heares.2008.10.006

Riecke L, Micheyl C, Vanbussel M, et al. (2011) Recalibration of the auditory continuity illusion: sensory and decisional effects. Hear Res 277:152–62. doi: 10.1016/j.heares.2011.01.013

Riecke L, van Opstal AJ, Goebel R, Formisano E (2007) Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. J Neurosci Off J Soc Neurosci 27:12684–9. doi: 10.1523/JNEUROSCI.2713-07.2007

Ross LA, Saint-Amour D, Leavitt VM, et al. (2007) Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cereb Cortex 17:1147–53. doi: 10.1093/cercor/bhl024

Rouger J, Lagleyre S, Fraysse B, et al. (2007) Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proc Natl Acad Sci U S A 104:7295–300. doi: 10.1073/pnas.0609419104

Saija JD, Akyürek EG, Andringa TC, Başkent D (2013) Perceptual Restoration of Degraded Speech Is Preserved with Advancing Age. J Assoc Res Otolaryngol. doi: 10.1007/s10162-013-0422-z

Sampaio ALL, Araújo MFS, Oliveira CACP (2011) New criteria of indication and selection of patients to cochlear implant. Int J Otolaryngol 2011:573968. doi: 10.1155/2011/573968

Samuel AG (1997) Lexical activation produces potent phonemic percepts. Cogn Psychol 32:97–127. doi: 10.1006/cogp.1997.0646

Schlichting L (2005) Peabody picture vocabulary test, 3th edition (PPVT-III-NL) Dutch version.

Schwab EC, Nusbaum HC, Pisoni DB (1985) Some effects of training on the perception of synthetic speech. Hum Factors 27:395–408.

Seitz AR, Vlahou EL, Gori S (2011) Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. Cognition 115:435–443. doi: doi:10.1016/j.cognition.2010.03.004

Shahin AJ, Bishop CW, Miller LM (2009) Neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44:1133–43. doi: 10.1016/j.neuroimage.2008.09.045

Shahin AJ, Kerlin JR, Bhat J, Miller LM (2012) Neural restoration of degraded audiovisual speech. Neuroimage 60:530–8. doi: 10.1016/j.neuroimage.2011.11.097

Shahin AJ, Miller LM (2009) Multisensory integration enhances phonemic restoration. J Acoust Soc Am 125:1744–50. doi: 10.1121/1.3075576

Shannon RV, Zeng FG, Kamath V, et al. (1995) Speech recognition with primarily temporal cues. Science 270:303–4.

Shiffrin RM, Schneider W (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychol Rev 84:127–190.

Shinn-Cunningham BG, Wang D (2008) Influences of auditory object formation on phonemic restoration. J Acoust Soc Am 123:295–301. doi: 10.1121/1.2804701

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

Sivonen P, Maess B, Lattner S, Friederici AD (2006b) Phonemic restoration in a sentence context: Evidence from early and late ERP effects. Brain Res 1121:177–189. doi: 10.1016/j.brainres.2006.08.123

Smalt CJ, Talavage TM, Pisoni DB, Svirsky MA (2011) Neural adaptation and perceptual learning using a portable real-time cochlear implant simulator in natural environments. 33rd Annu Int Conf IEEE EMBS 2011:1145–8. doi: 10.1109/IEMBS.2011.6090268

Song J-J, Lee H-J, Kang H, et al. (2014) Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. Brain Struct Funct. doi: 10.1007/s00429-013-0704-6

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

Stacey PC, Raine CH, O'Donoghue GM, et al. (2010) Effectiveness of computer-based auditory training for adult users of cochlear implants. Int J Audiol 49:347–56. doi: 10.3109/14992020903397838

Stacey PC, Summerfield AQ (2008) Phoneme-Based Training Strategies in Improving the Perception of Spectrally Distorted Speech. Hear Res 51:526–538.

Stevens G, Flaxman S, Brunskill E, et al. (2013) Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. Eur J Public Health 23:146–52. doi: 10.1093/eurpub/ckr176

Stickney GS, Zeng FG, Litovsky R, Assmann P (2004) Cochlear implant speech recognition with speech maskers. J Acoust Soc Am 116:1081–91. doi: 10.1121/1.1772399

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 50:437–44. doi: 10.1111/j.1467-9450.2009.00741.x

Sweetow R, Palmer CV (2005) Efficacy of individual auditory training in adults: a systematic review of the evidence. J Am Acad Audiol 16:494–504.

Trout J, Poser W (1990) auditory and visual influences on phonemic restoration. Lang Speech 33:121–135. doi: 10.1177/002383099003300202

Valkenier B, Duyne J, Andringa TC, Başkent D (2012) Audiovisual Perception of Congruent and Incongruent Dutch Front Vowels. J Speech, Lang Hear Res 55:1788–1801. doi: 10.1044/1092-4388(2012/11-0227)b

Verschuure J, Brocaar MP (1983) Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. Percept Psychophys 33:232–240.

Versfeld NJ, Daalder L, Festen JM, 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. doi: 10.1121/1.3483733

Warren RM (1970) Perceptual restoration of missing speech sounds. Science 167:392–393. doi: 10.1126/science.167.3917.392

Warren RM (1983) Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. J Audio Eng Soc 31:623–629.

Warren RM, Riener KR, Bashford JA, Brubaker BS (1995) Spectral redundancy: intelligibility of sentences heard through narrow spectral slits. Percept Psychophys 57:175–82.

Warren RM, Sherman GL (1974) Phonemic restorations based on subsequent context. Attention, Perception, Psychophys 16:150–156.

Wechsler D (2012) Wechsler Intelligence Scale, 4th edition (WAIS-¬IV-NL) Dutch version.

Wiersinga-Post E, Tomaskovic S, Slabu L, et al. (2010) Decreased BOLD responses in audiovisual processing. Neuroreport 21:1146–51. doi: 10.1097/WNR.0b013e328340cc47

Winkler I, Takegata R, Sussman E (2005) Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Brain Res Cogn Brain Res 25:291–299. doi: 10.1016/j.cogbrainres.2005.06.005

Wright BA, Zhang Y (2009) A review of the generalization of auditory learning. Philos Trans R Soc Lond B Biol Sci 364:301–311. doi: 10.1098/rstb.2008.0262

Zhang T, Dorman MF, Fu Q-J, Spahr A (2012) Auditory Training in Patients With Unilateral Cochlear Implant and Contralateral Acoustic Stimulation. Ear Hear 33:e70–79. doi: 10.1097/AUD.0b013e318259e5dd

Zhang T, Dorman MF, 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 GB, Schroeder CE, Poeppel D (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. J Neurosci 33:1417–26. doi: 10.1523/JNEUROSCI.3675-12.2013

SUMMARY

Introduction

This thesis presents theoretical and experimental work on auditory and cognitive mechanisms of top-down restoration of degraded speech. In daily life, speech is often interrupted by other sounds that draw our attention. The reason why normal-hearing listeners are able to have a conversation in background noise is that the perceptual system can rely on multiple high-level cognitive mechanisms to restore the degraded auditory input into meaningful speech, in which many factors are involved. Having a conversation in these challenging acoustic environments is very difficult for hearing-impaired people. They often report that they tend to avoid social situations with background noise and prefer face-to-face speech communication. We claim that for hearing-impaired listeners the incoming signal is so poor that the compensation mechanisms and strategies - from which normal-hearing listeners benefit sufficiently - do not always solve the problems they face in everyday speech communication.

The hope that drives the research presented in this thesis is that hearing-impaired listeners could be trained to learn to make better use of cognitive restoration mechanisms to improve their understanding of speech in complex acoustic environments. The main aim of the research presented in this thesis is to better understand the cognitive factors that influence, or are involved in, restoring degraded speech in such a way that the brain can form it into an intelligible speech stream. By studying the restoration mechanisms and capacities of the auditory system of normal hearing listeners we expected to gain more insight in how auditory perception of interrupted speech works in general. We used a well-controlled auditory scene, or rather, a well-controlled and well-understood auditory stimulus: periodically interrupted sentences with and without filler noise in the silent intervals. The brain is able to fill in for missing parts of speech, forming an intelligible percept of the speech stream. This mechanism is called *phonemic restoration*. The specific paradigm *phonemic restoration benefit* can be demonstrated by filling the silent intervals with noise bursts: the intelligibility of periodically interrupted speech increases. While this phenomenon has been known since the 1970s, the underlying mechanisms remained insufficiently studied until the studies presented in this thesis were performed.

Chapter 1: General introduction

In this chapter earlier research concerning the perception of interrupted speech is reviewed. Factors that contribute to the cognitive restoration process are mentioned. Definitions of “Phonemic restoration” and “Phonemic restoration benefit” of filler noise are given. The chapter is concluded with an outline of the thesis.

Chapter 2: Perceptual learning of interrupted speech

In this chapter, the auditory and cognitive mechanisms involved in the restoration process have been evaluated by determining the learning rate in understanding interrupted speech with and without filler noise, in order to gain more insight in how phonemic restoration works. Based on previous literature, we assumed that different cognitive mechanisms lead to learning at different rates, with different effort. We have explored whether the cognitive involvement

for understanding interrupted speech differs when filler noise is applied or not. The perceived listening effort of the participants was used as an indicator of the cognitive mechanisms involved. Another hypothesis was that the benefit of the filler noise would disappear at the end of training, if speech with silent intervals was an ecologically invalid speech stream. This could be tested using the experimental results of the participants trained with this particular stimulus. We found that training increased the overall performance significantly. However, the restoration benefit of filler noise did not diminish. This last finding implies that interrupted speech without filler noise in the silent intervals is ecologically valid, and the phonemic restoration benefit observed in this paradigm is not purely a side effect of the method used. Further, understanding interrupted speech can be learned to some extent through explicit training, regardless of additional filler noise in the silent intervals. The increase in intelligibility and the decrease in perceived mental effort were relatively similar between the groups tested with and without filler noise, suggesting similar cognitive mechanisms for the restoration of the two types of interruptions.

Chapter 3: Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities

The influence of a person's linguistic and cognitive skills on the effectiveness of phonemic restoration has been a topic of debate. Sentence context, linguistic rules and skills, vocabulary, and verbal comprehension were indirectly suggested to be of importance, but the exact mechanisms were not known. In this chapter, the nature of the cognitive restoration mechanisms is described. Measures of linguistic and cognitive skills have been studied to account for the individual differences in the use of top-down restoration mechanisms. The interrupted speech was played at normal speed and at slow speed (while preserving pitch). To account for the individual differences in the use of top-down restoration mechanisms we claimed that linguistic as well as cognitive skills would be important. If, for example, linguistic knowledge matters, relevant training methods could be implemented. We have measured receptive vocabulary by means of the Peabody Picture Vocabulary Test, representing linguistic skills, and full-scale intelligence score by means of the Wechsler Adult Intelligence Scale (a composition of the major components of intelligence: verbal comprehension, perceptual reasoning, working memory and processing speed), representing cognitive skills. The analysis of the data suggests that the restoration mechanisms of interrupted speech do make use of receptive vocabulary and verbal intelligence. The results imply the importance of linguistic factors for top-down restoration. An alternative explanation for our experimental findings is that a full-scale intelligence score is not sufficiently sensitive to capture effects from specific cognitive components on restoration mechanisms. To date, for hearing-impaired listeners these standardized and validated tests have not been used to study the underlying cognitive mechanisms of the restoration of interrupted speech.

In the second part of this thesis the implications of the perception of interrupted speech for cochlear implant users were studied. The ultimate goal was to contribute to training programs aiming at improving the speech perception by cochlear-implant users in background noise. The studies of the second part of this thesis are follow-up studies of these mentioned above.

Chapter 4: Perceptual learning of temporally interrupted spectrally degraded speech

This chapter is a follow-up of chapter 2 in terms of studying the phonemic restoration paradigm, with the main difference that acoustic simulations of cochlear-implant speech processing were used as stimuli. We investigated whether training with interrupted speech could lead to more effective use of restoration mechanisms with degraded speech streams. Effects of training on improvement in the benefit of filler noise and on improvement in intelligibility scores in clinical speech audiometry were investigated. The first hypothesis was that training with cochlear-implant simulations of interrupted sentences can induce more effective use of the high-level restoration mechanisms. Overall, the intensive training increased the understanding of interrupted speech with and without filler noise, implying that indeed listeners made better use of the speech cues in the audible speech portions. Secondly we hypothesized that the phonemic restoration benefit could be induced with training. As described in chapter 2, normal-hearing listeners benefit from filler noise in the silent intervals in case the interrupted sentences are not processed as it happens in a cochlear implant. However, the training with simulations of cochlear-implant processed speech did not induce the restoration benefit both during and after training. This hints that if the combination of degraded speech with filler noise creates misleading speech cues, these can perhaps not be overcome with training. A potential explanation for this is that the brain erroneously attributes (parts of) the filler noise to the noise like speech segments, as the noise and the speech in the cochlear-implant simulations sound qualitatively similar to each other, making a perceptual segregation of the two signals difficult. Finally, we argued that perceptual learning of interrupted sentences could be reflected in clinical speech audiometry with identification of words if listeners learned to make more effective use of cognitive restoration mechanisms during the training. The participants performed better with word-in-noise identification with cochlear-implant simulation in more favorable signal-to-noise ratios, as was expected. However, the performance in the clinical speech audiometry did not increase significantly after the intensive training with the interrupted and spectrally degraded sentences. This observation might indicate that linguistic factors beyond the word level may be even more relevant for the cognitive restoration mechanisms of speech perception.

Chapter 5: The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations

This chapter emphasized the enhancement of the intelligibility, and the restoration benefit of the filler noise, when the auditory stimuli are accompanied with visual speech cues. These cues are provided in the form of a video of the speaker, accompanying the acoustic speech signal. We assumed that these extra cues could enhance both the intelligibility and the phonemic restoration benefit of interrupted sentences. While increasing the intelligibility of interrupted sentences, visual cues did not increase the phonemic restoration benefit of filler noise per se. Finally, the effects of visual speech cues on the perception of simulations of both cochlear-implant and electric-acoustic stimulation processing were studied. The research questions were the same, i.e. whether visual speech cues can enhance the intelligibility, or induce or enhance a restoration benefit of filler noise in interrupted sentences that are degraded spectrally in this particular way. The results show that visual speech cues improve the intelligibility for both types of simulations. A significant restoration benefit of the filler noise was found for the simulations of electric-acoustic stimulation, but with the cochlear-implant simulations the accompanying video of the speaker does not provide the speech cues necessary to induce a restoration benefit. These results show that if the auditory signal is of insufficient quality, as is the case in cochlear-implant processed speech, listeners cannot effectively use the cognitive restoration mechanisms despite good speech perception results.

Chapter 6: General conclusions

Finally, at the end of this thesis we present new insights, accumulated in the course of working on this thesis. First of all, understanding interrupted speech can be learned to some extent regardless of additional spectral degradations, manipulation of the speed of the sentences, or addition of filler noise in the silent intervals. The benefit of the filler noise does not seem to be a side effect of participants being tested with the less ecologically valid interrupted speech with silent intervals, and the benefit of filler noise is persistent: long-term training, playback-rate manipulation and the addition of an accompanying video of the speaker does not influence the benefit of filler noise in interrupted speech. If a restoration benefit of the filler noise is not present in spectrotemporally degraded speech, a restoration benefit cannot be induced with training or with additional speech cues such as a video of the speaker. The phonemic restoration benefit can therefore be seen as hardwired in the auditory system.

Secondly, both the cognitive and auditory processes play an important role in the restoration of interrupted speech. Especially high-level linguistic mechanisms seem to have a large influence on the restoration of interrupted speech. Receptive vocabulary and verbal intelligence are shown to be significant predictors of successful restoration of interrupted sentences without spectral degradations. These *top-down* cognitive restoration mechanisms are shown to be less effective if the *bottom-up* auditory signal is of insufficient quality (as occurs in cochlear-implant speech processing).

Implications for cochlear implant users

Our overall results suggest that better perception of interrupted speech can indeed be achieved via training, even with spectrotemporal degradations of cochlear-implant speech transmission. Since linguistic skills play an important role in the restoration of spectrally degraded interrupted speech, cochlear-implant users can possibly be trained to improve their linguistic skills by reading books or solving crossword puzzles. Furthermore, providing relatively simple feedback, even the text of the sentence, seems to be an effective feedback to lead to successful learning. Finally, lip-reading, which is often available in daily speech communication, does help in speech perception for cochlear-implant users.

SAMENVATTING

Introductie

Dit proefschrift geeft het theoretische en experimentele werk weer rondom de auditieve en cognitieve mechanismen van het top-down herstel van gedegradeerde spraak. In het dagelijks leven wordt spraak vaak onderbroken door andere geluiden die onze aandacht trekken. De reden dat normaal horenden in staat zijn om een gesprek te voeren in achtergrondrumoer, is dat ze gebruik kunnen maken van verschillende hogere cognitieve herstelmechanismen die de gedegradeerde auditieve input herstellen tot betekenisvolle spraak. Een gesprek voeren in deze moeilijke akoestische omstandigheden is zeer lastig voor slechthorenden. Ze zeggen dan ook vaak dat ze de neiging hebben om sociale situaties met achtergrondrumoer te vermijden en de voorkeur geven aan één-op-één gesprekken. Wij stellen dat voor slechthorenden het te verwerken auditieve signaal zo slecht is dat compensatiemechanismen en compensatiestrategieën - waaraan normaal horenden voldoende steun hebben - niet altijd de problemen oplossen die ze dagelijks ondervinden. Voorafgaand aan het onderzoek gepresenteerd in dit proefschrift hadden wij de hoop dat slechthorenden getraind zouden kunnen worden om beter gebruik te leren maken van cognitieve herstelmechanismen, zodat hun spraakverstaan in complexe akoestische situaties zou verbeteren.

Het hoofddoel van dit onderzoek gepresenteerd in dit proefschrift is beter te begrijpen welke cognitieve factoren het herstellen van gedegradeerde spraak dusdanig beïnvloeden dat de hersenen het gebrekkige signaal kunnen omzetten in een begrijpelijk (spraak)signaal. Door het bestuderen van de herstelmechanismen en capaciteiten van het auditieve systeem van normaal horenden verwachtten we meer inzicht te krijgen in de auditieve waarneming van onderbroken spraak in het algemeen. Hierbij werd gebruik gemaakt van een auditieve scene, of beter gezegd een auditieve stimulus, die we goed begrijpen en die we goed kunnen manipuleren: periodiek onderbroken spraak met en zonder opvulruis in de stilte-intervallen. De hersenen zijn in staat om de ontbrekende stukken spraak te reconstrueren, zodat de spraak verstaanbaar wordt. Dit mechanisme noemen we *phonemic restoration*. Het effect van het specifieke verschijnsel *phonemic restoration benefit* kan worden gedemonstreerd door de stilte-intervallen op te vullen met luide ruis, waardoor het spraakverstaan van de periodiek onderbroken spraak verbetert. Dit fenomeen was al sinds de jaren ’70 bekend, echter, het onderliggende mechanisme bleef onvoldoende onderzocht tot de studies gepresenteerd in dit proefschrift werden uitgevoerd.

Hoofdstuk 1: Algemene introductie

Dit hoofdstuk geeft een overzicht van eerder onderzoek verricht naar het verstaan van onderbroken spraak. Factoren die bijdragen aan het proces van cognitief herstel worden beschreven. Tevens worden de definities van “Phonemic restoration” en van “Phonemic restoration benefit” van opvulruis gegeven. Het hoofdstuk wordt afgesloten met een beschrijving van de opbouw van het proefschrift.

Hoofdstuk 2: Het leren verstaan van onderbroken spraak

In dit hoofdstuk worden de auditieve en cognitieve mechanismen die betrokken zijn bij het herstelproces geëvalueerd. Dit werd gedaan door het leereffect te bestuderen in het verstaan van onderbroken spraak met en zonder opvulruis, waardoor meer inzicht werd verkregen in de werking van phonemic restoration. Gebaseerd op eerder onderzoek namen we aan dat verschillende achterliggende cognitieve mechanismen zouden resulteren in het leren met verschillende snelheid en inspanning van de twee soorten stimuli. We hebben onderzocht of de cognitieve betrokkenheid bij het verstaan van onderbroken spraak verandert met het al dan niet toevoegen van opvulruis. De luisterinspanning van de proefpersonen werd gebruikt als een indicatie voor de betrokken cognitieve mechanismen. Een andere hypothese was dat het voordeel van de opvulruis zou verdwijnen door training als spraak met stilte-intervallen een ecologisch valide spraaksignaal zou zijn. Dit werd onderzocht door het bestuderen van de experimentele resultaten van de proefpersonen getraind met deze specifieke stimulus. We vonden dat trainen het spraakverstaan in zijn totaliteit verbeterde.

Het voordeel van de opvulruis verdween echter niet. Deze uitkomst impliceert dat onderbroken spraak zonder opvulruis in de stilte intervallen een ecologisch valide signaal is en dat het voordeel van de opvulruis in dit onderzoeksmodel niet het resultaat is van de toegepaste methode. Verder kan het verstaan van onderbroken spraak tot op zekere hoogte geleerd worden door te trainen, onafhankelijk van het al dan niet opvullen van de stilte intervallen met ruis. De verbetering in het verstaan van de onderbroken spraak en de afname van de luisterinspanning was vergelijkbaar tussen de groepen getest met en zonder opvulruis. Dit suggereert gelijke cognitieve mechanismen voor het herstel van deze twee soorten van onderbreking.

Hoofdstuk 3: Individuele verschillen in top-down herstel van onderbroken spraak: de relatie naar linguïstische en cognitieve mogelijkheden

De invloed van individuele linguïstische en cognitieve vaardigheden op de effectiviteit van phonemic restoration is een onderwerp van discussie geweest. Er werd indirect gesuggereerd dat de context van de zinnen, linguïstische regels en vaardigheden, woordenschat en verbaal begrip van belang zou zijn, maar de precieze mechanismen waren niet bekend. In dit hoofdstuk wordt de aard van het cognitieve mechanisme beschreven. De linguïstische en cognitieve vaardigheden van de proefpersonen zijn onderzocht om de individuele verschillen in het gebruik van top-down herstelmechanismen te verklaren. De onderbroken spraak werd afgespeeld op normale en met vertraagde snelheid (met behoud van toonhoogte). Rekening houdend met de individuele verschillen van de proefpersonen in het gebruik van de top-down herstelmechanismen, stelden we dat zowel linguïstische als cognitieve vaardigheden belangrijk zouden zijn. Als bijvoorbeeld taalvaardigheid van invloed zou zijn, zouden relevante trainingsmethodes geïmplementeerd kunnen worden. De passieve woordenschat werd getest door middel van de Peabody Picture Vocabulary Test, een voorspeller van de algehele taalvaardigheid. De algemene intelligentie werd getest door middel van de Wechsler Adult Intelligence Scale (een samenstelling van de hoofdonderdelen van intelligentie: verbaal begrip, perceptueel redeneren, werkgeheugen en verwerkingssnelheid). De analyse van de data suggereert dat het herstelmechanisme van onderbroken spraak voornamelijk gebruik maakt van de woordenschat en van verbaal begrip. De resultaten impliceren het belang van linguïstische factoren voor het top-down herstel. Een andere verklaring voor onze experimentele resultaten is dat de algemene intelligentiescore niet voldoende gevoelig is om effecten van specifieke betrokken cognitieve onderdelen van de herstelmechanismen te vangen. Tot op heden is er echter voor slechthorenden nog geen onderzoek gedaan met deze gestandaardiseerde en gevalideerde testen om de achterliggende mechanismen van het herstel van onderbroken spraak te bestuderen.

In het tweede gedeelte van dit proefschrift zijn de gevolgen van onderbroken spraak voor gebruikers van een cochleair implantaat bestudeerd. Het uiteindelijke doel was om een bijdrage te leveren aan trainingsprogramma's gericht op het verbeteren van het spraakverstaan in rumoer van mensen met een cochleair implantaat. De studies gepresenteerd in dit gedeelte zijn een vervolg op de eerder genoemde studies.

Hoofdstuk 4: Het leren verstaan van temporeel onderbroken en spectraal gedegradeerde spraak

Dit hoofdstuk is een vervolg op hoofdstuk 2 met betrekking tot het bestuderen van het phonemic restoration paradigma, met het voornaamste verschil dat er gebruik is gemaakt van akoestische simulaties van de spraakbewerking zoals van een cochleair implantaat als stimuli. We onderzochten of training met onderbroken spraak kan leiden tot effectiever gebruik van herstelmechanismen van gedegradeerde spraaksignalen. Het effect van trainen op het verbeteren van het voordeel van de opvulruis en op het verbeteren van de verstaansscores bij klinische spraakaudiometrie werd onderzocht. De eerste hypothese was dat training met cochleair implantaat simulaties van onderbroken spraak tot een effectiever gebruik van cognitieve herstelmechanismen kan leiden. De resultaten lieten zien dat intensieve training het verstaan van onderbroken spraak verbeterde, zowel met en zonder opvulruis. Dit impliceert dat de proefpersonen inderdaad beter gebruik leerden maken van aanknopingspunten in de beschikbare spraakfragmenten. De tweede hypothese was dat het voordeel van de opvulruis geïnduceerd kan worden door middel van training. Zoals beschreven in hoofdstuk 2 hebben normaal horenden profijt van de opvulruis in de stilte intervallen als de zinnen niet bewerkt zijn zoals in een cochleair implantaat. Echter, de training met de simulaties van cochleair implantaat spraakbewerking induceerde niet het herstelmechanisme gedurende en na de training. Dit doet vermoeden dat de combinatie van gedegradeerde spraak met opvulruis misleidende spraak cues oplevert die niet ongedaan gemaakt kunnen worden door middel van training. Een mogelijke verklaring hiervoor is dat de hersenen abusievelijk (gedeelten van) de opvulruis opvatten als spraaksegmenten omdat deze spraaksegmenten zelf ook klinken als ruis. Het feit dat de opvulruis en de spraak van cochleair implantaat simulaties kwalitatief gelijk klinken maakt het splitsen van de twee signalen moeilijk. Tenslotte redeneerden we dat het leereffect van onderbroken spraak ook zijn positieve weerslag zou moeten hebben op de maximale score bij de klinisch gebruikelijke spraakaudiometrie met woorden, als de proefpersonen geleerd zouden hebben om effectiever gebruik te maken van de cognitieve herstelmechanismen gedurende de training. Zoals verwacht presteerden de proefpersonen beter bij de spraakaudiometrie met woorden-in-ruis met cochleair implantaat simulaties in (relatief) gunstige signaal-ruis verhoudingen. Echter, de resultaten van deze test verbeterden niet nadat de proefpersonen intensief getraind waren met spectraal gedegradeerde zinnen. Dit geeft de indruk dat linguïstische factoren hoger dan op woordniveau relevanter zijn voor de cognitieve herstelmechanismen van spraak.

Hoofdstuk 5: Het effect van visuele cues op het top-down herstel van temporeel onderbroken en spectraal gedegradeerde spraak

Dit hoofdstuk legt de nadruk op de verbetering van de verstaanbaarheid en op het voordeel van de opvulruis wanneer visuele cues worden toegevoegd aan de auditieve stimulus. De visuele cues werden aangeleverd in de vorm van een video van de spreker gelijktijdig met het auditieve spraaksignaal. We namen aan dat deze extra cues zowel de verstaanbaarheid als het voordeel van de opvulruis (phonemic restoration benefit) van onderbroken zinnen zou kunnen verbeteren. Hoewel de verstaanbaarheid van de onderbroken zinnen verbeterde, gaven de toegevoegde visuele cues geen toename van het voordeel van de opvulruis. Vervolgens werden de effecten van de visuele cues bestudeerd op de waarneming van simulaties van spraakbewerking zoals in zowel een cochleair implantaat alsook in elektro-akoestische stimulatie. De onderzoeksvragen waren dezelfde, namelijk of visuele spraak-cues de verstaanbaarheid kunnen verbeteren en of dit kan leiden tot het ontstaan of verbeteren van een voordeel van de opvulruis in onderbroken zinnen die spectraal zijn gedegradeerd op deze specifieke manier. De resultaten laten zien dat de visuele cues de verstaanbaarheid van beide type simulaties doen verbeteren. Een significant voordeel van de opvulruis werd gevonden voor de simulaties van de elektro-akoestische stimulatie, maar de visuele cues leverden niet de benodigde spraak cues om een voordeel van de opvulruis op te wekken bij simulatie van het cochleair implantaat. Deze resultaten laten zien dat als het auditieve signaal van onvoldoende kwaliteit is, zoals het geval is bij de spraakbewerking van een cochleair implantaat, gebruikers niet effectief gebruik kunnen maken van de cognitieve herstelmechanismen ondanks een relatief goed spraakverstaan.

Hoofdstuk 6: Algemene conclusies

Ter afsluiting van dit proefschrift presenteren we nieuwe inzichten die zijn opgedaan gedurende het schrijven van dit proefschrift. Ten eerste, het verstaan van onderbroken spraak kan tot op zekere hoogte geleerd worden, onafhankelijk van spectrale degradatie, manipulaties van de snelheid van de zinnen of van opvulruis in de stilte intervallen. Het voordeel van de opvulruis is geen bijkomstigheid van de opzet van het onderzoek waarin de proefpersonen getest zijn met het minder valide ecologische signaal onderbroken spraak met stilte intervallen. Het voordeel is persisterend: intensieve training, manipulatie van de spreeksnelheid en de toevoeging van een begeleidende video van de spreker beïnvloedt niet het voordeel van de opvulruis van onderbroken spraak. Als er geen voordeel van de opvulruis aanwezig is in spectraal gedegradeerde spraak kan dit voordeel ook niet verkregen worden middels training of met het toevoegen van spraak-cues zoals een video van de spreker. Het phonemic restoration benefit van de opvulruis kan daarom gezien worden als een basiseigenschap van het auditieve systeem.

Ten tweede, zowel cognitieve als auditieve processen spelen een belangrijke rol in het herstellen van onderbroken spraak. Met name high-level linguïstische mechanismen lijken erg belangrijk in het herstel van onderbroken spraak. Er is aangetoond dat woordenschat en verbaal begrip significante voorspellers zijn voor succesvol herstel van onderbroken spraak zonder spectrale degradatie. We hebben aangetoond dat deze *top-down* cognitieve herstelmechanismen minder effectief zijn als het *bottom-up* auditieve signaal van onvoldoende kwaliteit is zoals in spraakbewerking van het cochleair implantaat.

Gevolgen voor gebruikers van een cochleair implantaat

In het algemeen suggereren onze resultaten dat met training geleerd kan worden onderbroken spraak beter te verstaan, zelfs als deze spraak spectraal gedegradeerd is zoals gebeurt in de spraaktransmissie van het cochleair implantaat. Omdat linguïstische vaardigheden een belangrijke rol spelen in het herstel van spectraal gedegradeerde onderbroken spraak, kunnen mensen met een cochleair implantaat mogelijk getraind worden door deze linguïstische vaardigheden te trainen middels het lezen van boeken of het oplossen van kruiswoordpuzzels. Het aanbieden van relatief eenvoudige feedback, zoals de geschreven tekst van de zin, lijkt een effectieve feedback voor succesvol leren. Ten slotte, liplezen, dat doorgaans mogelijk is in de dagelijkse spraakcommunicatie, helpt wel degelijk bij het verstaan van onderbroken spraak voor gebruikers van een cochleair implantaat.

THANK YOU!

This thesis would not have been completed without the help of colleagues and friends. I cannot possibly list everyone individually, but I would never have been able to write the articles and this thesis without the help of professor Deniz Başkent: I am grateful for your time and knowledge throughout the years we have worked together. Professor Pim van Dijk, thank you for your critical reading of earlier versions of this thesis.

I am grateful to Michael Heil, for providing time and resources to let me continue and finish the research of this PhD project within Pento Speech and Hearing Center Zwolle and also to my direct colleagues Jessica Kuijper, Wim van Huffelen and Joop Leijendeckers, and my retired colleagues Age Hoekstra and Frank Jorritsma, for standing in for me while I was carrying out my research.

Pranesh Bhargava, thank you for the good days together programming in Matlab and working in the Aquarium: without you it would have been way less fun. Amarins Heeringa, Carina Pals, Jeanne Clarke, Christina Fuller, and Jefta Saija, thanks for the discussions and good times.

Peter van Hengel and Etienne Gaudrain, thanks for your help with the statistics. Annemieke ter Harmsel and Marije Sleurink, thanks for your help with planning appointments with participants and for collecting data. Age and Duifje Hoekstra, thank you for the valuable feedback on the final and earlier versions of this thesis.

I would like to give special thanks to my family. My wife Lian, and our daughters Silke and Lotte, supported and encouraged me. Thank you to my parents for supporting me.

Many thanks to all participants who made it happen!

CURRICULUM VITAE

2012 – present

Medical Physicist – Audiologist
Pento Speech and Hearing Center Zwolle

2010 – 2015

Ph.D. research, University Medical Center Groningen
Department of Otorhinolaryngology
Speech Perception Laboratory of prof. dr. D. Başkent

2008 – 2012

Post M.Sc. trainee Medical Physics – Audiology
Pento Speech and Hearing Center Zwolle & University Medical Center Groningen

2006 – 2008

M.Sc. Applied Physics, University of Groningen
Graduation research at the University of Groningen, Zernike Institute for Advanced Materials, Optical Condensed Matter Physics

2002 – 2006

B.Eng. Engineering Physics, Rijswijk University of Professional Technical Education
Graduation research at the University of Groningen, KVI – Center for Advanced Radiation Technology, Hadronic and Nuclear Physics

PUBLICATIONS

Articles

M.R. Benard & D. Başkent **(2015)**. The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations. *Hearing Research, 328, 24-33*.

M.R. Benard & D. Başkent **(2014)**. Perceptual learning of temporally interrupted spectrally degraded speech. *The Journal of the Acoustical Society of America*, *136*(3), 1344-1352.

M.R. Benard, J.S. Mensink & D. Başkent **(2014)**. Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *The Journal of the Acoustical Society of America*, *135*(2), EL88–94.

M.R. Benard & D. Başkent **(2013)**. Perceptual Learning of Interrupted Speech. *PLoS ONE*, *8*(3), e58149.

Conference proceedings

M.R. Benard & D. Başkent **(2014)**. Which linguistic skills and components of intelligence are involved in the top-down restoration of interrupted speech? Association for Research in Otolaryngology, 37th MidWinter Meeting, San Diego, USA. Podium presentation.

M.R. Benard & D. Başkent **(2013)**. Can interrupted speech be used as training stimuli for cochlear-implant users? Association for Research in Otolaryngology, 36th MidWinter Meeting, Baltimore, USA. Poster presentation.

M.R. Benard & D. Başkent **(2011)**. Perceptual learning of interrupted speech with cochlear-implant simulations. Conference on Implantable Auditory Prostheses, Pacific Grove, USA. Poster presentation.

Publication of this dissertation was supported by

Pento Speech and Hearing Centers
Advanced Bionics
Cochlear
De Nederlandse Vereniging voor Keel-, Neus- & Oorheelkunde en Heelkunde van het Hoofd-Halsgebied
EMID, Electro Medical Instruments BV Doesburg
Oticon Medical
Rijksuniversiteit Groningen (RuG)
School of Behavioural and Cognitive Neuroscience (BCN)
Universitair Medisch Centrum Groningen (UMCG)

Nederlandse Vereniging voor Keel-Neus-Oorheelkunde
en Heelkunde van het Hoofd-Halsgebied

emiD

oticon MEDICAL | Because sound matters

rijksuniversiteit groningen

bcn

umcg

MECHANISMS OF TOP-DOWN RESTORATION OF DEGRADED SPEECH

The present book is an investigation of the underlying mechanisms of top-down speech restoration in the presence of sound degradations in cochlear implants. The results show that both top-down and bottom-up processes play an important role in the restoration of interrupted speech. Especially high-level linguistic mechanisms seem to have a large influence on the restoration of interrupted speech. Receptive vocabulary and verbal intelligence are shown to be significant predictors of successful restoration of interrupted sentences without spectral degradations. These *top-down* restoration mechanisms are shown to be less effective if the *bottom-up* auditory signal is of insufficient quality (as occurs in cochlear implant speech processing).

IMPLICATIONS FOR COCHLEAR IMPLANT USERS

Our overall results suggest that better perception of interrupted speech can indeed be achieved via training, even with spectrotemporal degradations of cochlear implant speech transmission. Since linguistic skills play an important role in the restoration of spectrally degraded interrupted speech, cochlear implant users can possibly train themselves to improve their linguistic skills by reading books or solving crossword puzzles. Furthermore, providing relatively simple feedback, even the text of the sentence, seems to be an effective feedback to lead to successful learning. Finally, lip-reading, which is often available in daily speech communication, does help in speech perception for cochlear implant users.