The Development of Remote Speech Recognition Tests for Adult Cochlear Implant Users: The Effect of Presentation Mode of the Noise and a Reliable Method to Deliver Sound in Home Environments

Feike de Graaff^a Elke Huysmans^a Obaid ur Rehman Qazi^b
Filiep J. Vanpoucke^b Paul Merkus^a S. Theo Goverts^a Cas Smits^a

^aSection Ear and Hearing, Department of Otolaryngology, Head and Neck Surgery, and EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; ^bCochlear Technology Center Belgium, Mechelen, Belgium

Key Words

Cochlear implant · Remote testing · Speech recognition · Digits-in-noise · Consonant-vowel-consonant recognition · Personal audio cable · Direct connect · Discontinuous noise

Abstract

The number of cochlear implant (CI) users is increasing annually, resulting in an increase in the workload of implant centers in ongoing patient management and evaluation. Remote testing of speech recognition could be time-saving for both the implant centers as well as the patient. This study addresses two methodological challenges we encountered in the development of a remote speech recognition tool for adult CI users. First, we examined whether speech recognition in noise performance differed when the steady-state masking noise was presented throughout the test (i.e. continuous) instead of the standard clinical use for evaluation where the masking noise stops after each stimulus (i.e. discontinuous). A direct coupling between the audio port of a tablet computer to the accessory input of the sound proces-

sor with a personal audio cable was used. The setup was calibrated to facilitate presentation of stimuli at a predefined sound level. Finally, differences in frequency response between the audio cable and microphones were investigated.

© 2016 The Author(s)

Published by S. Karger AG, Basel

Introduction

The number of adult cochlear implant (CI) users is increasing annually. This growth is speeding up because the population of CI candidates increases due to changing regulations, expanding candidacy criteria, and technical improvements in CIs. The growing number of adult CI recipients, both newly implanted and experienced users, increases the workload of implant centers and, therefore, promotes the search for new and innovative ways to provide health care for these CI users.

Remote testing and programming could be time and cost saving for both the audiologist and patient. Various

remote applications for CI users have been studied in the past, such as intraoperative testing [Shapiro et al., 2008] and programming [Botros et al., 2013; McElveen Jr et al., 2010; Ramos et al., 2009; Wesarg et al., 2010]. Hughes et al. [2012] and Goehring et al. [2012] investigated the use of telehealth to measure speech recognition abilities in quiet and in noise for CI users. They found poorer speech recognition scores when testing in remote sites (i.e. small conference rooms with videoconference technology) compared to regular testing in a sound booth.

We developed remote tests to measure speech recognition in quiet and noise for adult CI recipients using a direct connection between the sound processor and audio port of a tablet computer. The current study addresses several challenges that we encountered in the development of a remote speech recognition tool. This article focuses on (1) the possible interaction of the advanced sound processing features in the Cochlear™ Nucleus® CP910 Sound Processor when the steady-state masking noise was presented throughout the test (continuous) instead of noise that stops after each stimulus (discontinuous) and (2) a reliable way to deliver sound to the processor.

The Effect of Discontinuous Noise on the Results of Speech-in-Noise Testing

Speech-in-noise tests generally use steady-state noise that starts 0.5–1 s before and ends after the speech stimulus (e.g. a word, digit triplet, or sentence). The use of discontinuous noise allows the subject to respond to the experimenter during the quiet period. However, modern hearing instruments, including CIs, contain adaptive sound processing features such as noise reduction and adaptive algorithms. During testing, the relatively slow-acting, advanced sound processing features need time to become fully active and may remain in a transition state after switching from the quiet response period to the short period where the stimulus is presented in noise. The results of the speech-in-noise test with discontinuous steady-state noise may be affected by these sound processor features and, therefore, may not reflect speech-in-noise recognition abilities in daily life. Thus, the first aim was to study the effect of the sound processor features on speech recognition test outcomes in CI users for a speech-in-noise test with the steady-state noise presented as either continuous or discontinuous.

We also investigated the effect of these two different presentation modes of the noise on speech recognition

outcomes in normal-hearing individuals to rule out possible other factors (i.e. absence of an auditory cue, habituation to noise, and startle reflexes). In the case where the noise is presented continuously during the test, subjects receive no auditory cue about the start of the speech stimulus. In contrast, when the discontinuous noise is used, subjects expect the speech stimulus to be presented approximately 1 s after the start of the noise. The absence of this auditory cue could result in poorer speech-in-noise outcomes in continuous noise compared to when presented in discontinuous noise. Therefore, in the continuous noise condition, we have implemented a visual cue on the tablet screen at the exact moment of presentation of the speech stimulus. It is, however, unclear whether this visual cue supports the attention processes in the same way as the auditory cue in discontinuous noise. It is possible that an increase in speech recognition outcomes might occur due to auditory habituation to continuous noise, and the onset of the noise in the test with discontinuous noise could cause a startle reflex each time the noise starts playing, which may, in turn, influence speech recognition scores. Thus, the second aim was to explore whether the differences between the presentation modes of the steady-state noise yield an effect on speech recognition scores in normal-hearing listeners.

Methods

Subjects

Twelve normal-hearing subjects (2 males and 10 females; mean age 27 years; range 22–41 years) and 16 CI users (8 males and 8 females; mean age 64 years; range 44–83 years) participated in this study. Normal-hearing subjects had pure-tone thresholds not exceeding 15 dB HL at any octave frequency between 500 and 4,000 Hz. All participants were native Dutch speakers. The CI patients acquired their severe hearing impairment after the age of 7, had at least 1 year of experience with their CI, and used the CP910 Sound Processor. This study was approved by the local medical ethical committee. All participants signed informed consent prior to the subject's participation.

Procedure

The digits-in-noise test [Smits et al., 2013] uses 24 digit triplets (e.g. 6-5-2) presented in steady-state noise, which are randomly chosen from a list of 120 triplets, to estimate the speech reception threshold (SRT). The SRT represents the signal-to-noise ratio (SNR) where the listener recognizes 50% of the triplets correctly. The digit triplets were presented at varying SNRs following an adaptive strategy, with a 2-dB step size. The overall presentation level was fixed at 65 dBA, with a start SNR of –4 dB for normal-hearing subjects and 0 dB for CI users. A digit triplet was judged to be correct when all digits were entered correctly. The SNR of a subsequent triplet depended on the correctness of the response on the previous triplet. For the normal-hearing participants, the SRT was calculated as the average SNR over triplets 5–25. Triplet 25 was not presented, but the SNR was based on the response cor-

Fig. 1. Box plots representing the distribution of digits-in-noise SRT scores with the noise presented either continuously during the whole test or discontinuously with quiet response periods in 12 normal-hearing subjects (**a**) and 16 CI users (**b**). The plots show upper and lower quartiles (boxes), medians (crossbars), and extremes of more than 1.5 times the inner quartile range (circles).

rectness and SNR of the preceding digit triplet. For the CI patients, 2 dummy triplets with a fixed SNR were added to make sure that the noise reduction algorithms had settled (approx. 5 s), while also keeping the attention of the test subject. The SRT was calculated as the average SNR over triplets 7–27. Again, triplet 27 was not presented but the SNR was based on the response correctness and SNR of the preceding digit triplet. All tests were performed in two test conditions, with the noise presented either continuously during the whole test or discontinuously with quiet response periods, in a sound-treated room. Participants performed the digits-in-noise test 3 times (practice, test, and retest) in each condition. Analyses were based on the outcomes of the test and retest measurements.

The tests for the normal-hearing subjects were performed with a laptop, with sound played by an external sound card (creative sound blaster X-Fi HD SB1240, Creative Labs) and delivered monaurally to the (subjectively) better ear through Sennheiser HDA200 headphones. The participants had to enter their response on a keyboard. The SRT for the continuous and discontinuous noise conditions was measured in one session.

The speech recognition tests for the CI users were administered with a tablet computer (Lenovo ThinkPad 10). The participants were seated in front of a loudspeaker (Genelec HT 205), which was connected to an external sound card (creative sound blaster X-Fi HD SB1240, Creative Labs). The CI users who normally used a contralateral hearing aid did not use this hearing aid during the tests. The contralateral ear was not occluded, but thresholds were at a level where no contribution to speech recognition could be expected from the unaided ear. The participants verbally repeated the digit triplets they recognized, which were entered by the experimenter. The SRT in continuous and discontinuous noises was measured in different sessions, as the measurements for this study were part of a larger study in which more speech recognition tests were assessed. As a consequence, the digits-in-noise tests with continuous noise were administered in the first session.

Results

The SRT scores of the two conditions averaged for the 12 normal-hearing and 16 CI participants are shown in figure 1.

The data were analyzed using a repeated-measures ANOVA with condition (noise continuously during the whole test vs. discontinuously) and measurement number (test, retest) as within-subject factors. The ANOVA for normal-hearing subjects yielded no significant main effect for condition ($F_{1,11} = 1.237$, $p = 0.290$) and test number ($F_{1,11} = 1.249$, $p = 0.288$), and no interaction between condition and measurement number ($F_{1,11} = 2.798$, $p = 0.123$). The ANOVA on the SRT data from the CI users showed no significant main effect of condition ($F_{1,15} = 0.047$, $p = 0.832$) and test number ($F_{1,15} = 0.002$, $p = 0.966$), and no interaction between condition and test number ($F_{1,15} = 2.122$, $p = 0.166$). These results suggest that there is no significant difference between SRTs measured with steady-state noise that is presented continuously during the whole test and steady-state noise presented with quiet response periods after each stimulus, for normal-hearing subjects and CI users with the CP910 Sound Processor.

Stimulus Delivery

Remote testing of speech recognition requires that the quality of loudspeakers of desktop computers, tablet computers or smartphones has no effect on the test results. Background noise and room acoustics are other impor-

Fig. 2. Experimental setup with a CI connected to a tablet computer using a PAC.

tant factors that are difficult to control in home settings and could affect the test results [Goehring et al., 2012; Hughes et al., 2012]. In our study, the audio signal was presented directly through the accessory input provided on the CP910 Sound Processor to avoid these unwanted effects. The direct coupling between the audio port of the tablet computer to the accessory input of the sound processor using a personal audio cable (PAC) was examined as a means to achieve the strict requirements needed for reliable home testing. Special attention was given to the calibration of the system and the frequency responses of the microphones and PAC. The experimental setup is shown in figure 2.

Calibration

The system was calibrated to assure that speech stimuli were presented at a predefined sound level (measured in dBA or dB SPL) to the sound processor at specified volume and sensitivity settings through the PAC. A CP910 Sound Processor, fitted with a map in slot 1 with volume and sensitivity settings of 10 and 12, respectively, and mixing ratio set to ‘accessory only’ was connected to the tablet computer with a PAC. The calibration noise file available for the speech materials used (digits-in-noise and consonant-vowel-consonant, CVC) were presented through a loudspeaker in a sound booth. The sound pressure levels were read out directly from the internal sound level meter of the sound processor. These levels were compared to measurements done by a Brüel and Kjaer Type 2250 sound level meter positioned at the same location in the room as the sound processor. The broad-band differences in levels (dB SPL and dBA) were within 1 dB for calibration noise. These differences were assumed to be due to the position of the processor microphones and their slightly different frequency responses compared to

the A-weighting curve. Therefore, it was concluded that the internal sound level meter of the sound processor can be used for accurate calibration of the level of incoming signals. For this purpose, the broad-band root mean square level of the input signal is adjusted (depending upon the speech material) such that the desired level is achieved when the input signal is presented through the PAC.

Frequency Response

When the PAC is connected, the sound processor shapes the incoming signal with a frequency response to mimic that of the microphone signal (pre-emphasis). The frequency response was measured using different pure tones, and the levels were read from the sound processor. A small difference was found between the frequency response of the signals that were delivered by the PAC and the microphones with a maximum deviation of approximately +3 dB at 6,000 Hz. A difference in frequency response could possibly have an effect on speech recognition scores. Therefore, we performed an experiment in which we compared the speech recognition scores in quiet and in noise with a PAC using the original speech material and speech material that was filtered to compensate for the difference in frequency response.

Methods

Subjects

Seven CI users participated in this study; all were part of the group of 16 CI users described earlier.

Procedure

Speech recognition tests included a test with monosyllabic words in quiet [Bosman and Smoorenburg, 1995] and digits-in-noise with continuous noise [Smits et al., 2013]. The monosyllabic words had a consonant-vowel-consonant structure, thus containing 3 phonemes. The words were spoken by a female speaker and

Fig. 3. a CVC recognition in quiet score in percentage phonemes correct (average of 3 lists) measured with unfiltered and filtered stimuli presented via a PAC in 7 CI patients. **b** Digits-in-noise SRT score (average of test and retest) measured with unfiltered and filtered stimuli presented via a PAC in 7 CI users.

presented in quiet at an intensity of 65 dB SPL. Three lists of 12 words were assessed. The procedure for the digits-in-noise test for CI users was explained earlier. For this experiment, the noise was presented continuously during the whole test.

A digital equalization filter was constructed to compensate for the differences in frequency characteristics between microphone input and PAC input. The audio files were filtered with this filter to ensure that the frequency response of the stimuli presented via the PAC matched the frequency response of the stimuli presented via the microphone. The speech recognition tests were first performed with the nonfiltered audio files, followed by the tests with filtered audio files. All tests were performed in a sound-treated booth with the stimuli presented via a PAC.

Results

The results of the speech recognition tests in quiet and in noise, measured with original and filtered stimuli presented with a PAC, are illustrated in figure 3.

A Wilcoxon signed-rank test showed that there was a significant difference in the CVC recognition scores obtained with the unfiltered (mean = 81.4%; SD = 16.5) and filtered (mean = 89.3%; SD = 9.7) audio files ($Z = -2.201$, $p < 0.05$). The results of the speech recognition tests in noise were subsequently analyzed. A Wilcoxon signed-rank test revealed no significant difference in SRT scores acquired with the PAC with unfiltered (mean = -5.5 dB SNR; SD = 1.3) and filtered (mean = -5.1 dB SNR; SD = 1.5) audio files ($Z = -1.183$, $p = 0.237$).

Discussion

The current study addresses two methodological challenges we encountered in the development of a remote speech recognition tool for adult CI users. First, we examined whether speech recognition in noise performance differed when the steady-state masking noise was presented continuously during the whole test instead of the standard use where the masking noise stops after each stimulus (discontinuous). We found no significant difference between those conditions for normal-hearing individuals, suggesting that the visual cue is appropriate to alert the listener that the next stimulus is to be presented. It also suggests that habituation to the noise or a startle reflex in the discontinuous noise tests with onset of the noise do not exist or have no effect on the test results. For the CI users, lower (better) SRTs for the test with continuous noise were expected because several features in the CP910 processor (i.e. AGC, ADRO, ASC, SNR-NR) are slow-acting features with time constants in the order of seconds, and some have been shown to improve speech recognition scores in noise (see Wolfe et al. [2015] for an overview). The results of the current study, however, showed no significant differences in speech recognition scores between continuous or discontinuous noise for CI users using the CP910 Sound Processor. A possible reason for this finding could be the low SNRs for the signals

in the digits-in-noise test that may reduce the effectiveness of noise reduction systems. Versfeld and Goverts [2013] demonstrated the effect of a carrier phrase on speech recognition scores in quiet and reported differences in scores for some hearing aids. This effect may also exist for CI users.

Second, we investigated the direct coupling between the tablet computer and sound processor using a PAC, by focusing on the calibration of the setup and differences in frequency response between PAC and microphones. Previous studies by Goehring et al. [2012] and Hughes et al. [2012] indicated that the characteristics of the test environment (i.e. a combination of background noise and reverberation) have an effect on speech recognition in CI users. We used a PAC to deliver sound directly to the sound processor and could therefore bypass possible negative effects of the test environment. The PAC, however, introduces a small difference in the frequency response compared to the frequency response of the microphones. This difference is typically not noted when listening through the PAC to, for instance, music, but it may affect formal audiological testing. The effect of the difference in frequency responses on speech recognition outcomes was examined with the use of filtered audio files. Performance of speech recognition in quiet was better with the filtered audio files than with the unfiltered audio files, whereas no difference was found in speech recognition in noise test scores. The difference in speech recognition scores in quiet might be due to a training effect throughout the different test sessions. However, prior to the CVC tests with filtered audio files, participants performed a multitude of CVC tests since the assessments were part of a larger study, which makes the likelihood of a training effect very small. There might also be an effect of the specific CVC list used in a test, despite the minimal phonemic differences among the lists [Bosman and Smoorenburg, 1995]. Different lists were used for each test in a different condition, but were the same for all participants. Another underlying cause might be that CI recipients normally use the microphone and may need to adapt to the slightly different spectrum of the sound that is delivered by the PAC. Possibly, this effect is more prominent for speech-in-quiet testing than for speech-in-noise testing where the masking noise is the main factor that corrupts the signal.

This study is part of a project targeting older users (60 years or older) of hearing implants and aims at allowing them to more effectively use their hearing implants in daily life. Implant centers are often located in larger cities, which requires many patients to travel long distances. The latter might be a problem when patients become old-

er and potentially less mobile. Remote testing has the potential to reduce the need for appointments and might, therefore, reduce the need for transportation. The average age of our participants was 64 years (range 44–83 years). All participants were able to perform the remote tests without any problems, and all participants reported positive experiences with the remote testing. It was considered easy to connect the PAC, launch the application and perform the tests, and the complete test block took no more than half an hour. One participant was unable to execute the tests himself (plug in the PAC and type the response) due to a medical condition; however, with the help of a family member who entered the answers, he was able to perform the tests.

In conclusion, the setup as used in this study gives promising results and provides a solid base for future studies on the remote assessment of speech recognition abilities in CI users in quiet and in noise. We have shown that speech-in-noise test outcomes are not influenced by the use of steady-state noise when presented continuously during the whole test. In addition, we have shown that stimuli can be presented with predefined levels using a direct connection via a PAC between a tablet computer and sound processor. In a future study, we will examine whether home assessment of speech recognition in CI users yields valid outcomes by comparing these to the outcomes of speech recognition assessment in a clinical setting. In addition, possible differences between stimulus presentation by a loudspeaker or PAC and assessment by an experimenter or self-assessment by the CI user will be investigated. Digital streaming of stimuli to the processor either via the remote control or Bluetooth might become available in the near future and may increase the ease of remote testing even further.

Acknowledgments

The authors would like to thank Ann de Bolster and Stefan Lievens from Cochlear Technology Center, Mechelen, for their contributions to the development and programming of the speech recognition tool.

Disclosure Statement

This research is part of the ‘Supporting Hearing in Elderly Citizens (SHiEC)’ project (www.shiec.eu), funded by the European Commission as part of the 6th call of the Ambient Assisted Living programme (project AAL-2013-6-065).

References

- Bosman AJ, Smoorenburg GF: Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Int J Audiol* 1995;34: 260–284.
- Botros A, Banna R, Maruthurkkara S: The next generation of Nucleus® fitting: a multiplatform approach towards universal cochlear implant management. *Int J Audiol* 2013;52: 485–494.
- Goehring JL, Hughes ML, Baudhuin JL, Valente DL, McCreery RW, Diaz GR, Sanford T, Harpster R: The effect of technology and testing environment on speech perception using telehealth with cochlear implant recipients. *J Speech Lang Hear Res* 2012;55:1373–1386.
- Hughes ML, Goehring JL, Baudhuin JL, Diaz GR, Sanford T, Harpster R, Valente DL: Use of telehealth for research and clinical measures in cochlear implant recipients: a validation study. *J Speech Lang Hear Res* 2012;55:1112–1127.
- McElveen JT Jr, Blackburn EL, Green JD Jr, McLearn PW, Thimsen DJ, Wilson BS: Remote programming of cochlear implants: a telecommunications model. *Otol Neurotol* 2010; 31:1035–1040.
- Ramos A, Rodríguez C, Martínez-Beneyto P, Pérez D, Gault A, Falcon JC, Boyle P: Use of telemedicine in the remote programming of cochlear implants. *Acta Otolaryngol* 2009; 129:533–540.
- Shapiro WH, Huang T, Shaw T, Roland JT Jr, Lalwani AK: Remote intraoperative monitoring during cochlear implant surgery is feasible and efficient. *Otol Neurotol* 2008;29:495–498.
- Smits C, Goverts ST, Festen JM: The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J Acoust Soc Am* 2013; 133:1693–1706.
- Versfeld NJ, Goverts ST: The effect of a carrier phrase on hearing aid amplification of single words in quiet. *Int J Audiol* 2013;52:189–193.
- Wesarg T, Wasowski A, Skarzynski H, Ramos A, Falcon Gonzalez JC, Kyriafinis G, Junge F, Novakovich A, Mauch H, Laszig R: Remote fitting in Nucleus cochlear implant recipients. *Acta Otolaryngol* 2010;130:1379–1388.
- Wolfe J, Neumann S, Marsh M, Schafer E, Lianos L, Gilden J, O'Neill L, Arkis P, Menapace C, Nel E, Jones M: Benefits of adaptive signal processing in a commercially available cochlear implant sound processor. *Otol Neurotol* 2015;36:1181–1190.