International Journal of Pediatric Otorhinolaryngology 113 (2018) 124–130

Contents lists available at ScienceDirect

International Journal of Pediatric Otorhinolaryngology

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

Spoken word recognition in noise in Mandarin-speaking pediatric cochlear implant users

Cuncun Ren[a,b], Jing Yang[c], Dingjun Zha[a], Ying Lin[a], Haihong Liu[d], Ying Kong[b], Sha Liu[b,**], Li Xu[e,*]

[a] Department of Otolaryngology, Xijing Hospital, Xi'an, China
[b] Beijing Tongren Hospital, Beijing Institute of Otolaryngology, Capital Medical University, Beijing, China
[c] Communication Sciences and Disorders, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[d] Beijing Children's Hospital, Capital Medical University, Beijing, China
[e] Communication Sciences and Disorders, Ohio University, Athens, OH, USA

ARTICLE INFO

Keywords:
Cochlear implant
Children
Word recognition
Noise
Mandarin Chinese

ABSTRACT

Objective: The purpose of the present study was to compare spoken word recognition performance in the presence of speech spectrum-shaped noise and four-talker babbles in Mandarin-speaking children with cochlear implants (CIs).

Methods: Participants included 33 children with unilateral CIs (with a mean age of 10.4 ± 2.9 years old and a mean length of CI use of 7.5 ± 3.0 years). The Standard Chinese version of Lexical Neighborhood Test was implemented in quiet, speech-spectrum-shaped noise (SSN), and four-talker babble (FTB). The signal-to-noise ratios (SNRs) were set at +5 and + 10 dB for both types of maskers. Participants responded by verbally repeating each word they heard and the response was scored as the percentage accuracy of recognition performance. A Generalized Linear Model (GLM) fitting, correlational tests, and a two-way repeated-measures ANOVA were conducted on the percent-correct data.

Results: Word recognition in quiet was on average 74.5% correct but dropped to 57.3% and 48.8% correct for SSN and FTB at 10 dB SNR, respectively, and 44.4% and 32.6% correct for SSN and FTB at 5 dB SNR, respectively. In both quiet and noise conditions, the participants showed lower recognition accuracy for the hard words than for the easy words. Disyllabic words were recognized with higher accuracy rates than were the monosyllabic words. The GLM analysis revealed that all four tested factors (masker type, SNR, lexical neighborhood feature, and lexical type) showed significant impacts on word recognition in children with CIs. Word recognition scores in the two types of maskers were significantly correlated for the disyllabic words at both SNRs and monosyllabic words at 10 dB SNR.

Conclusions: The present study demonstrated that the lexical features such as the lexical neighborhood characteristics and lexical type had significant effects on speech recognition performance in both quiet and noise conditions in pediatric CI users. Children with years of experience of CI use still encountered remarkable difficulties in everyday listening environment although their speech recognition in quiet reached relatively desired level. Fluctuating noise, such as speech babbles, caused greater challenge than steady-state noise for speech recognition in children with CIs.

1. Introduction

For patients with severe to profound hearing loss, cochlear implants (CIs) have become the most effective rehabilitation tool. Children with CIs have unprecedented access to important speech information that allows them to develop spoken language skills. As a result, CIs have dramatically improved speech perception ability for hearing-impaired children [1–6]. In recent years, many children with CIs have received education in mainstream schools. However, their auditory, linguistic, and cognitive outcome measures still show great variability in individual performances.

Previous research has shown that CI recipients have reasonably

* Corresponding author. Communication and Sciences and Disorders, Ohio University, Athens, OH 45701, USA.
** Corresponding author. Beijing Institute of Otolaryngology, Beijing Tongren Hospital, Capital Medical University, Beijing 100005, China.
E-mail addresses: liusha_liu@126.com (S. Liu), xul@ohio.edu (L. Xu).

https://doi.org/10.1016/j.ijporl.2018.07.039
Received 18 May 2018; Received in revised form 22 July 2018; Accepted 22 July 2018
Available online 24 July 2018
0165-5876/

good performance of speech perception in quiet [7]. However, natural acoustic environments often contain multiple sources of noises. Speech recognition tests in noise offer more representative outcomes of perceptual abilities in everyday listening environment for those individuals using auditory devices. Researchers found that hearing-impaired listeners presented considerable deficits in speech recognition in noisy environment [8–12]. For example, Dorman et al. [8] compared the sentence recognition in quiet, noise at 5 and 10 dB signal-to-noise ratios (SNRs) in 15 adults fitted with a CI in one ear and hearing aid in the contralateral ear. These patients were tested using electrical stimulation, acoustic stimulation, and combined stimulation. The results revealed that in all three conditions, the patients showed much lower sentence recognition accuracy in noise than in quiet and the recognition accuracy decreased as the SNR decreased from 10 to 5 dB. In a more recent study, Caldwell and Nittrouer [10] tested speech recognition in quiet and noise at −3, 0, +3 dB SNR in children with normal-hearing (NH), hearing aids (HA), or CIs. They found that all three groups of children showed poorer performance in noise than in quiet. In addition, when the SNR increased from −3 to +3 dB, all three groups of children demonstrated improved recognition accuracy and the magnitude of improvement was the greatest in the children with NH but the smallest in the children with CIs.

Speech spectrum-shaped noise and multi-talker babbles are most frequently used noise types in assessing individual speech recognition ability. When listeners are exposed in noise, two types of masking effects are usually involved. One type of masking effect is energetic masking which is also referred as “physical masking” that occurs when the neural excitation at the auditory periphery evoked by the masker exceeds the excitation produced by the target speech [13]. This masking effect is determined by the relative energies of the signal and the masker. By contrast, informational masking is one form of “perceptual masking” that can be “broadly defined as a degradation of auditory detection or discrimination of a signal embedded in a context of other similar sounds” [14]. Speech-shaped noise, a continuous steady-state noise, introduces interference of masker at the level of peripheral auditory processing and produces primarily energetic masking. Multi-talker babble, a fluctuant speech noise, introduces informational masking in addition to energetic masking because it is synthesized by mixing connected meaningful speech produced by a number of different speakers. The audible fragments presented in the speech babble masker interfere with a listener's ability to segregate and/or selectively attend to target speech. The effect of informational masking for speech recognition decreases as the number of talkers increases [15,16] because the acoustic waveform of the resulting masker babble is less confusable with target speech. In this situation, the masking effect is primarily caused by energetic masking [17]. On the other hand, a masking babble with fewer talkers produces masking release in NH listeners because NH listeners can benefit from the low-intensity parts of the competing noise to recognize speech information. Such ability is referred to as glimpse or listening in dips. However, listeners with hearing loss usually show very little benefit from these dips when tested in the same conditions in comparison to listeners with NH [18].

A few previous studies tested the recognition of pitch and disyllabic words presented with a two-talker babble or a steady-state white noise in adult CI users [9,12,19]. Yet, little data has been reported on the performance of speech recognition in different types of noise in children with CIs. Mao and Xu [20] reported lexical tone recognition in noise in both NH children and prelingually-deafened children with CIs. While children with CIs showed remarkably reduced tone recognition in noise, NH children also performed more poorly than NH adults in adverse listening conditions. This result was consistent with many other reports that children required a more optimal SNR to achieve a reasonably good performance in speech perception than did adults in the presence of various types of background noise [21–23]. The performance gap between children and adults appeared to be larger when the background noise was competing speech babble rather than the steady-state noise [22,24]. Given the different masking effects introduced by speech-spectrum-shaped noise and multi-talker babbles as well as children-adults differences in speech recognition in noise, the present study was designed to examine the performance of speech recognition in different types of noise in children with CI. Specifically, the research aim of the present study was two-folds: first, to determine how the factors of lexical density characteristics, lexical type, noise type, and SNR affect the spoken word recognition performance in children with CIs; and second, to examine whether children with CIs perform differently in steady-state speech-shaped noise and multi-talker speech babbles.

To address these research aims, a Standard-Chinese version of Lexical Neighborhood Test (LNT) was used in the present study. LNT was developed under the theoretical framework of Neighborhood Activation Model (NAM) which states that a stimulus word is recognized in a context of competing words activated in the memory that show similar acoustic-phonetic representations to the target word. The selection process is affected by the word frequency of the stimulus, the density and word frequency of the neighbors. LNT, based on NAM, was developed with child-friendly words selected from the corpus of Child Language Data Exchange System (CHILDES). This test includes two types of words: lexically easy words and lexically hard words. The former includes those with high word frequency and low neighborhood density and the latter includes those with low word frequency and high neighborhood density. Following the same principles of LNT, a Standard-Chinese version LNT was developed. It included monosyllabic and disyllabic words which were further divided into monosyllabic-easy (ME), monosyllabic-hard (MH), disyllabic-easy (DE), and disyllabic-hard (DH) words. Details regarding the development of Standard-Chinese LNT and the word density and frequency were provided in Liu et al. [25].

Liu et al. [2] used the Standard-Chinese LNT to test prelingually-deafened Mandarin-speaking children with CIs in quiet condition. This study included 230 pediatric Mandarin-speaking CI users with a wide range of chronological age and length of implant use. The results showed that the children with CIs had higher recognition accuracy for the easy words than for the hard words. Participants' age at implantation and length of device use also affected their recognition performance. Specifically, children with early implantation age and longer duration of device use showed higher recognition accuracy. However, compared with NH controls, the CI group still showed evident deficit in word recognition.

It is well established that preschoolers and school-aged children require a more desirable SNR than adults to achieve similar performance for speech recognition in the presence of steady-state noise or multi-talker babbles [21,24]. Ren et al. [26] used the Standard-Chinese LNT to test word recognition of NH Mandarin-speaking children in quiet and in speech-spectrum-shaped noise at 0 dB SNR. The authors found that the speech-spectrum-shaped noise caused decreased recognition accuracy in comparison to quiet condition. In both quiet and noise conditions, the NH Mandarin-speaking children showed higher word recognition scores for the easy words than for the hard words. In addition, the recognition accuracy in noise condition increased as a function of the participants' chronological age (from 3 to 6 years old).

Based on the findings on word recognition in children with CIs in quiet and children with NH in noise [2,26], it was reasonable to predict that children with CIs would perform worse in noisy environment than in quiet condition for word recognition and the recognition accuracy would decrease as the SNR decreases. In addition, the lexical characteristics of the stimulus words were expected to affect the performance of children with CIs in a similar way to the children with NH. That is, the easy words should be recognized with higher accuracy than the hard words. As the present study included two types of noise (speech spectrum-shaped and multi-talker babble) that employ different masking effects, it was of particular interest to determine to what