Proc. of the 1st Int. Conference on Challenges in Hearing Assistive Technology (CHAT-17), Stockholm, Sweden, August 19, 2017

An iOS-based speech audiometry for self-assessment of hearing status

Hongying Yang, Tao Song, Mingjie Song, Xihong Wu, Jing Chen

Department of Machine Intelligence, Speech and Hearing Research Center, and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China

yanghy@pku.edu.cn, chenj@cis.pku.edu.cn

Abstract

Speech audiometry is one of the methods for evaluating hearing status, and recently it's tended to be implemented on mobile phones with the development of internet and smart devices. Although several applications include the function of speech audiometry, the effectiveness of this function was rarely reported.

In this work, we developed an Apple iOS-based application (Hearing Assistant) with functions of pure-tone audiometry and speech audiometry for people speaking Mandarin Chinese. The speech material was designed to reflect listeners' audibility at specified frequencies. The reliability of the speech audiometry was analyzed and discussed according to the users' data collected through this application.

The results showed a significant but week correlation between pure-tone thresholds and scores of speech audiometry, indicating the speech audiometry can be used to discriminate impaired and normal hearing but it's not accurate enough to evaluate hearing thresholds quantitatively at each test frequency.

Index Terms: speech intelligibility, hearing impairment, mobile devices, iOS-based application

1. Introduction

Recently, there is increasing interest in developing a suitable method for self-assessment of hearing situation and conducting hearing compensation on smartphones. Some studies have concentrated on implementing pure tone audiometry (PTA) on smartphones, which lead to some apps (uHear, EarTrumpet, etc). However, PTA has critical requirement on test environment and stimuli calibration. Factors influencing PTA test on smartphones have been studied in previous works, including the test program [1], the test accuracy [2], the time cost [3], the reference sound level [4], the user types [5, 6], and the usage of specific extra-devices [7].

Some applications are available for speech audiometry with different test materials and procedures, for instance, uHear uses Acceptable Noise Level Speech Recognition Test, Siemens Hearing Test uses nonsense word recognition test in noise. All of these applications can show speech intelligibility and suggest brief rating of hearing impairment, but unlike PTA test, they can't supply any prescription for frequency-dependent hearing compensation. The present work aimed to develop an iOS-based speech audiometry for predicting the degree of hearing impairment at the specified frequencies.

The method of supra-threshold word identification was adopted in the speech audiometry, as this test is relatively robust for the variation of sound level. It has been reported that the speech perception performance remains constant for both normal-hearing and hearing-impaired people when sound level is in a reasonable range [8], and the speech recognition score decreases when stimuli level exceeds comfortable level [9]. A paradigm of identifying the spoken word in a confused word pair was used in the audiometry, in which the confused word pair was shown as text on the mobile screen. To select appropriated words as speech materials and produce confused word pairs, acoustic features of Mandarin initials and finals were studied.

The distinctive features of confused Mandarin initials and finals has been systematically studied in [10]. Based on this work, we produced monosyllabic word pairs to assess frequency-dependent hearing impairment by acoustic simulation with normal hearing listeners in a previous study [11]. The monosyllabic words were in consonant-vowel (CV) structure. For each pair, the syllables shared the same consonant or vowels, and kept the other phoneme different but easy to confuse, e.g. "/ba4/ vs./pa4/" or "/tao2/ vs. /tou2/" (the number represents the tonal pattern, and they were the same within a pair). Hearing impairment was simulated by processing stimuli with a serial of band-stop filters at different cut-off frequencies. Normal-hearing listeners took part in the experiment and their task was to select the correct syllable within each pair after one syllable was displayed. The results showed a significant correlation between frequency ranges of the band-stop filters (where hearing impairment occurred) and the frequency range of the distinctive feature for each pair, indicating the feasibility to predict the frequency-dependent hearing impairment by using speech audiometry with confused words pairs. However, these results were based on hearing-loss simulation in laboratory, and the frequency range was mainly below 2000 Hz.

In the present study, test material was determined as disyllabic words, as they're more intelligible than monosyllabic for Mandarin Chinese. The distinctive feature of the word pairs was determined as the frequency and intensity of formants, since speech formants are the most important acoustic feature for identifying a phoneme. Additionally, the general syllable structure of Chinese words, consonant-vowel, makes this acoustic feature quite critical for word identification. The first work for analyzing the strength of speech formant along frequencies among phonemes was finished by Fant and his colleagues [12] based on Swedish language. Whereafter, the conception of "Speech Banana" was developed, which showed the mean frequency of the first three formants and its mean intensity of a phoneme by a scatterplot in audiogram, and it has been available for English and Mandarin Chinese [13]. Generally, Speech Banana for Mandarin Chinese showed that the intensities of consonants are lower than vowels and the frequency of consonants (the frequency with local maximum energy) are higher than the frequency of vowels (the mean frequency of first three formant). It's also reported that the perception of consonants is harder than the perception of vowels, and it is easily affected by background noise or hearing impairment [14]. These studies suggest that the word

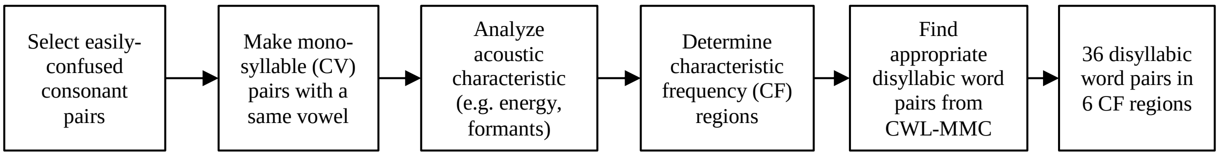

Figure 1: *The illustration shows the procedure for word pairs stimuli selection.*

identification within a confused pair would be dominant by the consonant identification, rather than by the vowel.

Therefore, for producing the speech materials, we firstly selected mono-syllable pairs that have CV structure and the same vowel. For each pair, consonants which are easily confused were selected, due to their energy distribution showing big difference for one specific frequency (or a narrow frequency range). In other words, the two confused consonants were adjacent along frequency but separate along hearing level in Speech Banana. These specific frequency or frequency region was defined as the characteristic frequency (CF) of the CV syllable pair. And then disyllabic word pairs consist of these CV syllable pairs were produced based on the following rules: 1) for each pair of disyllable words, one pair of confused CV syllable was assigned to the words, and the other syllable of the words were kept the same, e.g. “Ji2Zhong1 vs. Qi2Zhong1” (“concentrate vs. among” in English); 2) all words were common words in everyday life; 3) these words were assumed to be spondaic as two syllables of spondee were equally important and easy to identify. Word-pairs selected in this way could ensure that the only distinctive feature was concentrated at one certain CF. And then the wrong identification of the confused words can reflect the elevation of hearing threshold at the CF. Figure 1 shows the diagram of the whole procedure for the word pairs selection.

An iOS-based application (Hearing Assistant) was developed and released in China Apple Store, including functions of PTA and speech audiometry test. The effectiveness of the PTA test of this application has been testified and reported in [15]. To evaluate the validity of the speech audiometry test, the results of the speech audiometry were compared with that of PTA test among three user groups: normal-hearing (NH) group, hearing-impaired (HI) group and the all-user (AU) group, according to the data collected from 235 application users.

2. Method

2.1. Material and stimuli

Based on the speech Banana of Mandarin Chinese, the consonants were divided into six groups according to their locations along frequency axis [13], corresponding to six CF regions, around 250, 500, 1k, 2k, 4k, and 6k Hz. These CFs were determined to be matched with the test frequencies of PTA. Then, thirty-six disyllabic word pairs were selected from Common Word List of Modern Mandarin Chinese (CWL-MMC) [16], to produce 6 word pairs for each of the six CFs. These words are often common-used in daily conversation and their occurrence-frequencies were also balanced to ensure they could be equally identified when they were both audible. For each disyllabic word pair, their only difference is the consonant of the confused syllable, so the CF of syllables can also represent the CF the word pairs. Please notice, although there is one syllable the same between the two words within a pair, the corresponding characters of the two words can be different, because a syllable in Chinese Mandarin may have multiple meanings, which are dependent on the corresponding Chinese characters. These words were selected from the CWL-MMC by programming based on the rules above, and the output were checked manually.

The 72 words (36 pairs) were spoken by a young female native Chinese speaker, and recorded digitally onto a computer disk at the sampling rate of 44.1 kHz with 24-bit quantization. The amplitude of each stimulus was normalized in terms of root-mean-square (rms) pressure.

2.2. Implementation of iOS-based speech test

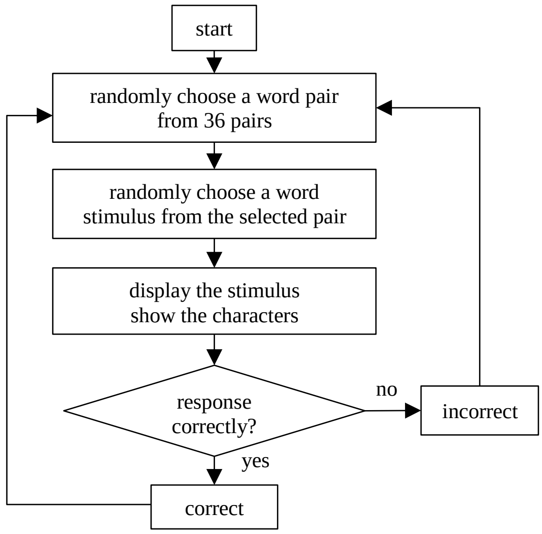

Figure 2: *The illustration shows the procedure of speech audiometry test in the application.*

The procedure of iOS-based speech test was as follows. At first, the listener pressed a “start” button to initial the procedure. Then the App began to play a random-chosen word stimulus through the earphone, and at the same time, the Chinese character of the word pair were shown on the screen. The listener was instructed to select the characters of the word he/she heard on the screen by pushing a button. The correct/incorrect response was accounted. Afterwards, another word pair was tested till the 36 pairs were all finished. Figure 2 shows diagram of this procedure. A progressing bar indicated the percent that had been finished. Usually, the whole test could be finished in ten minutes. After the test, the speech recognition scores were mapped to hearing level at each CF in the audiogram, where a higher score corresponded to a lower hearing level and a lower score correspond to a higher hearing level. A suggestion of the hearing situation was also showed with text.

Before this procedure, all users had to finish the PTA test, otherwise the function of speech audiometry could not be operated. To ensure the speech stimuli were displayed at a comfortable level, the sound level was adjusted at 30 dB higher than the mean hearing threshold for each user.

2.3. Implementation of iOS-based PTA

The procedure of the iOS-based PTA test was as described in the previous work [15]. Listeners were instructed to push a "begin" button to initiate the test, and push the button "I heard" if a pure tone was heard. The ascending method introduced in the standard of ISO 8253-1 [17] was implemented.

There was a preliminary calibration before the App was released on the App Store, and the consistency across different Apple devices has been testified by previous studies [15, 18]. These works confirmed that a relatively reliable PTA results could be collected from application users. Users can download the App from App Store and use the earphones to finish the tests with instruction, and there is no need for self-calibration.

2.4. Data collection and the users

The data used for analysis were from the 300 users who registered at the first five months. To ensure the reliability of the results, the data was prescreened by deleting some invalid data, which were collected from listeners who did not finish the whole test or did not go through the test as instructed, according to the following rules:

1) If the results of a PTA test were 90 dB HL for all frequencies, which means the button was not pressed at all and the PTA test procedure was not finished correctly, the data of this test was thought as invalid and deleted.
2) If the results of a speech test are 0 for all frequency range, which means the button was not pressed at all and the speech test procedure was not finished correctly, the data of this test was thought as invalid and deleted.

According to the criterions above, the data of 65 users were deleted. The data of 235 users were remained and divided into two groups by calculating the mean hearing threshold averaged across 0.5, 1, 2, 4 kHz [19]. Please notice that some invalid data might be remained even following the rules above, since these tests were unsupervised and it was difficult to ensure the user finish the task in a right way. The listeners whose mean hearing threshold lower than or equal to 25 dB HL fell into the normal-hearing (NH) group, and the listeners whose hearing threshold exceeded 25 dB HL fell into hearing-impaired (HI) group. Finally, there were 71 and 164 users for the NH and HL groups, respectively, and the valid data for the all 235 users was named as AU (all users).

3. Results

3.1. Data analysis

The mean speech intelligibility, that is the mean score of speech audiometry averaged across 6 CF regions, and the mean hearing threshold of PTA averaged across 0.5, 1, 2, 4 kHz for each user are shown in Figure 3. Every circle/cross represents one user's data, and circles represent the users of NH group; crosses represent the users of HI group. The distribution patterns of data are different between NH and HI groups. The data of NH group distribute mainly around 1 for speech intelligibility except some outliers, but the data distribute more widely along speech intelligibility for HI group.

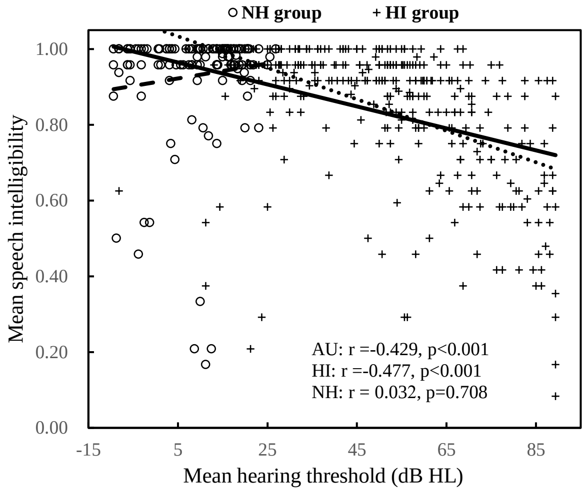

Figure 3: *The scatterplot represents mean speech intelligibility and mean hearing threshold for all users. The solid line is fitted for AU group, the dot line for HI and the dashed line for NH group.*

A correlation analysis between the mean speech intelligibility and mean hearing thresholds was conducted for the AU, NH and HI groups, respectively. The results revealed that the correlation was significant for AU group ($r=-0.429$, $p<0.001$) and HI group ($r=-0.477$, $p<0.001$). The negative r values indicated that the lower mean speech intelligibility was relevant to the higher mean hearing threshold. These results are consistent with previous studies that there is correlation between the phoneme recognition or word recognition performance and PTA, especially for hearing impaired listeners [20]. However, the correlation is not significant for NH group ($r=0.032$, $p=0.708$). Because the speech audiometry is specifically designed for HI listeners and the task is too easy for the normal-hearing, it is reasonable that their performance of speech audiometry is mainly around 1 and not significantly correlated to the hearing threshold.

When the users' performance was analyzed for each CF, the hearing thresholds at 3 kHz and 8 kHz of PTA were excluded, as the CFs for speech audiometry could only correspond to the other 6 frequencies. The correlation analysis between the speech intelligibility and the hearing thresholds for each of the six CFs was conducted, separately, and the r values for three user groups were listed in Table 1. The "whole" means that the analysis was conducted with the data for the all 6 CFs. The values which were marked with one asterisk represent the correlation is significant at the 0.05 level, and the values with two asterisks represent the significance at 0.01 level. The results showed that there was significant correlation between speech intelligibility and hearing threshold at each frequency for both HI and AU group, but not for NH group. This results were consistent with the results based on the mean speech intelligibility and mean hearing threshold. The speech intelligibility and the hearing threshold at each CF for HI group are shown in Figure 4. Every data point represents one HI user's data at one CF, and different symbols represent different frequency regions. The distribution patterns of each CF are similar. These results confirmed that the correlation between speech recognition scores and hearing thresholds was similarly in each CF region. For the significant correlations, the r values were all around -0.3, suggesting that the correlation was relatively weak at all CFs and it might be not accurate enough to evaluate audibility quantitatively at each test frequency.

Table 1: *The correlation of speech intelligibility and hearing thresholds in different CF regions for AU, NH and HI groups.*

Frequency (Hz)	**R**		
	AU	NH	HI
250	-.335**	0.092	-.329**
500	-.337**	0.178*	-.365**
1000	-.340**	-0.119	-.341**
2000	-.343**	0.013	-.385**
4000	-.282**	0.029	-.275**
6000	-.317**	0.020	-.313**
whole	-.413**	0.046	-.316**

**.Correlation is significant at the 0.01 level(2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

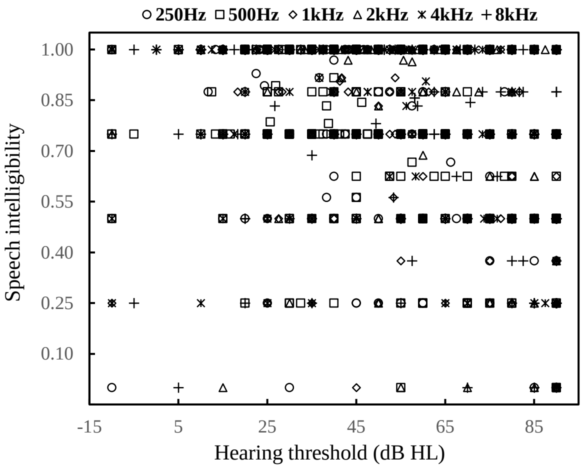

Figure 4: *The scatterplot represents speech intelligibility and hearing thresholds at each CFs.*

In summary, the significant correlation between speech intelligibility and hearing thresholds in HI group indicated that the speech audiometry with the word identification test in our App was valid in part and can help to assess individual hearing abilities on smart phones. However, although the correlation was significant, it is mild, and further work is required to improve the accuracy of the speech audiometry.

4. Discussion

The significant correlation values between speech intelligibility and hearing thresholds are all lower than 0.5, indicating the use of speech audiometry to evaluate the loss of hearing threshold could not be accurate as PTA. It could be explicated from three aspects.

Firstly, there is difficulty for user data's prescreening. Present screening criterion could only exclude the results from the tests that are not response correctly entirely, while the results from users who might response partly incorrectly could not be monitored. Besides, results from the tests that were conducted in noisy environment might also become the outliers for the whole group.

Secondly, when stimuli are disyllabic words and displayed without noise, the task of speech identification is still easy for those listeners with mild or moderate hearing impairment. For example, the proportion of the users who get speech intelligibility bigger than 0.6 is more than 90% of all the listeners. Hence the differences for individuals whose hearing impairment levels were mild or moderate were not reflected in the current results, and resulted into a relatively small correlation. In future work it is necessary to design a task whose difficulty could be adaptively adjusted according to listeners' proceeding performance.

Thirdly, the sound level for displaying speech stimuli was adjusted based on the user's individual hearing threshold. This manipulation was to ensure the listener can hear sound at a comfortable level individually, and to avoid the floor effect of speech intelligibility due to the very weak sound for HI users. However, this manipulation also brought in a sort of compensation for hearing impairment [21]. As a consequence, the performance of HI users was underestimated, and the correlation between the speech intelligibility and the hearing threshold was reduced. Moreover, for some subjects with mean hearing thresholds exceeding 40 or 50 dB HL, the stimuli level might induce a new problem for the speech perception performance, as it was reported that speech recognition scores could be reduced significantly with increasing sound level for some hearing-impaired subjects, due to loudness recruitment [22, 23]. It was also reported that the uncomfortable sound level could be high for some HI listeners [24].This individual difference between HI listeners was not taken into account in the present work.

The word pairs used in present work were selected based on Speech Banana of Mandarin Chinese, in which the location of a consonant along frequency was determined by the frequency at which the maximum energy is in the spectrogram. However, the energy of some consonants spread in a wide frequency range, and some other acoustic cue, e.g. duration, could be important for the word identification. This issue need more detailed and accurate analysis in the future work, to upgrade the present speech audiometry.

Listeners' performance for speech identification depends on several factors, including hearing threshold, frequency resolution and time resolution of auditory processing, and even high-level cognitive processing. It is important to discriminate the effects of these factors when speech audiometry is used to evaluate hearing impairment. Future work is required to use more suitable speech stimuli and more efficient test procedure for applying speech audiometry on smartphones.

5. Conclusions

A new iOS-based speech audiometry was produced, and the hearing impairment at 6 frequency regions (250-6000 Hz) could be evaluated by the identification of 36 confused disyllabic-word-pairs for people speaking Mandarin Chinese. The validity of this speech audiometry was analyzed by comparing with PTA based on 235 users. The results showed that the speech intelligibility has significant correlation with hearing threshold for hearing-impaired listeners but no remarkable correlation for normal-hearing listeners for each frequency region.

6. Acknowledgements

The work was supported by the National Natural Science Foundation of China (No. 61473008 and No. 11590773), and a Newton alumni funding by the Royal Society, UK.

7. References

[1] D. J. Van Tasell and P. Folkeard, "Reliability and accuracy of a method of adjustment for self-measurement of auditory thresholds," *Otology & neurotology*, vol. 34, no. 1, pp. 9, 2013.

[2] R. Al-Abri, M. Al-Balushi, A. Kolethekkat, D. Bhargava, A. Al-Alwi, H. Al-Bahlani, and M. Al-Garadi, "The Accuracy of IOS Device-based uHear as a Screening Tool for Hearing Loss: A Preliminary Study From the Middle East," *Oman Medical Journal*, vol. 31, no. 2, pp. 142, 2016.

[3] R. Levent, S.O. Tarik, E. Hulya, O. Ustun, and Y. M. Deniz, "Smartphone Based Audiometric Test for Confirming the Level of Hearing: Is It Useable in Underserved Areas?," *Journal of International Advanced Otology*, vol. 12, no. 1, pp. 61-66, 2016.

[4] M. Masalski, L. Kipiński, T. Grysiński, and T. Kręcicki, "Hearing Tests on Mobile Devices: Evaluation of the Reference Sound Level by Means of Biological Calibration," *Journal of Medical Internet Research*, vol. 18, no. 5, pp. e130, 2016.

[5] J. C. Yeung, S. Heley, Y. Beauregard, S. Champagne, and M. A. Bromwich, "Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones," *International Journal of Pediatric Otorhinolaryngology*, vol. 79, no. 8, pp. 1248, 2015.

[6] S. Abughanem, O. Handzel, L. Ness, M. Benartziblima, K. Faitghelbendorf, and M. Himmelfarb, "Smartphone-based audiometric test for screening hearing loss in the elderly," *European Archives of Oto-Rhino-Laryngology*, vol. 273, no. 2, pp. 333-339, 2016.

[7] M. Debevc, I. Kožuh, and H. Meier, "A Usability Requirements Analysis for Wireless Interaction and Connectivity for Elderly Hearing Aid Users," in *Human Factors in Computing and Informatics: First International Conference, SouthCHI 2013,*: Maribor, Slovenia, 2013 pp. 260-271.

[8] G. Walker and D. Byrne, "Reliability of speech intelligibility estimation for measuring speech reception thresholds in quiet and in noise," *Australian Journal of Audiology*, vol. 7, no.1, pp. 23-31, 1985.

[9] G.A. Studebaker, R.L. Sherbecoe, D.M. Mcdaniel, and C.A. Gwaltney, "Monosyllabic word recognition at higher-than-normal speech and noise levels," *Journal of the Acoustical Society of America*, vol. 105, no. 4, pp. 2431-2444, 1999.

[10] I. Mcloughlin, "Vowel intelligibility in Chinese," *IEEE Transactions on Audio Speech & Language Processing*, vol. 18, no. 1, pp. 117-125, 2010.

[11] J. Chen, H. Y. Yang, and X. H. Wu, "Evaluation of frequency-dependent hearing loss by words identification," *Journal of the Acoustical Society of America*, vol. 139, no. 4, pp. 1995-1995, 2016.

[12] G. Fant, *Acoustic analysis and synthesis of speech with applications to Swedish*. LM Ericcson, 1959.

[13] Y.C. Hung and Y.C. Ma. "Development of a Chinese Speech Banana," in *10th Asia Pacific Symposium on Cochlear Implants and Related Sciences. Beijing, China.*, 2015.

[14] P. M. Zurek and L. A. Delhorne, "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," *Journal of the Acoustical Society of America*, vol. 82, no. 5, pp. 1548, 1987.

[15] Y. Xing, Z. Fu, X.H. Wu, and J. Chen, "Evaluation of Apple iOS-based automated audiometry," in *22nd International Congress on Acoustics*: Buenos Aires. 2016.

[16] Y.M. Li and T.K. Wang, *Lexicon of Common Words in Contemporary Chinese*. The Commercial Press, 2008.

[17] International Organization for Standardization, "ISO 8253-1. Acoustics-Audiometric test methods-Part 1: Basic pure tone air and bone conduction threshold audiometry," Geneva. 1989.

[18] A.V. Hasselt, "Clinical evaluation of a computerized self-administered hearing test," *International Journal of Audiology*, vol. 51, no. 8, pp. 606-610, 2012.

[19] J. G. Clark, "Uses and abuses of hearing loss classification," *Asha*, vol. 23, no. 7, pp. 493-500, 1981.

[20] D. R. Moore, M. Edmondson-Jones, P. Dawes, H. Fortnum, A. Mccormack, R. H. Pierzycki, and K. J. Munro, "Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age," *PLoS ONE*, vol. 9, no. 9, pp. e107720-e107720, 2013.

[21] B. C. J. Moore, J. Marriage, J. Alcántara, and B.R. Glasberg, "Comparison of two adaptive procedures for fitting a multi-channel compression hearing aid," *International Journal of Audiology*, vol. 44, no. 6, pp. 345-357, 2005.

[22] D.D. Dirks, C.A. Kamm, J.R. Dubno, and T.M. Velde, "Speech recognition performance at loudness discomfort level," *Scandinavian Audiology*, vol. 10, no. 4, pp. 239-46, 1981.

[23] J. Jerger and S. Jerger, "Diagnostic significance of PB word functions," *Archives of Otolaryngology*, vol. 93, no. 6, pp. 573-580, 1971.

[24] H. Fletcher, "The Perception of Speech and Its Relation to Telephony," *Science*, vol. 108, no. 2816, pp. 682, 1950.