Comparison of the Automatic Speech Recognition System and the Traditional Method for Evaluating Speech Intelligibility among Children with Hearing Loss

Shuai Xu[1], Lu Yang[1], Yun Li[1], Hao Wu[1], Bei Li[1], and Zhi Huang[1]

[1]Department of Otolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital

August 28, 2020

Abstract

Objective The purpose of this investigation was to compare the efficacy of an automatic speech recognition (ASR) system and traditional methods for evaluating speech intelligibility in children with hearing loss. Design Observational Research Outcome Setting Department of Otolaryngology-Head and Neck Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, China Participants Ninety- four children aged 30 to 180 months with hearing loss who used hearing aids or cochlear implants prior to the examination and children with hearing loss who had congenital microtia. Main outcome measures Speech intelligibility evaluation,include the speech intelligibility score of traditional evaluation and automatic speech recognition system. Results For the speech intelligibility evaluation in the 94 children, the ASR system yielded speech intelligibility scores ranging from 0 to 100 (mean 64.3±31.1), and the traditional method yielded speech intelligibility scores ranging from 0 to 100 (mean 81.5±24.4. The correlation between the results of the traditional method and those of the ASR system was 0.895 ($p < 0.001$). Conclusion Speech intelligibility evaluation for children with hearing loss in the clinic can be performed using the ASR system. This method can increase the precision, objectivity, and efficacy of speech intelligibility assessments.

Comparison of the Automatic Speech Recognition System and the Traditional Method for Evaluating Speech Intelligibility among Children withHearing Loss

Abstract

Objective

The purpose of this investigation was to compare the efficacy of an automatic speech recognition (ASR) system and traditional methods for evaluating speech intelligibility in children with hearing loss.

Design

Observational Research Outcome

Setting

[removed for blind peer review]

Participants

Ninety- four children aged 30 to 180 months with hearing loss who used hearing aids or cochlear implants prior to the examination and children with hearing loss who had congenital microtia.

Main outcome measures

Speech intelligibility evaluation,include the speech intelligibility score of traditional evaluation and automatic speech recognition system.

Results

For the speech intelligibility evaluation in the 94 children, the ASR system yielded speech intelligibility scores ranging from 0 to 100 (mean 64.3±31.1), and the traditional method yielded speech intelligibility scores ranging from 0 to 100 (mean 81.5±24.4. The correlation between the results of the traditional method and those of the ASR system was 0.895 ($p < 0.001$).

Conclusion

Speech intelligibility evaluation for children with hearing loss in the clinic can be performed using the ASR system. This method can increase the precision, objectivity, and efficacy of speech intelligibility assessments.

Key points

1) Assessing speech intelligibility is a critical component in the assessment of communication efficacy.

2) We aim to compare the efficacy of an automatic speech recognition system and traditional methods for evaluating speech intelligibility in children with hearing loss.

3) In most cases, the automatic speech recognition score is lower than the traditional assessment score

4) The results of the speech intelligibility score of automatic speech recognition system were highly consistent with those of the traditional evaluation the speech intelligibility score of traditional

5) Speech intelligibility evaluation for children with hearing loss in the clinic can be performed using the automatic speech recognition system

Data availability statement Data available on request from the authors

Introduction

Hearing loss not only leads to restricted speech information input and understanding but also can alter speech output.[1]Speech intelligibility is an important aspect of speech-language output that allows a listener to understand what a speaker is saying. Therefore, speech intelligibility has even been considered to be the single most important practical index of oral language abilities for persons with hearing impairment.[2] This positive outlook has been bolstered by sizable gains in speech intelligibility after children with hearing loss receive sensory aids at relatively young ages.[3,4]

Traditional speech intelligibility assessments are solely based on subjective judgments.[5] Therefore, the accuracy of this assessment will strongly rely on the experience of each rater, which could affect the accuracy of evaluations. An automatic speech recognition (ASR) system could be used to predict speech intelligibility. The aim of this study was to compare the efficacy of an automatic speech recognition (ASR) system and the traditional method for evaluating speech intelligibility among children with hearing loss.

Materials and methods

Ethical considerations

This study was performed at [removed for blind peer review]. No formal ethical review by the Institutional Review Board was required. All patients provided written consent prior to this study.

Participants

This observational study was performed from February to December 2019 in [removed for blind peer review]. Ninety-four children who came to the hospital for verbal evaluation were recruited. Most of the subjects were children with hearing loss who used hearing aids or cochlear implants prior to the examination and children with hearing loss who had congenital microtia. All of the children exhibited a range of speech and language skills. Children in this study were able to identify all tested cards and cooperate to complete the test task.

No indications of mental disability, cleft palate, or pronunciation organ malformation were documented in the patients' records.

Procedure

The speech articulation test material consists of 25 disyllables that are common in children's daily lives. The target disyllable is written on a card with a vivid picture, and the child was asked to name the picture in the card alone or under guidance. Each of the disyllables was repeated three times. The raters and children were seated on the opposite side of each other around a table. They wrote down what they heard and were also asked to mark every unintelligible word with a cross. The speech intelligibility score was averaged across the four raters and obtained by traditional evaluation is marked as SIS^{TE} (Intelligibility Score of Traditional Evaluation). At the same time, an iPhone equipped with iFly software was placed 20 centimeters directly in front of the child in a quiet environment. The speech recognition system recorded the pronunciation of the child and then output the text. The speech intelligibility score of the ASR system is marked as SIS^{ASR}(Intelligibility Score of *ASR system*) and is based on the correctness of the output text.

Traditional evaluation (TE) of speech intelligibility

Traditional evaluation (TE) of speech intelligibility was measured by four adult raters. Based on the different degrees of familiarity with deaf speech, the raters were divided into three different levels: the level-1 rater is very familiar with the speech spoken by deaf people, the level-1 rater is marked as SIS^1; the level-2 rater has limited experience with the speech spoken by deaf people, and marked as SIS^2; the level-3 rater has no experience with the speech spoken by the children with hearing loss, and marked as SIS^3. In our study, the four raters contained two level-3 raters, one level-1 rater and one level-2 rater.

ASR system evaluation of speech intelligibility

In this study, we used version 4.8 of iflynote software (Automatic Speech Recognition System, Hefei, Anhui, China) developed by the iFly company. iFly developed a speech recognition system using a feed-forward sequential memory network (FSMN), which is a deep convolutional neural network in which image recognition is introduced, and its recognition framework is developed with a core spectrum recognition technique created by the company.

Statistical analysis

This study is an observational study. Descriptive statistics are presented as the mean and standard deviation (SD) and median and interquartile range (IQR) when appropriate. Interobserver agreement was analyzed by calculating intraclass correlation coefficients for the composite SIS^{ASR} and the three SIS^{TE}. Item-level interobserver agreement was assessed by calculating Cohen's $\varkappa$ statistic and the percentage of absolute agreement. A value of 0.7 was taken as an acceptable level of agreement for both the ICC and Cohen's $\varkappa$, preferably for the lower border of the 95% CI.

Results

In total, 94 children (47 girls and 47 boys) with hearing loss were recruited, and their ages ranged from 30 to 180 months old (mean 79.6±37.3 months, median 72 months). All 94 children completed the test. In brief, SIS^1 were the highest, followed by SIS^2 and then SIS^3. The interquartile range (IQR) for SIS^{ASR} was 40 to 92, with a median of 72. The IQR for SIS^1 was 77.5 to 100, with a median of 96. The IQR for SIS^2 was 74 to 100, with a median of 94. The IQR for SIS^3 was 71.4 to 98.7, with a median of 89. The scores of the tests are shown in Table 1.

To intuitively analyze the differences between different raters, 20 of the 94 children were randomly sampled in the final analysis using the SPSS software program. Among them, the SIS^{ASR} for 17 children was lower than the subjective assessments. Only the SIS^{ASR} for two children was higher than the SIS^1. Only the SIS^{ASR} for one child was higher than the SIS^3. Figure 1 shows the scores of subjective evaluations (including three levels) and the scores of the automatic speech recognition system.

The correlations between the SIS^{ASR} and the SIS^{TE} are shown in Figure 2. The overall trend showed that the higher the SIS^{ASR}, the higher the SIS^{TE}, indicating a positive correlation between the scores.

SIS^{ASR} showed perfect consistency with the SIS^{TE}, as shown in Table 2. The lowest correlation between a rater and the mean of the other two levels was 0.928 (SIS^1 vs SIS^3), and the highest was 0.971 (SIS^2 vs SIS^3). Table 2 also shows the correlations between different raters. The correlation between the traditional evaluation group and automatic speech recognition is 0.883. This means that the agreement between the human raters and the machine and the agreement among the humans alone can be regarded as significant.

Discussion

Synopsis of key findings

Until now, recordings of only 94 children have been evaluated, and the results indicated that the method could yield precise information in evaluating intelligibility. The ASR score is lower than the traditional assessment score. In most cases, the ASR recognition gap is large, but when the score is higher, the gap decreases. The possible reasons are as follows. The intelligibility of speech depends on five factors: the speaker, the speech (content familiarity), the content length, the grammatical complexity or perplexity and the input medium.[6] When the ASR system assesses the intelligibility of speech from an individual, it compares the spectral and temporal characteristics of the signal of the acoustic speech samples with according transliteration. The internal recognition system had been trained with acoustic information from normal speech speakers, so the system will be only able to accurately recognize speech that is highly intelligible and will be more error-prone when given speech that differs greatly from normal speech.

Strengths and

limitations

Assessing speech intelligibility is a critical component in the assessment of communication efficacy[2]. In traditional evaluation the speaker remains the most influencing factor because people could judge by "feeling".[7-10] In contrast, ASR system calculations are based on the input voice information without being affected by other factors. Despite this limitation, the results of the ASR system were highly consistent with those of the traditional evaluation.

The limitation of ASR is that the evaluation score is too low when the given speech is less intelligible. In most cases, the ASR score is lower than the traditional assessment score. It does not mean that ASR assessment does not show the true level of the child; on the contrary, the ASR assessment is more sensitive to nonstandard pronunciation from children because it can truly reflect the current speech level of the children. However, it should be noted that different ASR software programs and different versions might also affect the results, causing a certain impact because of different internal dictionary systems.

Clinical applicability

We believe that the clinical intelligibility of children with hearing loss can be evaluated using the ASR system. Further research is in progress to enhance the possibilities of different versions of automatic speech recognition, as it can also be of special interest as application in a medical field. In addition, the ASR system can also try to assess the articulation intelligibility of Mandarin-speaking children, such as cleft lip and palate, dysarthria, childhood apraxia of speech and aphasia. It can also further complete the intelligence of diagnosis and directly match the recognition result with the correct answer to obtain the final score.

Conclusions

Automatic speech recognition for speech intelligibility evaluation in children with hearing loss is a valuable means for research and clinical practice. It enables the quantification of the quality of speech. An automatic evaluation will lead to quick results without time-consuming perceptive evaluation, and the results will be accurate and reliable.

References

1 Hoffman, M. F., Cejas, I. & Quittner, A. L. Health-Related Quality of Life Instruments for Children With Cochlear Implants: Development of Child and Parent-Proxy Measures. *Ear Hear* . 2019;40(3): 592-604.

2 Chin, S. B., Bergeson, T. R. & Phan, J. Speech intelligibility and prosody production in children with cochlear implants. *J Commun Disord* . 2012;45(5): 355-366.

3 Chin, S. B., Tsai, P. L. & Gao, S. Connected Speech Intelligibility of ChildrenWith Cochlear Implants and Children With Normal Hearing.*American Journal of Speech-Language Pathology* . 2003;12(4): 440-451.

4 Ertmer, D. J. & Jung, J. Prelinguistic vocal development in young cochlear implant recipients and typically developing infants: year 1 of robust hearing experience. *J Deaf Stud Deaf Educ* . 2012;17(1): 116-132.

5 Wolk, D. B. S. The relative effects of three expressive methods upon the speech intelligibility of profoundly deaf speakers. *Journal of Communication Disorders* . 1986;19(3): 209-218.

6 Fontan, L. *et al.* Automatic Speech Recognition Predicts Speech Intelligibility and Comprehension for Listeners With Simulated Age-Related Hearing Loss. *J Speech Lang Hear Res* . 2017;60(9): 2394-2405.

7 Landa, S. *et al.* Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding. *Int J Speech Lang Pathol* . 2014;16(4): 408-416.

8 Miller, N. Measuring up to speech intelligibility. *Int J Lang Commun Disord* . 2013;48(6): 601-612.

9 Ziegler, W. & Zierdt, A. Telediagnostic assessment of intelligibility in dysarthria: a pilot investigation of MVP-online. *J Commun Disord* . 2008;41(6): 553-577.

10 D'Innocenzo, J., Tjaden, K. & Greenman, G. Intelligibility in dysarthria: effects of listener familiarity and speaking condition.*Clin Linguist Phon* . 2006;20(9): 659-675.

Hosted file

`Table 1.docx` available at https://authorea.com/users/353900/articles/477639-comparison-of-the-automatic-speech-recognition-system-and-the-traditional-method-for-evaluating-speech-intelligibility-among-children-with-hearing-loss

Hosted file

`Table 2.docx` available at https://authorea.com/users/353900/articles/477639-comparison-of-the-automatic-speech-recognition-system-and-the-traditional-method-for-evaluating-speech-intelligibility-among-children-with-hearing-loss

Hosted file

`Fig. 1.docx` available at https://authorea.com/users/353900/articles/477639-comparison-of-the-automatic-speech-recognition-system-and-the-traditional-method-for-evaluating-speech-intelligibility-among-children-with-hearing-loss

Hosted file

`Fig. 2.docx` available at https://authorea.com/users/353900/articles/477639-comparison-of-the-automatic-speech-recognition-system-and-the-traditional-method-for-evaluating-speech-intelligibility-among-children-with-hearing-loss