Speech Reception Threshold Measurement Using Automatic Speech Recognition

Emre Yılmaz
ESAT, KU Leuven
Leuven, Belgium

Joris Pelemans
ESAT, KU Leuven
Leuven, Belgium

Stefan Lievens
Cochlear Technology Center
Mechelen, Belgium

Hugo Van hamme
ESAT, KU Leuven
Leuven, Belgium

Abstract

Various speech tests have been proposed for measuring hearing abilities of both normal hearing and hearing-impaired people. One well-known measure is the speech reception threshold (SRT) defining the signal-to-noise ratio (SNR) level at which the speech recognition rate of a person is 50%. The SRT measurement is relevant for quantifying the hearing abilities of normal listeners, periodic screenings of cochlear implant (CI) patients, algorithm development in hearing aids and CIs and psychoacoustic research. In this work, we demonstrate our efforts on integrating an automatic speech recognizer into a conventional SRT measurement software utilized by audiologists. Adopting such an automatic evaluation scheme is expected to reduce human effort which can then be invested in more vital tasks, e.g. psychoacoustic research or providing additional assistance to CI patients. The proposed system has shown to be viable providing sufficiently accurate SRT estimates.

1. Introduction

Speech reception threshold (SRT) measurements have been used for evaluating a listener's hearing capabilities and diagnosing hearing loss [14]. In a clinical setting, the SRT value is a subjective measure for quantifying the hearing ability of patients with cochlear implants (CI) in order to adjust the CI parameters and analyze the impact of new developments in CI devices on the patient's hearing abilities [15]. Moreover, these measurements provide useful data for psychoacoustic research [13].

In practice, the SRT measurements are performed by repeated and adaptive tests of the up-down type as described in [9] performed by an audiologist. The speech material used during these tests has to be designed carefully in order to obtain accurate SRT estimates [10]. Several Dutch speech tests have also been proposed for determining a patient's SRT, e.g. NVA-tests [19] and LIST-tests [18]. During these tests, words or sentences which are embedded in different levels of noise are presented to the patients and they are asked to repeat what they have heard. The responses are evaluated by an audiologist who decides if patients properly repeat the presented word or sentence. LIST tests consist of ten sentences that are presented to a patient at a certain noise level. For each sentence, two to five keywords are defined. Each keyword in the patient's response is evaluated by the audiologist and if all keywords were reproduced correctly (incorrectly), the noise level in which the following sentence is embedded is increased (decreased) by 2 dB resulting in a more (less) challenging recognition task. After presenting ten sentences, the SRT value is obtained by averaging the signal-to-noise ratio (SNR) levels at which the last six sentences are presented. This SRT value corresponds to the SNR value where 50% of the words are understood correctly by the patient.

Automation of these tests was investigated in [16, 8] by letting the patients type what they have heard while accounting for spelling errors. A rehabilitation tool for CI users using automatic speech recognition (ASR) is described in [11]. CI patients are encouraged to repeat spoken sentences upon which correctness feedback is provided using ASR. The proposed system for SRT tests is similar in recognition task, but differs in the language model constraints since the main task is to detect the keywords rather than recognition of the complete utterance. As the expected utterances are known in the scope of this paper, the use of deterministic language models is feasible. For this purpose, a flexible nonsequential finite state grammar (FSG) has been used in order to be able to accept responses with incorrect word order.

To the best of our knowledge, this is the first effort to establish an automatic evaluation scheme for SRT measurements. The SRT measurement procedure was identified as one that was particularly apt for automation since it seems feasible to set up an automatic speech evaluation method that makes significantly fewer errors than the patient under test who has a recognition error around 50% by definition. Once the error rate of the speech recognizer is small enough compared to 50%, these errors are expected not to affect the test outcome significantly. The initial experimental findings towards such an automated system using an automatic

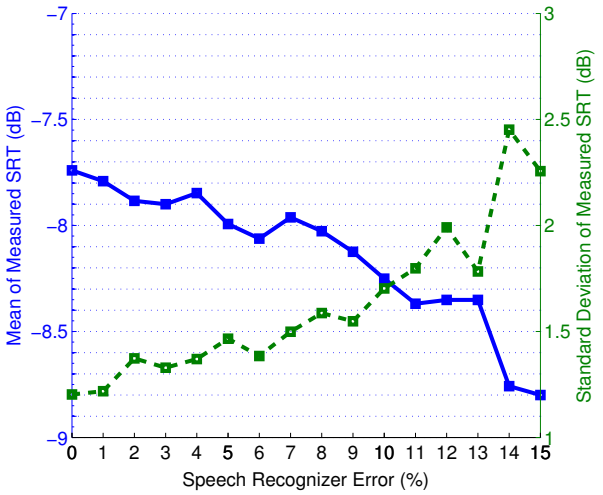


Figure 1. Impact of Recognizer Errors on the SRT Measurement

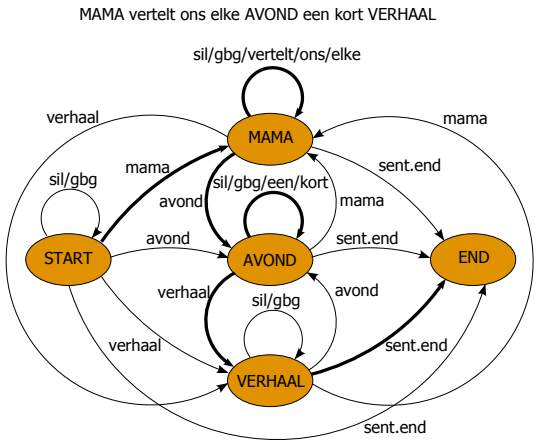


Figure 2. Example of an FSG model for a LIST sentence

speech recognizer are given in [4]. Particularly, the impact of recognizer errors on measured SRT values has been investigated which is presented in Figure 1. From these results, it can be concluded that the recognizer errors have a negligible impact on the measured SRT value.

Practically, the benefits of using the automated system are twofold. Firstly, an automated test provides the benefits of an objective and repeatable measurement compared to an audiologist whose evaluation may be biased. Moreover, automating this procedure saves a great amount of time in which audiologists could intensify their assistance to CI patients and psychoacoustic research.

The rest of the paper is organized as follows. Section 2 introduces the speech recognizer's architecture. The implementation details are described in Section 3. The paper is concluded in Section 4.

2 Automatic Speech Evaluation Scheme

2.1 ASR overview

A two-layered HMM-based recognition system has been used for obtaining the word-level recognition output. A similar recognizer has been applied to a reading level assessment task of pupils and provided a significantly better ROC curve compared to a single-layered recognizer [5]. In the first layer, a phone recognizer generates a phone lattice using task-independent acoustic and language models. These models can be general models that are trained on the data of the target language. In the second layer, task-dependent information is provided in the form of an FSG describing lexical and grammatical knowledge. This structure comes with increased modularity as the generic phone recognizer can be used for any recognition task provided that the task-specific information is contained in the second stage [7]. Using the task-dependent information incorporated in the FSGs, the phone lattice obtained in the previous step is decoded into a word level recognition result which can further be processed to obtain the keywords that have been uttered.

2.2 Task-specific Language Models

The SRT measurement procedure is well-structured in the sense that the recognizer has access to the sentence that is presented to the patients via headphones. As the sentences that have to be recognized are known in advance, using FSGs is feasible for this recognition task. Due to the nature of SRT measurement tests, it is a requirement to have higher flexibility in the FSG as the test subject is only scored on the selected keywords in the sentence and is allowed to repeat the keywords in an arbitrary order. Non-keywords can be skipped, inserted or substituted.

An example FSG is illustrated for the Dutch sentence "MAMA vertelt ons elke AVOND een kort VERHAAL" (MOM reads us a short STORY every NIGHT) in Figure 2, where keywords are written in uppercase characters. The arcs that models the correct sentence are given in bold. The start and end nodes are marked as "START" and "END" respectively. All other nodes are labeled with the keywords: visiting a state indicates the associated keyword was detected. State transitions occur upon a match between a word or phrase model and a partial path in the phone lattice generated by the first layer. Non-keywords, silence (marked as "sil") and garbage words (marked as "gbg") result in a self-transition. Garbage words stand for any unanticipated utterances which are different from the presented sentence. To avoid being it often preferred over other arcs, it is penalized with a *garbage model cost* that is incurred once upon entry.

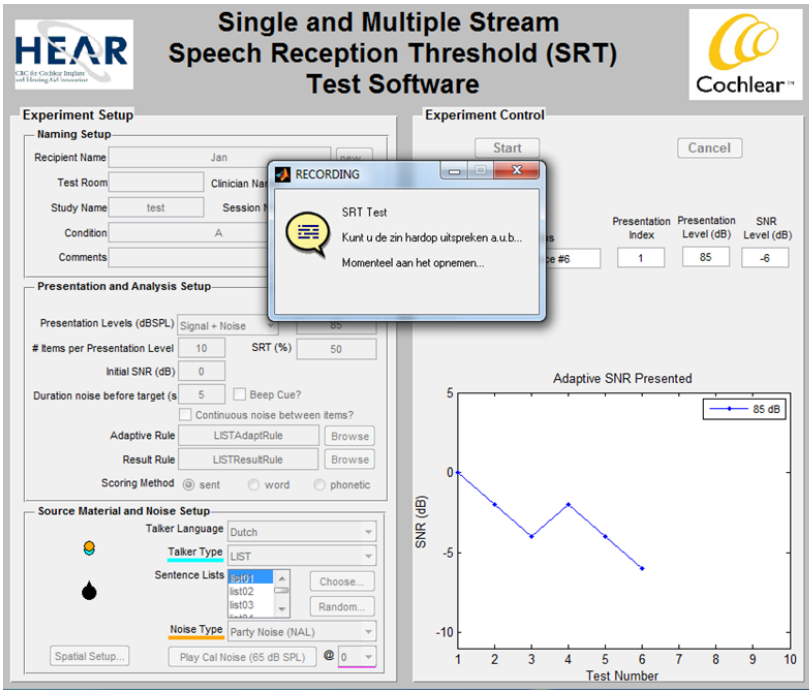


Figure 3. Recording the patient's response

3 Implementation Details

3.1 The Speech Recognizer

The proposed system uses the Dutch recognizer developed as a part of the SPRAAK recognition toolkit[1]. The acoustic models were trained on the CoGen database [3] which contains 7 hours of read speech. The speaker independent acoustic models are semicontinuous HMMs with tied Gaussians consisting of 576 states and 10635 Gaussians. The task-independent language model consists of a trigram phoneme sequence model derived from a Dutch database with correctly read sentences [6]. The preprocessing is based on Mel-spectrum analysis and includes cepstral mean subtraction and discriminant analysis (MIDA) [2].

3.2 Baseline SRT Measurement Software

The baseline SRT measurement software has been described in [1]. The measurement procedure is implemented in MATLAB with a user interface containing several input textboxes. In this software, the audiologist evaluates the pronunciation of the patient after each sentence and marks the correctedly pronounced keywords manually. The SNR level of the upcoming sentence is adjusted depending on the evaluation. After all ten sentences are presented, the SRT level is calculated by averaging the last six SNR levels.

[1] `www.spraak.org`

3.3 Modifications Towards an Automatic System

The software is modified in a way that patient responses are recorded for a variable duration depending on the duration of the presented sentence. A screenshot of the software recording the patient's response is illustrated in Figure 3. Then, the recording is sent via HTTP to a RESTful web service which performs keyword detection as described in Section 2. Both the client and server were built using the CLAM application wrapper [17] with which we have already built several ASR services [12]. In this way, the SRT measurement can be performed by the patient without the guidance of the audiologist after a brief tutorial about the measurement procedure. A demonstration of the automatic SRT measurement procedure is available in `http://www.esat.kuleuven.be/psi/spraak/demo/srt/`.

4 Conclusion

An automated evaluation scheme for speech reception threshold measurements has been implemented using automatic speech recognition technology. The proposed setup is based on conventional SRT measurement software as it is used by audiologists today. The future work includes an in-depth analysis of the system with cochlear implant patients and investigation of the possible practical limitations.

References

[1] P. W. Dawson, S. J. Mauger, and A. A. Hersbach. Clinical evaluation of signal-to-noise ratio-based noise reduction in nucleus cochlear implant recipients. *Ear and Hearing*, 32(3):382–390, 2011.

[2] K. Demuynck, J. Duchateau, and D. Van Compernolle. Optimal feature sub-space selection based on discriminant analysis. In *Proc. Eurospeech*, pages 1311–1314, 1999.

[3] K. Demuynck, D. Van Compernolle, C. Van Hove, and J.-P. Martens. CoGen een corpus gesproken Nederlands voor spraaktechnologisch onderzoek - eindverslag. *Tech. Rep. K.U. Leuven - ESAT & Universiteit Gent*, 1997.

[4] H. Deprez, E. Yılmaz, S. Lievens, and H. Van hamme. Automating speech reception threshold measurements using automatic speech recognition. In *4th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, pages 35–40, Grenoble, France, Aug. 2013.

[5] J. Duchateau, K. Demuynck, and H. Van hamme. Evaluation of phone lattice based speech decoding. In *Proc. INTERSPEECH*, pages 1179–1182, Brighton, UK, Sept. 2009.

[6] J. Duchateau, Y. O. Kong, L. Cleuren, L. Latacz, J. Roelens, A. Samir, K. Demuynck, P. Ghesquière, W. Verhelst, et al. Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules. *Speech Communication*, 51(10):985–994, 2009.

[7] J. Duchateau, M. Wigham, K. Demuynck, and H. Van hamme. A flexible recognizer architecture in a reading tutor for children. In *Proc. of the ITRW on Speech Recognition and Intrinsic Variation*, pages 330–331, Toulouse, France, May 2006.

[8] T. Francart, M. Moonen, and J. Wouters. Automatic testing of speech recognition. *International Journal of Audiology*, 48(2):80–90, 2009.

[9] H. Levitt. Adaptive testing in audiology. *Scand. Audiol. Suppl.*, 6(6):241–291, 1978.

[10] M. Nilsson, S. D. Soli, and J. A. Sullivan. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.*, 95(2):1085–1099, 1994.

[11] W. Nogueira, F. Vanpoucke, P. Dykmans, L. De Raeve, H. Van Hamme, and J. Roelens. Speech recognition technology in CI rehabilitation. *Cochlear Implants International*, 11(Supplement 1):449–453, 2010.

[12] J. Pelemans, K. Demuynck, H. Van hamme, and P. Wambacq. Speech Recognition Web Services for Dutch. In *the International Conference on Language Resources and Evaluation (LREC)*, May 2014.

[13] R. W. Peters, B. C. Moore, and T. Baer. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J. Acoust. Soc. America*, 103(1):577–587, 1998.

[14] R. Plomp and A. M. Mimpen. Speech-reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. America*, 66(5):1333–1342, 1979.

[15] M. K. Qin and A. J. Oxenham. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J. Acoust. Soc. America*, 114(1):446–454, 2003.

[16] H. Terband and R. Drullman. Study of an automated procedure for a dutch sentence test for the measurement of the speech reception threshold in noise. *J. Acoust. Soc. Am.*, 124(5):3225–3234, 2008.

[17] M. van Gompel. *CLAM: Computational Linguistics Application Mediator. Documentation. ILK Technical Report 12-02*, 2012. http://ilk.uvt.nl/downloads/pub/papers/ilk.1202.pdf.

[18] A. Van Wieringen and J. Wouters. LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *International journal of audiology*, 47(6):348–355, 2008.

[19] J. Wouters, W. Damman, and A. J. Bosman. Vlaamse opname van woordenlijsten voor spraakaudiometrie. *Logopedie: informatiemedium van de Vlaamse vereniging voor logopedisten*, 7(6):28–34, 1994.