Speech Communication 99 (2018) 47–55

Contents lists available at ScienceDirect

Speech Communication

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

Effect of talker variability on hearing aid benefit with closed word recognition test[☆]

Christophe Lesimple*, Barbara Simon, Julie Tantau

Bernafon AG, Berne, Switzerland

ARTICLE INFO

Keywords:
Speech intelligibility
Response time
Hearing aids benefit
Talker variability

ABSTRACT

The relationship between hearing aid benefit measured in a clinical setting and hearing aid performance in daily listening situations is not clearly established. A significant aspect of realistic listening situations is the variety of talkers that people normally encounter. This is a daily listening obstacle that is not easily recreated in a clinical setting; however, with the appropriate speech testing material the results could contribute to a better understanding of the benefit of amplification in realistic listening situations. Our present study investigated the interaction between talker variability and amplification benefit. Nine experienced hearing aid wearers with moderate to severe hearing loss were tested in aided and unaided conditions with the original and a multi-talker version of the WAKO word recognition test. Word recognition scores and response times were simultaneously measured for each test condition and test version. Variability of the test material was quantified in terms of the fundamental frequency and duration of each test word. The statistical analysis used a logistic mixed effect regression model on the word recognition scores and a linear mixed effect regression model on the response times. Introducing talker variability in the test material led to a significant decrease in amplification benefit for word recognition accuracy. These results suggest that using speech materials closer to that experienced in daily life might change the magnitude of measured benefit from hearing aids amplification.

1. Introduction

Speech testing is used regularly in research as well as in the clinic to demonstrate the benefit of hearing aids. The significant results achieved in a simulated clinical environment often do not equal those reported by clients using the hearing aids in the real world. We must then ask ourselves the question: How do our test results transfer to the real world? It might be challenging to give a simple answer to that question as the speech test parameters (classified in stimulus, presentation, response and subject variables) directly influence its outcomes (Theunissen et al., 2009). The hearing aid evaluation can also include questionnaires designed to reflect the listeners' satisfaction under various conditions in daily life listening environments. However, the correlation between the aided speech test results and self-reported hearing aid outcomes might not be systematic (Humes et al., 2009; Ng et al., 2013; Humes et al., 2017). A possible explanation might be found in the properties of the evaluation tools i.e. how well the test captures aspects of speech in daily listening conditions.

Speech testing in the clinic generally consists of one speaker with the option to add background noise. The background noise may include various voices or just a broadband noise matching the target signal, but the speaker is generally the same voice throughout the entire test. This would be realistic if everyone had only one conversation partner at all times. However, the reality is that there are often multiple people speaking to us and sometimes all at once. To add further difficulty, the speakers are not always speaking the same language or may not be native speakers of the language in which they are conversing. McCloy et al. (2015) found that variation in speech intelligibility can also be explained by the talker origin once acoustical measures of the talker's speech where taken into account in the analysis. In addition to their accent, people have different methods of annunciation, speaking rates, and different levels of volume and pitch. Introducing talker variability into speech test material lead to poorer intelligibility for normal hearing listeners (Mullennix et al., 1989) as well as for hearing impaired listeners (Kirk et al., 1997). For the test with hearing impaired listeners, the word-recognition performance measured with multi-talker test material was better correlated with self-reported listening ability. These findings suggest that a multi-talker speech test might reflect more aspects of speech perception and difficulties met in daily listening situations as compared to standardized clinical tests with one talker. A

[☆] Parts of this article have been presented at the SPIN (Speech in Noise) workshop 2016 in Groningen, NL.

* Corresponding author.
E-mail address: cles@bernafon.com (C. Lesimple).

https://doi.org/10.1016/j.specom.2018.02.007
Received 30 June 2017; Received in revised form 8 February 2018; Accepted 12 February 2018
Available online 09 March 2018

typical example is a hearing aid user listening to a trained speaker, e.g. news anchor, compared to a program with multiple talkers. These situations provide different difficulty levels which may increase or decrease the users' intelligibility.

Another aspect of evaluating speech perception is that the scores are most often recorded in percent correct of the presented material (phonemes, words, or sentences). However, speech perception is multidimensional by nature and other aspects like sound quality or listening effort might also contribute to the differences between test conditions (Devocht et al., 2017) or hearing aid features, e.g. noise reduction (Brons et al., 2014). The listening effort dimension can be evaluated with various methods like subjective, behavioral, or physiological tests (Ohlenfrost et al., 2017). One interesting method for measuring listening effort is the use of response time during a speech intelligibility task (Houben et al., 2013; Pals et al., 2015). Interpretation of changes in response time can be found in the ease of language understanding (ELU) model (Rönnberg et al., 2013). The ELU model assumes that the cognitive processes, which make auditory speech input understandable in optimal listening conditions, are fast and automatic within an episodic buffer. The episodic buffer is a processing component that matches the incoming phonological information with existing language representations stored in the long-term memory (LTM) to quickly understand speech. This process is implicit and fast only with an incoming signal that is not degraded by noise (Houben et al., 2013) or distorted by a hearing loss (Carroll et al., 2016). A degraded auditory input signal makes this association more complex, which requires additional explicit cognitive processes. This effect can be measured with an increased response time for hearing impaired listeners despite compensation for audibility as compared to normal hearing listeners (Carroll et al., 2016). Working memory capacity with semantic LTM and phonological processing is required to understand a degraded input signal. These explicit processes run on a slower time scale measured in seconds. This time can be shortened by improving the signal-to-noise ratio of the incoming signal (Houben et al., 2013) or by simplifying the listening situation by reducing the number of talkers (Mullennix et al., 1989). It is possible that by combining percent correct scores with response time measures, the sensitivity of a test can be improved as shown by Mackersie et al. (1999) with their results using the modified rhyme test.

Adding response time as an outcome scale increases the potential to give additional information on other aspects of speech perception. Under specific test conditions, it has been shown that word recognition accuracy decreases and response latency increases when talker variability is introduced in word recognition tests compared to a single talker condition for normal hearing (Mullennix et al., 1989; Bent and Holt, 2013) and for hearing impaired listeners (Kirk et al., 1997). The use of a single professional speaker within a commercially available speech test might underestimate the real difficulties met in daily life by hearing impaired listeners and limits the ability to generalize findings on speech perception (Clark, 1973). We suggest that this might explain why the relationship between reported subjective hearing aid benefit and speech recognition improvement with clinical tests is not systematically and clearly established (Cord et al., 2000; Mendel, 2007).

Creating speech test material that includes one aspect of the variability of speech met in daily listening situations has the potential to:

- Decrease speech perception performance on different scales such as word recognition and response time.
- Reduce the ceiling effect for test conditions closer to those met by hearing aid wearers in daily life (Smeds et al., 2015).
- Show better correlation with the subjective benefit and speech perception scores obtained during tests in controlled lab environments.

While the effect of multi-talkers and benefit of amplification on speech recognition are well documented separately, the combination of both effects, to our knowledge, is not addressed. Our research questions are (1) Is speech perception (word recognition and response time) affected when talker variability is introduced in the test material and (2) Is the effect of amplification influenced when some talker variability is introduced in a speech perception test? The first question should confirm findings from previous research, while the second one should give new insights about measuring the benefit of amplification when one additional aspect of daily listening condition is introduced in the speech test design.

2. Multi-talker test material

In an effort to add some variability to a widely used speech test for German speakers, a multi-talker version of the WAKO rhyme test (von Wallenberg and Kollmeier, 1989) was recorded by four non-professional talkers. The original WAKO rhyme test was evaluated by Kollmeier et al. (2011) and was shown to be suitable to verify hearing aid benefit. Using a closed-set test has the advantage to precisely control and measure response times during a speech test (Mullennix et al., 1989; Mackersie et al., 1999). The original speech material for this experiment is provided by Hörtech GmbH in Oldenburg and will be labelled as the single-talker version with ST. The multi-talker version is intended for research and development purposes only and not to measure speech reception thresholds in a clinical practice. The recorded multi-talker material will be labelled with MT.

2.1. Test material selection and recordings

The original test stimuli consist of various lists of 47 original WAKO words spoken by a male speaker. Each list is built with consonant-vowel-consonant (CVC) words that are phonetically balanced and close to the German phoneme occurrence distribution (von Wallenberg and Kollmeier, 1989). The tested phoneme is the consonant that is either at the start or at the end of the target word. The tested phoneme is also categorized by the manner of articulation: stop, fricative, nasal, and lateral/trills. Four lists which have the same occurrence of each target consonant's manner (15 stops, 14 fricatives, 9 nasals, and 9 lateral/trills) were selected for the multi-talker recording.

The MT rhyme test material was produced using non-professional speakers: two male and two female native German speakers with either a German or a Swiss German (an Alemannic dialect) accent. This variability should represent some situations that are encountered in daily communication in the German speaking part of Switzerland. While the official language used at school, in the administration or in the media is German, most daily conversations are in Swiss German. Meyer et al. (2011) suggest that adding accent as factor, with 2 speakers per region, could increase error rates during phoneme recognition tests for normal hearing listeners.

The test material itself consists of four lists containing 47 words from the original WAKO rhyme test. Each word includes a carrier sentence: "Das Wort…" (The word…). The recording was made in a single-walled sound room with a Reloop sPOD USB microphone at a sampling rate of 44.1 kHz with 16-bit quantization. Talkers were instructed to read the test list with the carrier sentence for each word without making any particular effort for articulation or reading speed. They could repeat any item as often as needed in order to create an acceptable sample.

All the test items with the carrier sentence were saved in individual files of 1.5 s each and the end of the target word was aligned with the end of the test file. The level of all the files, including sentence and target word, was normalized on RMS values with Adobe Audition (version 1.5).

2.2. Test material characteristics

Speech intelligibility is influenced by various acoustic features,