Effect of a moving distractor on speech intelligibility in babble noise using a digit-triplet test

Rhoddy Viveros, Z. Ellen Peng, Florian Pausch, Janina Fels

Institute of Technical Acoustics, Medical Acoustic Group, RWTH Aachen University, Kopernikusstraße 5, 52074 Aachen, Germany, E-Mail: rvm@akustic.rwth-aachen.de

Abstract

The speech reception threshold (SRT) is measured to assess speech recognition under the presence of distracting noises. However, most laboratory test conditions do not reflect the complexities of real-world listening, such as listening with moving noise sources. In this study, a standard digit-triplet test for measuring SRT at 50% intelligibility is adopted to investigate the effect of moving babble noise source. A listening experiment with normal-hearing adults is performed, using binaural technologies to create a headphone-based acoustic virtual environment (VAE). The listeners are asked to identify digit-triplets spoken by the target talker, always positioned in front (0° azimuth), under the presence of static and moving babble noise also produced by the target talker. The SRTs are measured within-subject in both static and dynamic scenes under VAE. In the dynamic scene, four conditions consisting of varying directions of the dynamic trajectory are tested, by moving the babble source away from (from 0° to +/-90° azimuth) versus toward (from +/-90° to 0° azimuth) the target talker at a constant angular velocity. The SRTs measured in the dynamic scenes are compared to those measured in the static scenes, in which the babble source is positioned statically at 0°, +90°, and -90° azimuth.

Introduction

The major complaint of listeners with hearing loss is that they have difficulties in understanding speech with background noise. Conventional pure-tone audiometry is not sufficient to assess the problems in everyday life listening situations thus listening-in-noise tests are needed [7].

In real-life listening situations, we are confronted with multiple sound sources, either static or dynamic, that disturb intelligibility in speech perception. In natural acoustic scenes, conversations may become very difficult to understand with masking noise that has dynamic movements in space. Due to technical limitations in creating moving sound sources in listening experiments, little is yet known on the masking effect of a moving noise source.

One of the frequently used tests to assess intelligibility under background noise is the listening-in-noise test. The purpose of the test is to find the speech reception threshold (SRT) which is the specific signal-to-noise ratio (SNR) at which the listener is able to understand 50% of the message (e.g., digits, groups of vowels and consonants, sentences, stories). SRT values are given in units of dB.

This study seeks to the effect of moving noise distractors on speech intelligibility using a digit-triplet test.

The effect of moving sound sources has been studied in several publications [4], [5], [8], [9]. However, little is known about the binaural measurement of intelligibility under moving babble-noise.

In the current study, the SRT of each participant was measured using a listening-in-noise test which uses digit triplets. Due to the repeated test conditions in this investigation, a set of 14 parallel digit-triplet tests were developed at the Institute of Technical Acoustic (ITA) Aachen based on recommendations proposed by Akeroyd et al. (2015) [1] and Zokoll at al. (2012) [10].

Although the use of the digits may seem less representative of real-life listening, the digits are in the category of highly familiar words and most frequently spoken words. Additionally, such speech materials require low linguistic abilities of the listeners and can be easily understood albeit auditory disabilities. Moreover, the simple digit-triplet test is more appropriate than a sentence test for the evaluation of hearing aids or cochlear implants fittings [7], which is appropriate for continuing research in the future. Digit tests were also commonly used in tests of cognitive abilities, such as attention capacity, which is very well aligned with the investigations of the cocktail party effect in auditory perception [2], [3].

In this investigation, the intelligibility of digit-triplets is investigated under the presence of a babble noise source either statically located or dynamically moving along a horizontal trajectory in the virtual scene. The virtual sound sources were presented binaurally via headphones in a sound attenuated listening booth. For the binaural reproduction, a set of head-related transfer functions (HRTF), measured from the ITA artificial head, was convolved with the stimuli to be rendered in free-field conditions. All virtual sound sources were simulated using the real-time software Virtual Acoustics (VA), developed at the ITA.

Method

Stimuli

Stimuli consist of digit-triplets in German. The digit-triplet test was developed under the guidelines from [1], [7] and [10] and consists of three monosyllabic digits (0-9, excluding 7) recorded by a female German native speaker. In the digit-triplet test, the number of syllables for the digits needs to be considered to avoid a certain digit being recognized purely by its unique number of syllables. As a result, the German digit-triplet test does not contain the number seven (unique digit with two syllables).

The masking noise was a randomized superposition of all digits in the test, with a random delay (up to 4 s) between successive repetitions of the speech items. This results in a quasi-stationary noise that has the same long-term average spectrum as the target speech.

After optimization and evaluation of the digit-triplets, 14 lists each containing 24 unique digit-triplets were created with different digit-triplets, provided that each digit in each position must appear three times. In the validation with human listeners, only seven lists were tested due to long experimental condition. It is assumed that the psychometric functions derived from the seven lists are similar to the other seven lists untested, since all 14 lists were created in the same way.

The results of the digit-triplet test optimization are shown in Table 1, where S_{digit} is the slope of the psychometric function and the SRT is the mean SRT with the corresponding error σ.

Table 1: Results of the ITA German digit-triplet test optimization

	S_{digit} [%/dB]	SRT ± σ [dB SNR]
Before Optimization	18.0	-9.1 ± 2.4
After digit selection	17.7	-9.1 ± 2.1
After intelligibility adjustment	17.7	-9.1 ± 0.3

The results of the digit-triplet test evaluation are shown in Table 2, where S is the mean slope of the psychometric function with its error σ and the SRT is the mean SRT with the corresponding error σ.

Table 2: Results of the ITA German digit-triplet test evaluation

	S ± σ [%/dB]	SRT ± σ [dB SNR]
Analysis by digits	12.8 ± 1.8	-12.2 ± 0.3
Analysis by triplets	18.8 ± 2.1	-11.9 ± 0.2

Test Conditions

Seven different cases were investigated in the listening-in-noise test. The target was always positioned in front of the listener (0° azimuth), but the distractor babble-noise can be in positions 0°, -90°, +90° or moving toward or away to the front. The worst case for intelligibility with strongest masking effect is when the sources are co-located, that is target and distractor in the same position (0° azimuth). The most advantageous case is when the target and distractor are in different spatial positions, that is target 0° and distractor ±90°. However, the masking effect is unknown when the distractor is moving.

Binaural Reproduction

The listening experiment took place in a sound attenuated listening booth at ITA which has a room volume of V ≈ 10.5 m^3 (l x w x h [m^3] = 2.3 x 2.3 x 2.0).

The stimuli were presented through headphones. For the binaural reproduction, head-related transfer functions (HRTFs) were used, rendering a sound source under free field conditions. The used HRTFset consists of ITA artificial head HRTF measurements with a 3° x 3° resolution in azimuth and elevation, respectively, measured with a loudspeaker mounted on an arm at a radius of 1.86 m [6].

Virtual acoustic scenes with static and moving sound sources were created using the Virtual Acoustic (VA) program, developed from the ITA.

Subjects

Twenty-eight young adult listeners (thirteen female) completed the listening experiment. All listeners were tested for normal-hearing of <20 dB hearing level between 125 and 8000 Hz. German was the native language for all subjects.

Procedure

Listeners were seated in front of a display and used a keyboard for data input. The used GUI and test routine was developed in Matlab/VA to play back the digits and the babble noise, to record and grade the response, to adjust the SNR after each trial is completed, and to store all the results.

The noise level was varied adaptively with a step size of 2 dB. After the first directional change, the variation of the step size was reduced to 1 dB. The starting level of the noise and the digit-triplets was set to 0 dB SNR.

The SRT was measured using all 14 digit-triplet lists in which the seven test cases were repeated for with and without head movements from listeners in the virtual scene. In this paper, only the results of the seven cases tested without allowing head movement in the virtual scene is reported.

The seven cases differ in the position and the movement of the distractor, with the target always positioned in front (0° azimuth). The cases are:

- Distractor is located statically in the front
- Distractor moves away from the front to the right
- Distractor moves away from the front to the left
- Distractor moves toward the front from the right
- Distractor moves toward the front from the left
- Distractor locates statically at the right
- Distractor locates statically at the left

Figure 1 provides a graphical illustration of all seven test cases.

All digit-triplet lists and the test conditions were presented to participants in a counterbalanced manner using a Latin Square.

The participants received the stimulus, which consisted of a list of 24 digit-triplets per case. After each digit-triplet, they must type in the digits in the same order as presented aurally using a numeric keypad. While keeping the babble-noise level constant at 70 dB, the target sound source level was changed adaptively using a one-up-one-down adaptive procedure to track the SRT at 50% speech intelligibility.

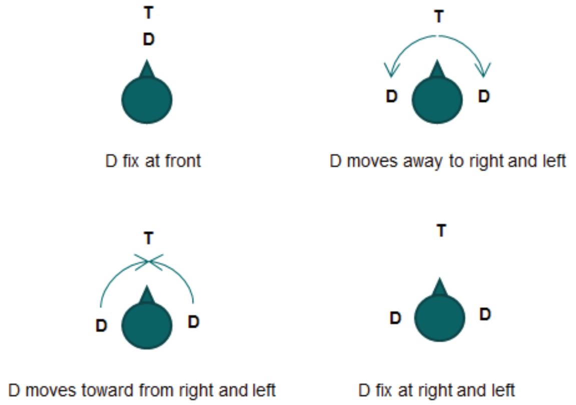

Figure 1: Graphical representation of all test cases. T denotes target sound source and D denotes distractor sound source.

Results

The mean SRT results from the 28 participants are shown in Figure 2. Release from masking is shown in Figure 3, which was calculated by subtracting the SRT from the reference “distractor front” condition from each of the remaining conditions.

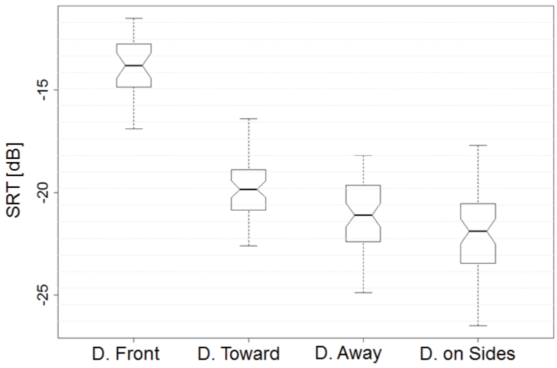

Figure 2: Comparison of mean SRT between cases.

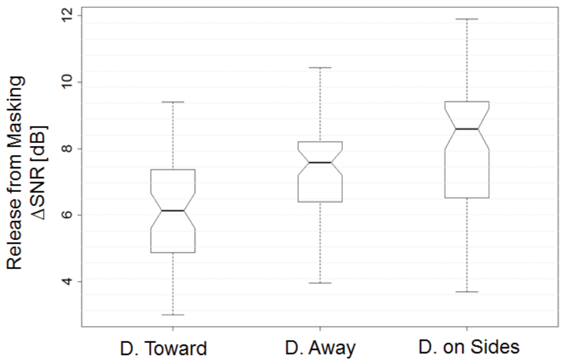

Figure 3: Comparison of release from masking in dB.

The data was analyzed in a repeated measures analysis of variance (ANOVA) and yielded significant main effects for distractor movement, $F(2,110) = 37.14$, $p<0.0001$.

The post-hoc analysis using pairwise comparisons are shown in Table 3. The Bonferroni adjustment was applied for multiple comparisons. The mean SRT was significantly different between all possible pairs of the test conditions, after averaging across ($p<0.05$).

Table 3: Significance value of pairwise t-test using Bonferroni adjustment

	Toward	Away
Away	p<0.0001	-
On Sides	p<0.0001	p=0.041

In addition, the percentage of correct answers between the first and the third digit was analyzed, in the cases of moving distractor towards and away, as shown in Figures 4 and 5. The data were analyzed using a paired t-test. In both cases, there exists a significant difference ($p<0.0001$) between the percentage of correct answers of the first and the third digit.

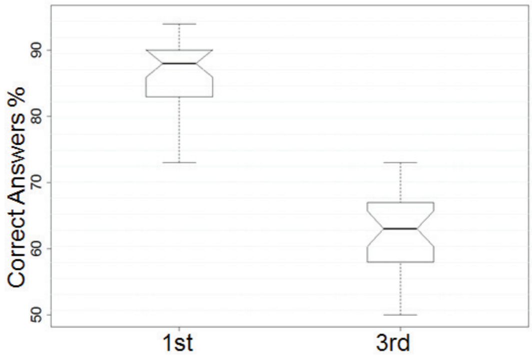

Figure 4: Distractor moving towards target. Percentage of correct answers for the first digit and the third.

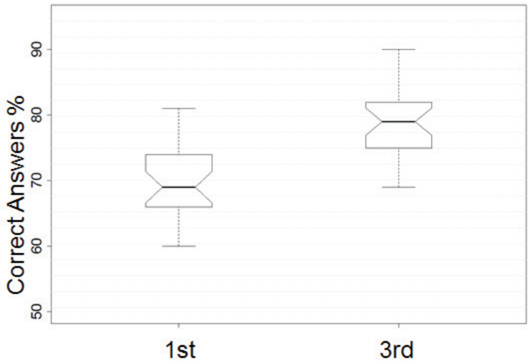

Figure 5: Distractor moving away from target. Percentage of correct answers for the first digit and the third.

Conclusions and Discussion

A German digit-triplet test with 14 parallel lists was developed following the guidelines of previous works. Masking noise that matches the long-term average speech spectrum of the digits was created. The digit-in-noise test measures the digit-triplet SRT using an adaptive procedure.

The main results of the listening experiment are:

The intelligibility (mean SRT values) across all the cases was found significantly different from each other.

Perceptually, the most difficult case is when the target and distractor are co-located.

If the target and distractor are separated by ±90° the SRT was lower, likely due to the increased differences in the interaural cues (e.g., ITD and ILD) from which the listeners can rely on in separating auditory streams of the target from the distractor.

Less release from masking was found for the dynamic trajectory with distractor moving toward than moving away from the target, likely due to a difference of masking energy between movements. Figure 6 shows the masking levels of the first digit (level M2, distractor in position ±90°) and the third digit (level M1, distractor in position 0°), when the distractor moves toward the target. . In Figure 7, the masking levels of the first digit (level M1, distractor in position 0°) and the third digit (level M2, distractor in position ±90°), when the distractor moves away. It is noticeable that the overall masking energy is higher when the distractor is moving toward the target than when moving away from the target.

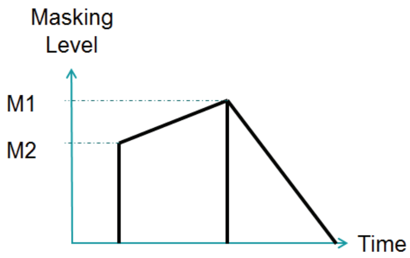

Figure 6: Distractor moving toward. Masking level of 1st and 3rd digit.

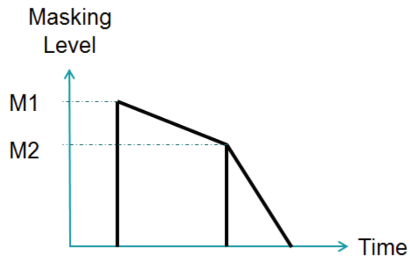

Figure 7: Distractor moving away. Masking level of 1st and 3rd digit.

It was demonstrated that the effect of a moving distractor in speech intelligibility test using digit-triplets test is significantly different between the different movements. Therefore, moving sound sources as well as their direction of movement with respect to a target source should be taken into account to design more realistic listening-in-noise tests.

References

[1] Akeroyd M.A., Arlinger S., Bentler R.A.: ICRA recommendations for the construction of multilingual speech tests. Int J Audiol (2015), 54(S2).

[2] Best, V., Ozmeral, E. J., Kopco, N., et al.: Object continuity enhances selective auditory attention. Proc Natl Acad Sci USA 105 (2008), 13174 –13178.

[3] Best, V., Shinn-Cunningham, B.G., Ozmeral, E.J., Kopco, N.: Exploring the benefit of auditory spatial continuity. J Acoust Soc Am 127 (2010), el258–el264.

[4] Carlile, S., & Best, V.: Discrimination of sound source velocity in human listeners. The Journal of the Acoustical Society of America 111(2) (2002), 1026-1035.

[5] Evans, M.J.: Synthesising moving sounds. Proceedings of the International Conference of Auditory Display ICAD'98, Glasgow, Scotland, 1-4 Nov. 1998, British Computer Society, UK.

[6] Pausch, F., Aspöck, L., & Fels, J.: Evaluation of binaural sound reproduction systems with focus on perceptual quality. Proceedings of Fortschritte der Akustik: 41 (2015). Deutsche Jahrestagung für Akustik, Nürnberg.

[7] Smits C, Theo Goverts S, Festen J.M.: The digits-in-noise test: assessing auditory speech recognition abilities in noise. J Acoust Soc Am 133 (2013), 1693–1706.

[8] Vasilenko, Y.A. and Shestopalova, L.B.: Moving sound source discrimination in humans: Mismatch negativity and psychophysics. Human Physiology 36(2) (2010), 139– 146.

[9] Weissgerber, T., Rader, T., Baumann, U.: Impact of a Moving Noise Masker on Speech Perception in Cochlear Implant Users. PLOS ONE (2015), vol. 10, no 5.

[10] Zokoll M.A., Wagener K.C., Brand T., Buschermöhle M. & Kollmeier B.: Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype. Int J Audiol 51 (2012), 697–707.