The benefit of head movements of normal listeners in a dynamic speech-in-noise task with virtual acoustics

Rhoddy Viveros, Janina Fels

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

Introduction

The head movements in a speech intelligibility task are studied since Kock (1950). He found that turning the head away from the speech source can provide benefits and was the first to map out thresholds of speech intelligibility in noise as a function of head orientation away from the speech (Grange and Culling, 2016). Even with the findings of Kock several studies proposed that for clinical test the listener should be placed in front of the speech source all the time, arguing that is a more natural listening attitude (Plomp, 1986; Bronkhorst and Plomp, 1990; Koehnke and Besing, 1996).

After more than 60 years since the findings of Kock, Brimijoin *et al.* (2012) were one of the first to measure head orientation during a speech-in-noise task. They evaluated the use of head orientation as a listening strategy for speech comprehension, only for asymmetric hearing impaired participants. The stimuli were presented from one of the loudspeakers arranged in a ring; a speech-shaped noise was used as a distractor source and an adaptive procedure was used to find the speech reception threshold (SRT). They reported a high variability in listener's head orientation and in the majority of the cases the orientation was different from the ideal.

Other similar study developed by Grange and Culling (2016) presented a comparative analysis of the predictive model of head orientation benefits (Jelfs *et al.*, 2011) and the spontaneous head orientation from listeners. The aim was to investigate if normal hearing listeners adopt an appropriate head orientation spontaneously. After the experiment a subset of participants was tested post-instructions, informing about the possible benefits of head orientation. As a result, only the 56% of the trial listeners spontaneously move the head more than 10° (a reference from short or long head movements) and in general, the participants did not make optimal use of their head orientation to improve the intelligibility. The predicted model in the asymmetrical cases revealed that the best head orientation is almost in between the two sound sources and the worse orientation is when the two sources are in the same cone of confusion.

Brimijoin *et al.* (2012) commented that listeners in the real world face an acoustic environment that rarely consists of a single target sound and a single, localizable distractor. 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. Our study compared between static and dynamic reproduction to investigate if listeners could use head movements to try to maximize their intelligibility during a dynamic acoustic scene with the distractor in movement. Were analyzed 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) and the spatial release from masking (SRM) which are the benefits of spatial separation of distractor from the target sound source.

In the current study, a headphone-based binaural audio reproduction system is used to present the speech target and distractor at six different scenes: target always fixes at 0◦ azimuth and distractor collocated with the target or moving away from 0◦ to 15◦, 30◦, 45◦, 60◦ and 90◦.

For the binaural reproduction, a set of head-related transfer functions (HRTF), measured from the ITA artificial head (Schmitz, 1995), 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

Speech stimuli were digit-triplets in German, consists of three monosyllabic digits (0-9, excluding 7) recorded by a female German native speaker (Viveros *et al.*, 2016).

The distractor 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 (Zokoll *et al.*, 2012). The complete length of each trial was 4 s (digit-triplet with babble-noise in the background).

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 HRTF-set consists of ITA artificial head HRTF measurements with a 1° resolution in azimuth and elevation.

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

Test Conditions

Six different cases were investigated in the speech-in-noise test. The target was always positioned in front of the listener (0° azimuth), but the babble-noise distractor can be in position 0° or moving away to the front 15°, 30°, 45°, 60° and 90°.

All cases were tested in two different conditions: static and dynamic reproduction. For dynamic reproduction in the virtual acoustic scenes, the binaural auralization was conducted in real-time and listener's head movement updating the HRTFs inside the virtual scene. An array of four optical cameras in the listening booth was utilized to track a sensor attached on a cap worn by the participant to track the head position. For static reproduction, the listener's head movements do not update the HRTFs in the virtual scene.

After the entire test, the participants have tested again but only in one condition post-instructions, informing about of possible benefits of head orientation. The only case was distractor moving away to the front 90°.

Subjects

Twenty-eight young adult listeners (thirteen female) completed the listening experiment divided into two different groups of fourteen. 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 4 dB. After the first directional change, the variation of the step size was reduced to 2 dB. The test terminated following six directional changes. The starting level of the noise and the digit-triplets was set to 0 dB SNR and the initial distractor noise level was played back at 70 dB SPL.

The speech reception threshold (SRT) was determined by averaging the SNRs of the last four reversals.

Each group of participants performed different cases: one group attended the cases with distractor fix in 0°, moving 15°, 30° and 45° (among other cases not reported in this study). The other group attended the cases with distractor fix in 0°, moving 30°, 60° and 90° (among other cases not reported).

Both groups together performed the six cases related to this study. Figure 1 provides a graphical illustration of all the cases.

The aim of this study is to compare between static and dynamic reproduction, exploring if listeners could potentially use head movements to try to maximize their intelligibility.

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 maximal 24 digit-triplets per case. After each digit-triplet, they must type 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.

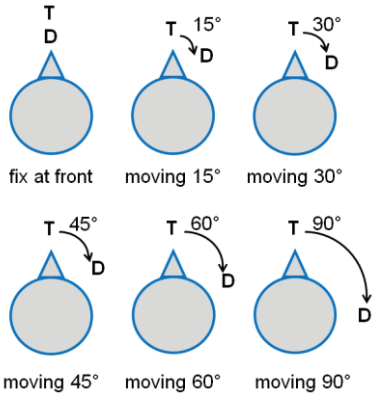

Figure 1: Graphical representation of all cases and the distractor location. T denotes target sound source and D is distractor source position or movement during the trial.

Results

The graphic in figure 2 shows the amount of spontaneous head movements larger than 5° (in blue). Is possible to observe that in the majority of the cases around the 50% of the listeners move the head away from the target position (0° azimuth) more than 5° and after the post-instructions (90° dyn.*) the amount of large head movements increase up to 75%.

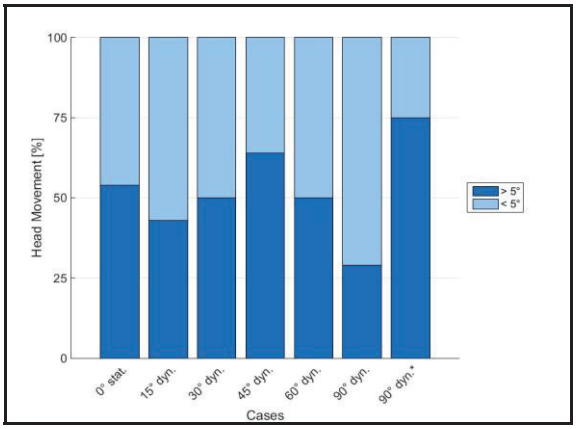

Figure 2: Graphical representation in percent of head movements across the cases. In blue the % of head movements larger than 5°.

The SRT data was analyzed in a repeated measures analysis of variance (ANOVA). The effect of listener head movement was found to be statistically not significant, $F(1, 13) = 0.4218$, $p > 0.05$, suggesting a not different SRT between dynamic and static binaural reproduction. The interaction is

shown in figure 3, which show the SRT between a static and dynamic condition for all distractor trajectories.

A similar two-way ANOVA was fitted to the SRM data. The effect of listener head movement was found to be statistically not significant, $F(1, 13) = 2.683$, $p > 0.05$, but the post-hoc analysis using a Bonferroni adjustment, show that the mean SRM was significantly different between static and dynamic reproduction ($p<0.05$). The interaction is presented in figure 4, which show the SRM between a static and dynamic condition for all distractor trajectories.

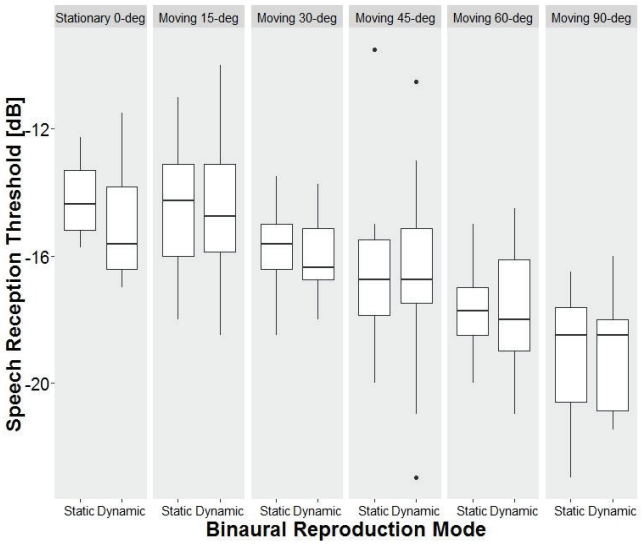

Figure 3: Speech reception thresholds at 50 % speech intelligibility measured for each distractor trajectory: (a) stationary at 0°, (b) moving away target 15°, (c) moving away target 30°, (d) moving away target 45° (e) moving away target 60°, and (f) moving away target 90°.In each distractor trajectory condition, the SRT was plotted separately for static and dynamic reproduction.

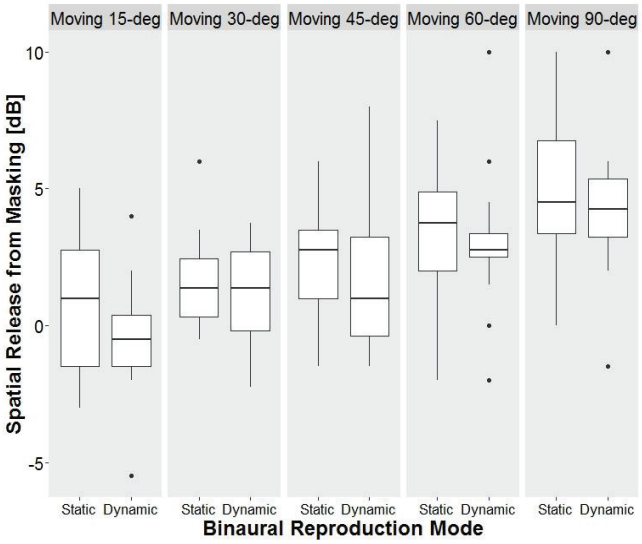

Figure 4: Spatial release from masking measured for each distractor trajectory: (a) moving away target 15°, (b) moving away from target 30°, (c) moving away target 45° (d) moving away from target 60°, and (e) moving away target 90°.In each distractor trajectory condition, the SRM was plotted separately for static and dynamic reproduction.

In addition, the t-test analysis post-instructions (only in the case distractor moving away to the front 90°) for SRT and SRM show a significant difference ($p<0.05$) between static and dynamic reproduction, meaning SRT for static reproduction is significantly lower than SRT for dynamic reproduction and SRM for static reproduction is significantly higher than SRM for dynamic reproduction.

Discussion and Conclusions

It is known that there are important benefits to speech intelligibility in noise available from orienting the head appropriately, but the results show that these benefits are not spontaneously used by the listeners.

A possible explanation is that the listeners do not have a natural strategy to confront this speech understanding in noise condition. In the majority of the cases around the 50% of the participants move the head more than 5° but using different strategies. Figure 5 show the head movements (orientation in yaw degrees, negative values means movements to the right) of a single listener during a specific case (distractor moving 90°), also are included the SNR values after each trial showing the adaptive procedure followed by the participant. In the first three trials, the listener moves the head like searching the best orientation, after that he/she decided oriented the head at 20°-22° to the right during the rest of the case. It is important to say that this strategy reported the best result in SRT across all participants under this case. Also important is to point that in this case the three digits target are played when the distractor is at around 22°, 45° and 68° respectively and the strategy to hold the head oriented at 20°-22° was the most effective because the two sound sources never are in the same hemisphere, avoiding the possibility that the two sound sources being in the same cone of confusion. Other participants with movements larger than 5° employed different strategies (see figure 6). In this case is possible to observe that the head orientation was to the right but also to the left, producing detrimental results as exposed by the predicted model in the asymmetrical cases of Grange and Culling (2016).

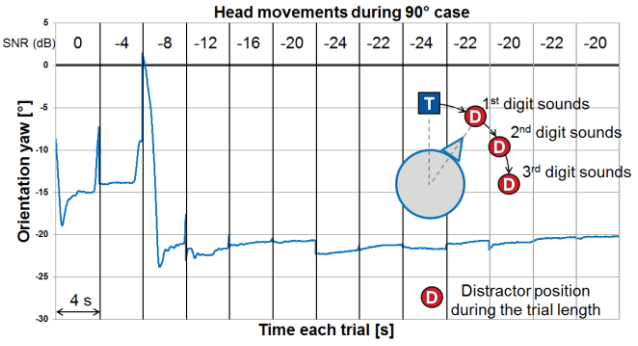

Figure 5: Head orientation of a single listener after each trial during the case distractor moving 90°. Best strategy reported in SRT under this condition.

Other explanation associated with the results obtained could be related to interference between cues. Bronkhorst and Plomp (1988) indicated that not only head shadow has a

considerable effect on free-field speech intelligibility in noise (basic case of a single stationary target and distractor), but also that the ILD caused by head shadow interfere with unmasking through ITD. Thus, in this study with distractor in movement (more complex acoustic scene) is possible to suppose that the interference could be higher, causing the poor results.

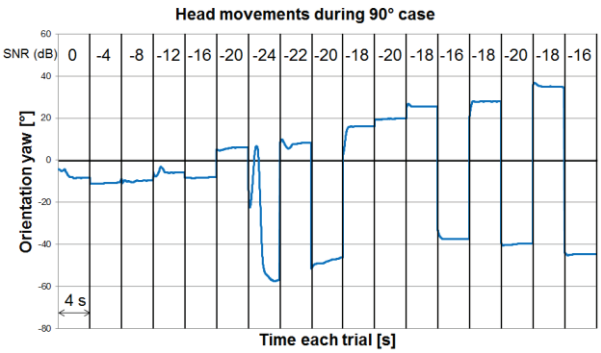

Figure 6: Head orientation of a single listener after each trial during the case distractor moving 90°. One of the worse results in SRT reported under this condition.

Despite the comparative results between static and dynamic reproduction in spontaneous and post-instructions head movements, it is suggested the use of dynamic reproduction for clinical tests, arguing that is a more accurate measurement of the intelligibility because is considering the irrational head movements of listeners in real complex situations (see figure 7).

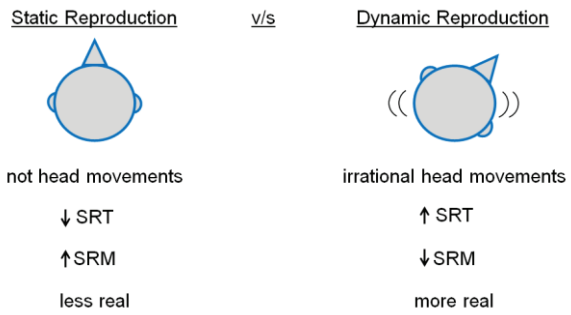

Figure 7: Summary of differences between static and dynamic reproduction.

The main results of the listening experiment are:

The intelligibility (mean SRT values) across all the cases was found not significantly different from static and dynamic reproduction.

After the post-hoc analysis using Bonferroni adjustment, the SRM across all the cases was found significantly different between static and dynamic reproduction.

After post-instructions, informing about of possible benefits of head orientation in case of distractor moving 90°, the 75% of listeners move the head more than 5°.

A t-test analysis of SRT and SRM shows a significant difference between static and dynamic reproduction in the case post-instructions.

$$SRT\ 90°_{dynamic}{}^{*} > SRT\ 90°_{static}{}^{*}$$

$$SRM\ 90°_{dynamic}{}^{*} < SRM\ 90°_{static}{}^{*}$$

References

Brimijoin, W., Mcshefferty, D., and Akeroyd, M. (**2012**). "Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task," Hear. Res. **280**, 162–168.

Bronkhorst, A., and Plomp, R. (**1990**). "A clinical test for the assessment of binaural speech perception in noise," Int. J. Audiol. **29**, 275–285.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J Acoust Soc Am. **83(4)**, 1508-1516.

Grange, J. A., and Culling, J. F. (**2016**). "The benefit of head orientation to speech intelligibility in noise," The Journal of the Acoustical Society of America **139**, 703-712.

Jelfs, S., Culling, J., and Lavandier, M. (**2011**). "Revision and validation of a binaural model for speech intelligibility in noise," Hear. Res. **275**, 96–104.

Kock, W. (**1950**). "Binaural localization and masking," J. Acoust. Soc. Am. **22**, 801-804.

Koehnke, J., and Besing, J. (**1996**). "A procedure for testing speech intelligibility in a virtual listening environment," Ear Hear. **17**, 211–217.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Schmitz, A. (**1995**). "A new digital art head measuring system," Acta Acustica, united with Acustica **81.4** 416-420.

Viveros, R., Peng, Z. E., Pausch, F., and Fels, J. (**2016**). "Effect of a moving distractor on speech intelligibility in babble noise using a digit-triplet test," Proceedings of Fortschritte der Akustik: 42 (2016). Deutsche Jahrestagung für Akustik, Aachen.

Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M., and Kollmeier, B. (**2012**). "Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype," International Journal of Audiology **51**, 697-707.