The effect of binaural processing techniques on speech quality ratings of assistive listening devices in different room acoustics conditions

Johan Odelius and Örjan Johansson

Luleå University of Technology, Dept Human Work Sciences, Div of Sound and Vibration, SE-97187 Luleå, Sweden
johan.odelius@ltu.se
1 Introduction

External microphone systems, referred to as assistive listening devices (ALD), are used to support communication in classrooms for hearing impaired students. The objective is to investigate different mixtures of sounds picked up by the ALD (T mode) and the hearing aid microphone (M mode). A listening experiment was conducted with 10 hearing impaired students. Response variables were self-assessments of speech intelligibility and preference. Design variable was mode combinations of T, M and M+T on right and left ear, respectively. Stimuli were generated using the room acoustic modelling software CATT Acoustic. Target source was continuous female speech in front of the listening position. Brownian noise and male speech were used as masking sounds. T mode corresponded to a recording from a microphone positioned in front of the target source and M mode corresponded to a binaural recording at the listening position. Stimuli were presented to the subjects wearing hearing aids by two loudspeakers utilizing cross-talk cancellation. Both M and T mode presented to one or the other ear, respectively, and M+T mode presented to both ears was significantly assessed better compared to only listening in M mode.

2 Methods

Speech intelligibility and preference was assessed in a listening experiment based on Multi Stimulus test with Hidden Reference and Anchor (MUSHRA) [13]. Stimuli were generated using the room acoustic modelling software CATT Acoustic [14].

2.1 Subjects

Ten hearing impaired students, eight girls and two boys, participated in the study. The age of the subjects ranged from 16 to 20 years. The participating students were fitted bilaterally and their unaided best ear pure-tone average (PTA, average of the hearing levels at 0.5, 1, and 2 kHz) ranged from 38-83 dB HL (median 70 dB HL). One subject had an interaural asymmetry equal or more than 15 dB.

2.2 Room acoustic model

A rectangular classroom 9.4x8.7x3.1 m was modelled in CATT Acoustics, see fig.1. Target speech source and listening position were placed on opposite sides of a table, at a distance of 1.5 m and 0 degrees azimuth. Two masker speech sources were positioned at either side behind the listening position. A masker noise source was also placed in one of the top corners in the room. The room was furnished and absorption and diffusion coefficients were set to create good room acoustic characteristics, see reverberation times in Table 1. Two receiver models were utilized in CATT;
one omnidirectional 0.2 m in front of the target speech source at -40 degrees elevation and one binaural at the listening position. ITA kunstkopf artificial head was used as binaural model [14]. The directivity of a singer [15] was applied to all speech sources.

![Male speech masker positions, Target female speech position, Noise masker position, Listening position](image)

Fig.1 Classroom model in CATT Acoustics.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1k</th>
<th>2k</th>
<th>4k</th>
</tr>
</thead>
<tbody>
<tr>
<td>T30 (s)</td>
<td>0.46</td>
<td>0.43</td>
<td>0.36</td>
<td>0.32</td>
<td>0.26</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 1 Reverberation time in the room acoustic model. T30 is derived from ray tracing using least-square fits to the decay in the interval -5 to -35 dB.

### 2.3 Stimuli

Anechoic recordings of female and male speech was used and convolved in CATT Acoustics at the target speech and masker speech positions, respectively. The target female speech was continuous sections between 12-15 s. For the male speech masker, random cuttings were used and randomly assign to one of the two masker speech positions. The equivalent sound pressure level of the male speech masker was set 3 dBA below the level of the female speech target at the listening position. Brownian noise (red noise) was convolved in the corner masker position and set 25 dBA below the female speech level. The M and T mode was equalized to the same A-weighted level. Further were both modes limited in frequency range, in octave bands, from 125 Hz to 4000 Hz. Eleven different mixtures of M and T mode were created, see Table 2.

### 2.4 Procedure

The listening experiment was carried out in a classroom at the students’ school. The room had a reverberation time of 0.6 s and a background noise level of 27 dBA. Cross-talk cancellation technique was used in order to enable dichotic listening. The two channel stimuli were presented, utilizing Lexicon MC-1, by two ADAM-S2A speakers in front of the listener. Subjects wore their hearing aids, set in M mode, during the experiment. The speakers were placed in the centre of the room to avoid early reflections, especially from the sides. Sufficient binaural reproduction was subjectively ensured.

The experiment was run in conjunction with another listening experiment evaluating different ALD solutions. The same user interface and stimuli were used in both studies. The order of the two experiments was randomized and one practice sessions introduced the subjects with user interface and type of stimuli. The sound pressure level of the stimuli in the experiment was subjectively set to normal speech levels. In the practise session the subjects had the opportunity to adjust the output level.

The stimuli was split up in two sessions, see Table 2. Stimulus 7 and 12 (M mode) are the same and was accordingly used as an anchor in both sessions. The sessions were run twice, with two different tasks (liberally translated from Swedish):

a) **Give 100 points to the sound where you most easily can hear what the female voice is saying. Then give the other sounds points. Lesser points the more difficult you think it gets to hear what the female voice is saying.**

b) **Give 100 points to the sound you like the most. Then give the other sounds points. Lesser points the worse you think the sound is.**

Each of the sounds had a slider where the score could be set from 0 to 100 points. The subjects could freely decide the order and how many times to play each sound. Each sound was a random selection of the female speech between 12-15 s. The two sessions with the two tasks, i.e. four runs, where carried out at random order.

### 3 Results

The analysis of the results will focus on speech intelligibility (SI), which was assessed with the ability to follow the female speech in the presence of a speech and a noise masker. An ANOVA of the difference between SI and preference showed no significant differences between stimuli. Fig. 2 and 3 shows the box plot for SI for the two sessions, respectively.
For further comparisons the data was normalised with regard to mean and standard deviation [13]:

\[ Z_i = \frac{(x_i - m_{si})}{s_{si}} + m_s \]

where:
- \( Z_i \): normalised result
- \( x_i \): score of subject \( i \)
- \( m_{si} \): mean score for subject \( i \) in session \( s \)
- \( m_s \): mean score of all subjects in session \( s \)
- \( s_s \): standard deviation for all subjects in session \( s \)
- \( s_{si} \): standard deviation for subject \( i \) in session \( s \).

A one-way ANOVA was performed for each session on the normalised scores, respectively. Significance level was chosen to 5%. Tukey HSD intervals, confidence level 95%, were used to derive which stimuli that differ from each other.

**Session I (stimulus 1-7):** The score means differed significantly between the different stimuli. Stimuli 1-2 and 5-6 received higher points than stimulus 7, see fig.4.

**Session II (stimulus 8-12):** The score means differed significantly between the different stimuli. The score increased with increasing quantity of T mode. Stimuli 8-10 scored higher points than stimulus 12, see fig.5. Further was stimulus 8 better than stimuli 10-11 and stimulus 9 was better than stimulus 11.

To compare session I with T mode (stimulus 8) and M+T mode (stimulus 10) in session II, a third ANOVA was performed on the difference to the anchor, i.e. M mode (stimulus 7 in session I and stimulus 14 in session II). Stimulus 8 received significantly higher points than stimuli 3-4, see fig.6.

A two-way ANOVA was also performed to see if there were any effect and interaction effect due to different degree of hearing loss. The subjects were categorised in two groups; PTA above and below 55 dB HL. No significant effects on hearing loss could be found.
4 Discussion

The results showed the advantage of an ALD system with regard to the audibility of a single talker in presence of masking sounds. A result that is in agreement with previous reported studies on both IL and FM systems. The binaural advantage in M mode when compared to T mode is lesser than the advantage of a short microphone distance. The compromise between good audibility and retrieving information from the surroundings remains. The results in this experiment do support that a combination of M+T mode is a feasible solution. Using one hearing aid in M mode and the other in T mode significantly increased speech intelligibility when compared to using both hearing aids in M mode, see fig. 4. Also the combination of M+T mode in one ear and T mode in the other ear was in advantage when compared to M mode. M+T mode in both ears was also rated better than M mode although it received lesser score than only listening in T mode, see fig. 5. Since T mode is adjusted on group basis, one can also discuss the advantage of an M+T mode due to the possibility of individual adjustment and the gained control over one’s own perception.

Multiple stimuli test is a powerful tool for the assessment of different ALD solutions. Self assessment of speech intelligibility was also chosen compared to a word recognition test in order to emphasis that it is the technology that is evaluated not the students’ hearing ability. It can be argued to what degree the scale from 0 to 100 is appropriate. As illustrated in the box plots for sessions I and II, fig. 2 and 3, the subjects’ usage of the scale varied. The data was consequently not normally distributed and the variance varied among the different factors. The suggested normalisation [13] is thus necessary. Non-parametric analysis could also be discussed since the method is a combination of ranking and magnitude assessments.

There were no difference in scores for speech intelligibility and preference. Hence, in this context the subjects preferred the stimulus with highest intelligibility. In future studies it would be preferable to also assess sound quality parameters. It has been shown that hearing aid users prefer aids with good speech quality rather than with high speech intelligibility [16].

Since the preferable mixture of T/FM and M mode are dependent on auditory scenario [7], further studies aims to better replicate classroom situations with multiple talkers. In addition, it would be preferable to assess the binaural advantage in localization and segregation tasks.

Acknowledgments

I would like to express my gratitude to the students participating in this study and to Håkan Bergkvist and Leila Sandmon for their support.

References