New pedagogical views on assistive listening devices have enlightened the technology’s impact on communication patterns in the classroom. Supporting a more dialogue-oriented and participatory learning environment, the attitude towards and the requirements of assistive listening devices [1].

The basic function of an assistive listening system is to enhance the speech in order to maintain acceptable speech intelligibility. According to this specific requirement, the sound quality of a listening device is more adequate termed sound-transmission quality. In a wider perspective, the output could also be analysed concerning the quality of the speech, where, according to Blauert and Jekosch [2], high speech quality is to which amount of the sound itself is suitable to match given expectations, demands and necessities. Regarding speech-transmission quality, Blauert and Jekosch also discusses perceptual, behaviour and psychological measures to sound transmission or reproducing systems. The sense of presence has been proposed measuring the amount of involvement.

The quality parameters discussed so far have focused on the speech as the signal of interest. Choosing an assistive technology, e.g. a teleconference system, with the purpose to improve the students’ possibilities to discuss and interact, there is a prominent risk that higher levels of ambient sound sources are picked up.

In a classroom, ambient sounds could be categorized into three types:

- **Activity noise**: Speech, footsteps, chairs, tables, doors, etc.
- **Installation noise**: Ventilation, computers and other equipment.
- **External noise**: Schoolyard, adjacent classrooms, traffic, etc.

According to Lundquist [3], activity noise is often considered the most annoying. Footsteps, chairs, and tables are examples of structure-borne noise, which, due to their temporal characteristic, may be a critical part of the sound climate in the classroom [3].

The most obvious aspect of ambient sounds is annoyance. Since sound is a carrier of information, another aspect could be the positive effect of receiving information about the acoustical environment and what happens in the surroundings. Analogous, directional microphones in hearing aids are sometimes discussed concerning communication isolation, i.e. the inability to detect off-axis acoustic signals could cause communication isolation [4]. The importance of auditory presence is also discussed regarding Virtual Reality technology [5]. Consequently, the feeling of presence concerning booth speech and ambient sounds is of interest judging upon the quality of a listening device. One must thus realize the reduced ability among individuals with a hearing impairment to locate and suppress competing sound sources.

Based on concepts behind sound quality adding a pedagogical interactive perspective, three dimensions are proposed defining communication quality in the context of assistive listening devices for students with a hearing impairment. The three dimensions are:
1) Speech intelligibility
2) Ambient sounds; a) Annoyance b) Feeling of presence
3) Parameters concerning the students’ performance given the aim and task of the pedagogical situation, e.g. participation.

One could argue about to which extent the proposed definition fulfill the features of speech quality. In the proposed definition, the requirement in focus is how the students recognize and comprehend the acoustical events in a classroom rather than judging upon the speech signal itself.

Speech intelligibility is commonly assessed as percentage correctly repeated phonemes, words or sentences. To capture the more natural situation, where a listener, in a noisy environment, tries to understand a spoken message, the just-follow-conversation (JFC) method has been utilized [6-8]. The test is performed by letting the listener adjust the level of speech or masker until he or she is just able to follow what is being said. This is perhaps how most listeners will react when trying to understand what is being said, e.g. by moving closer to listener [6]. Higher speech-to-noise ratio is needed in a JFC test compared to a 50-percentage speech recognition test, on average 10.5 dB [6] or 11.5 dB [8]. It should also be noticed that good correlation between the methods were achieved using speech as masking signal. The correlation was, however, poor when using filtered noise as a masker.

As concluded by Hygge et al. [7], subjects with a hearing impairment are equally affected by speech and noise backgrounds in a JFC test. Normal hearing subjects are, on the other hand, less affected by background speech than by noise. This is explained by the reduced temporal resolution for subjects with a hearing impairment. Moore [9] has summarized the outcome of several studies showing the amount to which the speech reception threshold (SRT) in background sounds is higher for subjects with a cochlear hearing loss. According to Moore, the difference in SRT is 6-12 dB and 2.5-7 dB, respectively, using speech and speech-shaped noise as background sound.

The objective of this study was to develop a method for evaluation of an assistive listening device regarding communication quality in a real environment. An experiment, where normal hearing subjects evaluated speech intelligibility and annoyance, was performed to investigate the effects and interaction effects of room acoustics, signal processing techniques, speech signals, and masking signals. Speech intelligibility was assessed using a JFC approach and ambient sounds were assessed concerning annoyance.

### Method

The methodical approach is shown in Figure 1. The speech and masking signals were filtered by an impulse response generated from a room acoustic model. The output signal was then generated using signal processing techniques, i.e. simulating the output of an assistive listening device. The computations were done in Matlab.

![Figure 1: The principal approach of the method.](image)

The experiment was based on a two level factorial design (2^5) divided into two blocks. Design variables were; (A) room acoustic properties, (B) microphone distance, (C) masking signal, (D) amplifier, and (E) microphone directivity. The variables are compiled in Table 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>-1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Room acoustic properties</td>
<td>Room 1</td>
<td>Room 2</td>
</tr>
<tr>
<td>B Microphone distance</td>
<td>0.75 m</td>
<td>1.5 m</td>
</tr>
<tr>
<td>C Masking signal</td>
<td>SSN^1</td>
<td>SBN^2</td>
</tr>
<tr>
<td>D Amplifier</td>
<td>Linear</td>
<td>Compression</td>
</tr>
<tr>
<td>E Microphone directivity</td>
<td>Omni</td>
<td>Array</td>
</tr>
</tbody>
</table>

^1Speech-spectrum random noise. ^2Structure-borne noise.

Response variables were; (I) adjusted masker level in the JFC-test, and (II) annoyance rating. The adjusted masker level was analysed as the A-weighted signal-to-noise ratio, SNR(A), of the input signals.

\[
\text{SNR}(A) = L_{\text{speech}}^{\text{Aeq}} - L_{\text{masker}}^{\text{Aeq}}
\] (1)

### Subjects

Ten normal hearing subjects, five women and five men, participated in the experiment. The age of the subjects ranged from 26 to 52 years (M=34 years).

### Stimuli

#### 2.2.1 Speech signal

The speech signal was a recording of a woman reading a continuous story from a fiction book. The equivalent A-weighted sound pressure level, \( L_{\text{eq}} \), was set to 55 dBA (\( L_{\text{eq}} = 62 \text{ dB} \)), which is the standard speech spectrum level for normal vocal effort [10].
2.2.2 Masking signal

*Speech-spectrum random noise (SSN):* Low-pass filtered white noise with a frequency spectrum approximating the long-term average spectrum of speech (-12 dB/octave with cut-off frequency at 1 kHz). In steps of 1 dB, twenty-one SSN signals were generated where LAeq varied from 48-68 dBA.

*Structure-borne noise (SBN):* A 0.25 seconds long recording of a metal cylinder striking a wooden table. The characteristic of the signal is shown in Figure 2. The impulse was repeated with frequency 2.5 Hz. As for SSN, twenty-one signals in steps of 1 dB were generated. LAeq varied from 64-84 dBA.

![Figure 2: The recordings of one beat of a metal cylinder at a wooden table plotted in time and its 1/3 octave spectrum.](image)

2.3 Room acoustic model

A rectangular classroom 7x8.5x3 m was modelled, with two acoustic properties, using the software CATT [11], see Figure 3. The reverberation times are defined in Table 2.

<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>Room 1</td>
<td>0.73</td>
<td>0.57</td>
<td>0.61</td>
<td>0.58</td>
<td>0.58</td>
<td>0.57</td>
</tr>
<tr>
<td>Room 2</td>
<td>0.35</td>
<td>0.58</td>
<td>0.60</td>
<td>0.50</td>
<td>0.46</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Table 2: Reverberation time T₆₀ (s).

![Figure 3: A schematic picture of the room acoustic model.](image)

The revised speech-transmission-index (STI) [13] was 66.7 in room 1 and 68.0 in room 2. The signal source was placed 1 m from the microphone at a height of 1.2 m. Increasing the microphone distance from 0.75 m to 1.5 m, the SNR(A) was decreased 4.3 dB.

2.4 Amplifier

A linear amplifier and an amplifier utilizing compression were assessed. The compression parameters were chosen as shown in Table 5.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Ratio</th>
<th>Attack time</th>
<th>Release time</th>
</tr>
</thead>
<tbody>
<tr>
<td>65 dB</td>
<td>5:1</td>
<td>10 ms</td>
<td>100 ms</td>
</tr>
</tbody>
</table>

Table 5: Compression setup.

2.5 Microphone directivity

The microphone at the height of 1 m was used as the omni directional alternative. The microphone array output was processed using a delay-and-sum beamformer. The beamformer was steered to ϕ′ = 12°, corresponding to a source at height 1.2 m and radius 1 m. The beamformer was implemented applying time delays, which for the nth microphone is given as

$$\tau_n = \frac{(n-1)d \sin \phi'}{c} \quad n = 1, \ldots, 5 \quad c = 343 \text{ m/s}$$

To avoid spatial aliasing the sensor spacing must be less than half the wavelength. Consequently, spatial
aliasing occurred for frequencies above 1700 Hz. The directivity index of the array, computed at a distance of 1 m, is shown in Table 6.

Table 6: Directivity index of the array by octave bands.

<table>
<thead>
<tr>
<th>Octave (Hz)</th>
<th>DI (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>0.1</td>
</tr>
<tr>
<td>250</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>3.1</td>
</tr>
<tr>
<td>1k</td>
<td>6.2</td>
</tr>
<tr>
<td>2k</td>
<td>8.3</td>
</tr>
<tr>
<td>4k</td>
<td>5.6</td>
</tr>
</tbody>
</table>

2.6 Procedure

The listening experiment was carried out in an anechoic room. Sound stimuli were presented using headphones. Dividing the 2^5 design into two blocks, each subject performed 16 test runs randomly arranged in each block. All subjects performed a training session of two runs. The stimuli used in the training session were randomly picked from the 16 test runs, where one stimulus contained a SSN masker and one contained a SBN masker. In the JFC task the subjects were instructed to adjust the level of the background sound until they, with full concentration, were able to follow what was being said. Using a computer interface the subjects were able to adjust the masking signal in 1 dB or 3 dB steps within the 20 dB dynamic range of each masker. All stimuli were 30 seconds long and the adjusted value was registered at the end of each stimulus. In conjunction with each JFC task, the subjects were asked to estimate the annoyance of the masking signal on an eleven-point scale. The experiment could be illustrated as in Figure 4.

![Figure 4. A schematic picture of the experimental procedure. A-E denotes the five design variables.](image)

The following instructions were given to the subjects: “You will be listening to a women reading from a book in 16 sequences of each 30 seconds. The female voice will be presented together with background sounds. The background sounds are either noise or a “knocking”-sound. Your task is to listen to the voice and adjust the level of the background sound to a level where you with full concentration can follow what is being said. Your adjusted value will be registered when 30 seconds have passed. When you are finished adjusting the level your task is to rate the annoyance of the background sound.”

3 Result

3.1 Just-follow-conversation

The group average of adjusted SNR(A), Equation (1), for the different masking sounds are shown in Figure 5.

![Figure 5. Adjusted SNR(A) for the SSN and SBN masker.](image)

Table 7. The definitions of the test conditions.

<table>
<thead>
<tr>
<th></th>
<th>Room 1</th>
<th>0.75 m</th>
<th>Linear</th>
<th>Omni</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Room 2</td>
<td>1.5 m</td>
<td>Linear</td>
<td>Omni</td>
</tr>
<tr>
<td>2</td>
<td>Room 2</td>
<td>0.75 m</td>
<td>Compression</td>
<td>Omni</td>
</tr>
<tr>
<td>3</td>
<td>Room 1</td>
<td>1.5 m</td>
<td>Compression</td>
<td>Omni</td>
</tr>
<tr>
<td>4</td>
<td>Room 2</td>
<td>0.75 m</td>
<td>Linear</td>
<td>Array</td>
</tr>
<tr>
<td>5</td>
<td>Room 1</td>
<td>1.5 m</td>
<td>Linear</td>
<td>Array</td>
</tr>
<tr>
<td>6</td>
<td>Room 1</td>
<td>0.75 m</td>
<td>Compression</td>
<td>Array</td>
</tr>
<tr>
<td>7</td>
<td>Room 2</td>
<td>1.5 m</td>
<td>Compression</td>
<td>Array</td>
</tr>
<tr>
<td>8</td>
<td>Room 2</td>
<td>0.75 m</td>
<td>Linear</td>
<td>Omni</td>
</tr>
<tr>
<td>9</td>
<td>Room 1</td>
<td>1.5 m</td>
<td>Linear</td>
<td>Omni</td>
</tr>
<tr>
<td>10</td>
<td>Room 1</td>
<td>0.75 m</td>
<td>Compression</td>
<td>Omni</td>
</tr>
<tr>
<td>11</td>
<td>Room 2</td>
<td>1.5 m</td>
<td>Compression</td>
<td>Omni</td>
</tr>
<tr>
<td>12</td>
<td>Room 1</td>
<td>0.75 m</td>
<td>Linear</td>
<td>Array</td>
</tr>
<tr>
<td>13</td>
<td>Room 2</td>
<td>1.5 m</td>
<td>Linear</td>
<td>Array</td>
</tr>
<tr>
<td>14</td>
<td>Room 2</td>
<td>0.75 m</td>
<td>Compression</td>
<td>Array</td>
</tr>
<tr>
<td>15</td>
<td>Room 1</td>
<td>1.5 m</td>
<td>Compression</td>
<td>Array</td>
</tr>
</tbody>
</table>

The average of the adjusted SNR(A) was -11.3 dBA (Range: -29 to 7 dBA). The average of SNR(A) for the SSN masker and the SBN masker was -1.5 dBA (Range: -12 to 7 dBA) and -21.0 dBA (Range: -29 to -9 dBA), respectively.

Analysing the estimated effects of the design variables, microphone distance (B) and masker (C) were the only factors with a significant effect (p<0.05). There were no significant interaction effects or a block effect. The effects are presented in Figure 6. The estimated effect
of factor B was 3.3 dBA and regarding factor C the estimated effect was -19.5 dBA. The significant effects imply that when the microphone distance was increased from 0.75 m to 1.5 m the listener needed 3.3 dB better SNR(A) in order to follow the conversation. Further, the listener needed 19.5 dB lower SNR(A) for the SBN masker compared to the SSN masker.

Further, the listener needed 19.5 dB lower SNR(A) for the SBN masker compared to the SSN masker.

### 3.2 Annoyance

The averaged annoyance was 7.2, measured on the eleven-point scale. The only significant effect (p<0.05) was type of masker (C), where the estimated effect was 2.4. Thus, the subjects estimated the annoyance 2.4 point higher for the SBN masker compared to the SSN masker. The estimated effects are shown in Figure 7.

![Pareto Chart for Annoyance](image)

**Figure 7.** Estimated effects regarding the annoyance ratings.

### 4 Discussion

In accordance with the 4.3 dB deterioration in SNR(A), the listener needed 3.3 dB higher SNR(A) to comprehend what was being said when the microphone distance was increased to 1.5 m. The importance of microphone distance imply that despite the big potential in different signal processing techniques microphone distance is essential for speech intelligibility concerning assistive listening devices. However, since there was no significant effect applying the microphone array, one could conclude that the microphone array utilized in the experiment was not well suited for this application. The result could be explained by the frequency dependent gain and spatial aliasing for high frequencies. Figure 5 indicates that the array performed better at the microphone distance of 0.75 m compared to 1.5 m. The impact of compression to speech intelligibility is disputed, see e.g. [9]. In this experiment it could be concluded that the compression had no impact on speech intelligibility. Since compression is used to match the restricted dynamic range of the listener, the effect of compression is more adequate evaluated using subjects with a sensorineural hearing loss. It could also be noticed that compression, due to the slow performance, presumable had a modest impact on the SBN masker.

The small differences in room acoustic properties did not result in a significant effect. Studies [14, 15] have argued that the most important factor to room acoustic design is background noise. This implies that the variations in room acoustics must be more prominent to produce effects in comparison to the effect of ambient noise level.

The 19.5 dBA higher allowed noise level for the SBN masker could be explained by the difference in its temporal characteristics compared to the SSN masker. The result is in agreement with previous studies [7], where normal hearing subjects performed better when the masker was speech compared to a speech-spectrum noise masker. Previous studies also conclude that subjects with a hearing impairment are equally affected by speech and noise maskers. Consequently, the difference between SBN and SSN will probably be smaller for subjects with a hearing impairment.

The SBN masker was rated 2.4 point more annoying than the SSN masker. The correlation between the level of the adjusted masker and the annoyance ratings was significant (r=0.62). The higher rated annoyance due to the SBN masker is thus partly a consequence of the higher levels adjusted in the JFC test. However, the increased annoyance ratings could be considered small in relation to the 19.5 dBA difference in adjusted masker level.

Since the listeners only could adjust the level within the 20 dB dynamic range, one must consider the number of times when the masker level was adjusted to the maximum or minimum level. The maximum masker level was used 19 times (12%) and the minimum masker level was used 24 times (15%). All the maximum masker levels and 14 of the minimum masker levels occurred when the SBN masker was used. There was no such skewed distribution concerning the other design variables. Consequently, there might be effects not captured in the experiment, especially concerning the SBN masker.

In addition to the concluded importance of microphone distance and type of masker, the experiment has shown the potential of the proposed method. Increasing the possibility to find significant effects, some
improvements regarding the experimental design must be considered. Besides increasing the number of subjects, it would be preferable not to have such a big difference in characteristics of the masking signals. A better structure-borne sound is thus needed. It is also important to have a wider adjustable dynamic range. Another important aspect is to find a better response variable that captures the features of the sound presented to the listener.

To fully evaluate the different effects, it would be interesting to present the processed sound to subjects with a hearing impairment using an induction loop system and the hearing aid induction pick-up coil. Simultaneously the sounds would be reproduced using loudspeakers. Accordingly, the subjects would judge upon the quality of the output of the device in the actual acoustical environment. This approach will also more easily enable the subjects to rate additional dimensions to communication quality.

References


