What Audio Quality Attributes Affect the Viewers’ Preference, Comparing Overhead and Underneath Boom Microphone Techniques

Simon J. Nyqvist

Audio Technology, bachelor's level
2018

Luleå University of Technology
Department of Arts, Communication and Education
Abstract

The boom microphone has been used to record dialogue on a film set for decades. As a boom operator it can be difficult in certain situations to acquire an ideal placement for the boom microphone. It depends on the frame and the lighting on the set. There are situations where the boom operator is forced to place the microphone underneath the actors instead of the standard overhead position. To investigate the difference between these techniques, and what audio quality attributes that affect the viewers’ choice of preference, a listening test has been constructed. The subjects were instructed to choose preferences in four trials of stimuli that had been recorded with both microphone techniques. To collect further data the subjects were also instructed to describe the differences they perceived between the stimuli. A statistically significant result was not found, except for one of the trials, the trained listeners preferred the underneath microphone technique for the second female stimuli. A difference between the techniques was difficult to determine from the answers given by the subjects. The conclusion was made that there is not a clear preference between the microphone techniques. Even though one of the trials was significant a clear difference could not be found in the analysis of the qualitative data.
Contents
1. Introduction .................................................................................................................. 3
  1.1. Background ............................................................................................................. 3
    1.1.1. The boom microphone .................................................................................... 3
    1.1.2. The professional view .................................................................................... 4
    1.1.3. Audio quality, earlier research and dialogue evaluation ................................. 5
  1.2. Purpose .................................................................................................................... 7
2. Method .......................................................................................................................... 9
  2.1. Recording and creation of stimuli .......................................................................... 9
    2.1.1. Recording environment and recording equipment ........................................... 9
    2.1.2. The scene and the actors ................................................................................ 9
    2.1.3. Post-production of the film clips ..................................................................... 10
    2.1.4. Dialogue editing and exportation .................................................................... 11
  2.2. Pre study ................................................................................................................ 12
  2.3. Listening test ......................................................................................................... 12
    2.3.1. Listening environment ................................................................................... 12
    2.3.2. Subjects and test procedure .......................................................................... 13
3. Results and analysis ................................................................................................... 15
  3.1. The A/B-test ......................................................................................................... 15
  3.2. The answers from the open question .................................................................... 18
    3.2.1. Attributes in general ...................................................................................... 18
    3.2.2. Other factors in choice of preference .............................................................. 20
  3.3. The answers from the statistically significant trial ............................................... 20
  3.4. Comparison between the overhead and underneath microphone technique ........ 21
4. Discussion ..................................................................................................................... 24
  4.1. The method .......................................................................................................... 24
  4.2. The results ............................................................................................................. 24
  4.3. Conclusion ............................................................................................................. 25
  4.4. Further research ................................................................................................... 25
Acknowledgements ........................................................................................................ 26
5. References .................................................................................................................... 27
Appendix A: The questionnaire for the listening test .................................................... 28
Appendix B: The playback order for the subjects ........................................................... 29
Appendix C: Table of Cumulative Binomial Probability ................................................ 30
1. Introduction

1.1. Background

Dialogue in film has a direct storytelling role. Human voices are usually the most significant elements in a film’s sound track (Goussios, Sevastiadis & Kalliris, 2007). This means that the dialogue is a crucial part of a film’s sound track. If it is poorly captured, or poorly post-processed, the story of the film will be lost for the viewer.

Dialogues are recorded on set by the sound crew that typically consists of a Production Sound Recordist and a Boom Operator. These productions sound recordings are used in the post production for the film’s sound track (Goussios et al. 2007). The limitations on a film set when it comes to capturing dialog can be different. Goussios et al. (2007) writes: “Several recording techniques and equipment are used in outdoor and indoor recordings for Motion Picture. The choices are usually characterized from subjectivity and technical limitations irrelevant to the desired final sound quality.”.

As a boom operator it’s sometimes impossible to get close enough with the microphone to capture high quality dialogue recordings. Lavaliere microphones are also used but these doesn’t provide the same audio quality as a boom microphone. It can also be difficult for the boom operator to place the microphone in the right angle to the speaker, and sometimes it’s necessary to place the microphone underneath the speaker instead of overhead. These factors will eventually affect the audio quality of the dialogue in the final sound track. Other common problems on a film set are noisy environments such as weather (wind, rain), traffic, people etc. Some of these problems can be solved on set, but that is not the case for all of these. Which means that the final sound track will be affected.

1.1.1. The boom microphone

Although the boom microphone may be old fashioned, they usually work best for dialogue recording. The ideal position is on a boom above the frameline and in the center of the frame. The reason for this is that the perspective matches that of the camera. If an actors happens to turn their head and face out of the frame the change heard when going from “on mic” to “off mic” is natural. The on-mic position sounds clearer and less reverberant, the off mic sound duller and more reverberant (Holman, 2013).

A boom microphone is a directional condenser microphone placed on a boom pole. The microphone is often referred to as a “shotgun microphone”. These microphones are used because the design and the length of the interference tube lead to better discrimination of the microphone; and usually, the more discriminating it becomes in rejecting off-axis sound. These characteristics make directional microphones preferable for film applications and especially for recording dialogue on set. (Goussios et al., 2007).
When using a boom microphone, on an outdoor set, the two biggest problems are wind and vibration. An effective way to control these problems is the use of a pistol-grip shock-mount with a basket wind screen (referred to as a zeppelin). Since wind susceptibility is relatively high for directional microphones it is mandatory to use a wind screen. Additionally a hairy coat windscreen can be used. It is also recommended to use a foam windscreen. Since fast movement and panning of the boom microphone can lead to unwanted noise due to air passing through the microphones interference tube’s slots (Goussios et al., 2007).

The responsibility of the boom operator is placement of the boom microphone. The use of a boom pole allows control of the microphone at a distance away from the actors (Goussios et al., 2007). A challenge for boom operator is the dimensions of the frames. It’s crucial as a boom operator to know the exact dimensions of the frames, in order to avoid placing the microphone or boom pole in the frame when shooting. It is also important to be aware of lighting on set, in order to keep the shadow of the boom microphone out of the frame (Goussios et al., 2007).

According to Goussios et al. (2007) there is a low and mid frequency increase when using the underneath technique that can be due to resonances of the human body. While using the overhead technique the voice is recorded with a richer timbre.

Holman (2013) states that a microphone centered underneath the frameline can be considered as a fallback position, although it is not as desirable. The reason for this is that it often sounds “chesty”, thus emphasizing midbass, and less clear than above. In this situation the microphone has to be closer to the floor meaning that it will receive reflections off the floor that can color the sound.

1.1.2. The professional view

Tony Österholm, a professional boom operator, sound recordist, supervising sound designer and film mixer, with 10 years of experience, was asked, by the author, about the differences between the overhead and the underneath microphone techniques. The purpose of this was to get an understanding of what the professional view is, regarding this question.

The main point brought up is that it is a matter of a sound. An example given regarding the overhead technique was that it has been used for decades in film dialogue recording. Meaning that the preferences for this microphone technique, and the dialogue recorded with this technique, has been developed through the years. Another example given, that doesn’t regard the dialogue sound, was the creation and recording of footsteps on grass. In real life, walking on grass does not sound much. In film sound grass footsteps are created by walking on old recording tape, which makes the recognizable
sound of footsteps on grass. This method has also been used for decades which means that the viewers are used to this sound (personal communication, February 16, 2018).

Another factor regarding overhead and underneath microphone technique is the characteristics of the sound, the timbre. In dialogue editing all the dialogue is eq matched for the sake of consistency. It is possible to eq match dialogue recorded with the underneath technique to dialogue recorded with the overhead technique, the problem here is that the different techniques will have different timbres. This means that, if the boom operator starts with the use of the underneath technique in the beginning of a scene, it is necessary to consider to keep using the underneath technique throughout the whole scene even though it would be possible to switch to overhead for the next frame (personal communication, February 16, 2018).

The overhead techniques is regarded as the standard technique from a professional view point. The reason for this is that it is what the viewers are most used to, because the overhead technique has been used for a long time. It is also the most versitale because it leaves space for the actors to do their job. It the underneath technique is used it is more likely that the boom pole will be in the way, thus preventing actors from performing certain movements that may be necessary for the scene (personal communication, February 16, 2018).

1.1.3. Audio quality, earlier research and dialogue evaluation

Blauert & Jekosch (2012) states the following: “Sound quality is a complex and multilayered phenomenon. When analyzing or modeling the formation process of sound-quality judgments, a variety of quality elements and quality features have to be taken into account, whereby the actual relevance and salience of each of the them is situation dependent.”. When evaluating the sound quality of dialogue there’s a lot to take in to consideration. What kind of attributes/qualities should be assessed and how should these attributes/qualities be evaluated. Ehn (2016) created a listening test where subjects were instructed to evaluate three different qualities in film dialogue.

In recent year there has been an increase in the number of complaints concerning the dialogue intelligibility of both movie and TV sound tracks. Leading to networks (e.g. BBC) having to publicly defend their productions (Mapp, 2016).

There are different ways the audio quality of the dialogue can be evaluated. In an experiment done by Goussios et al. (2007) different test signals were recorded with different types of microphone techniques, and microphones accessories, used for capturing dialogue sound. The authors wanted to present results of comparative recordings to give answers to every-day-practice problems. The test
signals recorded consisted of white noise, frequency sweeps and human voice. The recordings were done both outdoors and indoors in a professional studio, with different microphones techniques and different types of accessories. These recordings were analyzed by two different frequency analysis software applications thus giving a technical result of the different recordings. The test results are presented in graphs that compares frequency curves of the different recordings. Giving a clear picture of how the different techniques and accessories are affecting the recorded test signals.

In the listening test created by Ehn (2016) the test subjects were asked to rate three different qualities in film dialogue, in two different listening positions. The positions were defined as a sweet spot, the ideal listening position, and an off-center listening spot defined as the less ideal listening position. The purpose of this experiment was to investigate if untrained listeners could perceive a difference in audio quality when switching between the sweet spot and the off-center position. The qualities that the subjects rated were depth, stability and clarity. These three qualities were chosen because they were deemed non-complex enough for untrained listeners to listen for and to rate. Prior to the test the qualities were explained and defined to the test subjects to make it clear what to listen for during the test. This test did not show a statistically significant results.

Audio is very multidimensional both in nature and perception. It is possible to measure the physical characteristics of an audio signal in the acoustic or electrical domain. However, this characterization does not tell how the human auditory system will interpret and quantify it. An alternative to assess how listeners perceive and audio signal is to ask the listeners to quantify the experience. In other words, listening tests are the most common form of perceptual evaluation of audio (Bech & Zacharov, 2006).

Bech & Zacharov (2006) also present a list of attributes that can be used when evaluating audio quality. A few examples of these are: Tense/Sharp, Dark-bright and mechanic (Speech quality). Sense of depth, Sense of space and Sense of movement (Spatial sound quality in loudspeakers). Sense of distance, Sense of direction and Sense of movement (Spatial sound quality in headphones).

Another important factor in the evaluation of dialogue is the viewers of the movies or the television shows. These viewers are evaluating the audio quality at home. Ruddick (2017) reports that the creators of the TV series Blue Plant 2 used a normal television instead of a movie theatre or a studio when mixing the final sound track for the show. This because of concerns that viewers would complain about the narration not being audible. The BBC has faced an increase of complaints about sound in its programs, were viewers are saying that the dialogue cannot be heard due to loud music (Ruddick, 2017).
One reason for the poor intelligibility according to the program producers is the poor set up or quality of the viewers TV sets. Because of this, a test were acoustical measurements and frequency response of television loudspeakers were measured to investigate if this could be the cause of the significant increase in complaints about the intelligibility (Mapp, 2016).

Mapp (2016) came to the conclusion that the viewers living rooms provide an acoustically good environment for speech intelligibility. Different modes or setups of the televisions showed that some settings could potentially improve the speech intelligibility by attenuating lower mid peaks and accentuating the consonant region. However, a problem was found in the frequency response. Some television manufacturers place the speakers below the screen pointing downwards. This showed a strong rise in comb filtering.

The viewers have also asked why the problems with speech intelligibility are not picked up and corrected at the production stage. It is however likely that the loudspeakers and the acoustic responses that the producers use would not expose the degradations that the measurements have highlighted. Another factor is that familiarity with the material can have a significant effect on perceived intelligibility, which cannot be compensated for (Mapp, 2016).

1.2. Purpose

As a sound recordist on set it’s important to record the dialogue as good and clear as possible. A well recorded dialogue with good quality is easier to work with in the dialogue editing in post-production. An ideal microphone placement for the boom operator is as close to the actors as possible. This is easier to do in close up shots. One thing to have in mind when booming at a close distance is to keep the same distance to the actor all the time, also when panning the microphone to another actor. This is due to the inverse square law. Goussios et al. (2007) writes: “This means that if desired ‘close’ position can be achieved, it has to be the same for all the recordings of the takes of the scene and the same after all the movement and panning while recording different actors. The closer the distance the easier it becomes for the level to fluctuate and experience deeps and peaks in the sound pressure level.” If it’s hard for the boom operator to achieve this it’s important for the set sound recordist to keep the input levels balanced.

The purpose of this study is to investigate the difference between the overhead and the underneath boom microphone techniques, and what audio quality attributes that affect the viewers’ choice of preference. When filming on set, the choice of using overhead or underneath boom is often decided by the frame rather than because of audio quality consideration. The reason for this is that in some situation it is not possible for the boom operator to use the standard overhead technique. Reasons for
this being that the boom microphone may be visible in the shot, or the more common one that the shadow from the boom microphone is visible in the shot, thus forcing the boom operator to use the underneath technique. This study could be of help for boom operators and sound recordist regarding the choice of whether to use one or the other technique.
2. Method

To investigate if viewers can perceive a difference between the microphone techniques it was decided to use a preference test. The subjects would listen to pairs of stimuli one with the overhead technique and the other with the underneath technique and pick a preference between these two followed up by a question why the choose one over the other.

2.1. Recording and creation of stimuli

The stimuli created for the study was two different film clips, both film clips being about 1:30 minutes long. These film clips were edited and cut down to four shorter clips between 18 and 25 seconds long. These four shorter film clips were the ones later used in a pre study and a listening test.

2.1.1. Recording environment and recording equipment

The film clips were recorded in an interior at LTU campus Piteå in a medium big room, approximately 7 by 8 meters, with a ceiling height of 3 meters. This environment was chosen because of its neutral state. A neutral environment was chosen because the possibilities of creating a sound design that would not be dependent of the environment.

When filming the clips it was necessary to do a close up frame. This made it possible to keep a microphone distance for both overhead and underneath boom microphone technique at 40 cm at an angle of 45 degrees. Without the microphones being visible in the shot. Both overhead and underneath microphone technique were recorded with these premises, simultaneously with one male actor and one female actor.

Recording equipment:

Camera: Panasonic Lumix G7
Recorder: Sound Devices 702T
Audio format: PCM Wave 24 bit of depth and 48 kHz sample rate
Microphones: Neumann KM184 + pop filter

2.1.2. The scene and the actors

In the scene the actors were put in an interview situation. The interviewer asked the actors what the most important characteristics are in their field of profession. The actors answered the question in a neutral voice on a medium level with no change in dynamics. In other words no actual acting or dramatic elements were inserted in to the direction of the film clips.
As mentioned earlier two different recordings were made, one with a male actor and the second with a female actor. The male actor had a tenor voice and the female actor had an alto voice. Meaning that the male actor’s voice did not contain much low frequency content, although compared to the female voice, which was not that high in pitch. The male voice contained significantly more low frequency content.

2.1.3. Post-production of the film clips

To make the film clips more ecologically valid a sound design was made. The reason for this is that in a real film, the dialogue track is also accompanied by other sound elements. These elements being: ambient sounds, sound effects and music. Meaning that only a dry dialogue track wouldn’t represent a real film situation.

The sound design created for the film clips was an office like environment. Because of the filming locations’ neutral environment this was possible, however even though the environment was neutral it can easily be seen that it is not a crowded or a stressful environment were a lot of things is happening. Because of this the sound design made was a calm and neutral interior environment. The sound design contained the following elements:

Ambient sounds
For the ambient sounds a room tone track and a valla track was used. A room tone is a neutral tone that exists in most rooms, this particular one contained neutral static noise and a sound of a quiet fan. A valla track is a soundtrack that contains human voice. In this particular track voices can be heard as well as footsteps in a calm environment

Sound effects
Since there is nothing else happening in the shot expect from the actors talking. The sound effects are off-screen effects. The purpose of these is to accompany the ambient sound track to make a more living sound environment. The off-screen effects for this sounds are footsteps, a door that is being open and shut as well as a telephone ringing

Music
To further enhance the sound design a music track were added. The music track in the film clips was heavily processed to give it the sound of a small radio that was playing back the music in the room
2.1.4. Dialogue editing and exportation

The only editing done to the dialogue track was a high pass filter at 80 Hz. The reason for this was to get rid of potential low frequency rumble in the recording. Other than that the level of the dialogue as well as the other elements were perceptually matched to the same level for all of the film clips with the clip gain tool found in Pro Tools 12. Figure 1 and figure 2 show screenshots of the captured frame for the stimuli.

All the editing and sound design were made in Pro Tools 12. When the editing was done the film clips were exported to a quicktime file, using Pro Tools, with an audio format of PCM Wave, 24 bits of depth and 48 kHz sample rate. These were the files that later was used in the listening test.

Figure 1: Screenshot of the frame used for the male stimuli
2.2. Pre study
An informal pre study was done with three audio engineers. The purpose of the pre study was to investigate if the subjects clearly could pick a preference and if they could identify attributes that differed between the stimuli. The subjects was able to freely switch between the overhead and the underneath microphone technique. The pre study showed that it was possible to choose a preference between the techniques. The subjects could also identify similar differences in attributes between the techniques.

The attributes identified were:
Bass frequency content
Mid frequency content
Intelligibility
Presence

2.3. Listening test
2.3.1. Listening environment
The listening test was conducted in room L158 (the screening room) at LTU Campus Piteå. It was decided to conduct the test in this room because its cinema like environment. The listening room is built to mimic a small scale cinema with 28 seats in four rows. A 5.1 audio setup is installed in the
room however this test was played back in stereo. The rooms is six by eight meters, with a ceiling height of six meters. To shorten reverb time the rooms is acoustically treated. It is also sound isolated to minimize external noise leaking in through the walls. Figure 3 shows a representation of the listening environment.

![Figure 3: The listening environment](image)

2.3.2. Subjects and test procedure

20 subjects participated in the listening test. 13 of these listeners were trained listeners which, in this study, means that they were all students at the audio technology program. The remaining seven were untrained listeners and students at some of the other programs. The test was done with one subject at a time and took about 10-15 minutes to complete. The subject were asked to sit in the sweet spot in the screening room and watch and listen to the film clips played back to them with the amplifier set to -22 dB, which gave an SPL of 62 dB in the sweet spot. The instructions given to the subject were that the subject would watch and listen to eight film clips in four pairs and that they should focus on the dialogue in the film clips. For each pair the subject would write
down a preference, A or B, with the first clip in the pair being A. Afterwards the subject would answer the following question: *describe why you chose A over B or vice versa*. The questionnaire is found in appendix A. The stimuli order was randomized meaning that all the subjects listened to the eight film clips, and the overhead and underneath microphone, in different order. The order of the playback is found in appendix B.

The purpose of the question was to collect qualitative data and with that, investigate if the answers were similar between the subjects on why they chose one stimulus over another. Another interesting factor was to investigate if there was a difference in the answers depending on if the subject was a trained or untrained listener. The second purpose was to be able to tie the answers from the subjects to already existing audio quality attributes.

The original alternative to the question was to take the attributes from the pre study and have the subjects rate them on a scale. It was later on decided to let the subjects, in their own words, describe the differences they heard in the sound. The argument for this is that if the subjects would have been given attributes to rate it is possible that the subjects would be concentrating too much on the already given attributes. It was also more interesting for the study to gather information that eventually could be tied to already existing attributes.

**Playback equipment:**
Speakers: Bowers & Wilkins 803
Subwoofer: Bowers & Wilkins DB1
Speaker amplifier/processor: Classé SSP-800
Playback software: VLC Media Player
SPL-Meter: Norsonic nor131
3. Results and analysis

3.1. The A/B-test

To determine if the results from the preference A/B tests are statistically significant a binomial distribution analysis was used. Table 1 shows the results from all of the subjects, table 2 shows the results from the trained listeners and table 3 shows the results from the untrained listeners. A table of cumulative binomial distribution can be found in appendix C.

Table 1: Results for all subjects, the red cells show the disqualified subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>Male 1 O</th>
<th>Male 1 U</th>
<th>Male 2 O</th>
<th>Male 2 U</th>
<th>Female 1 O</th>
<th>Female 1 U</th>
<th>Female 2 O</th>
<th>Female 2 U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=19</td>
<td>n=19</td>
<td>n=20</td>
<td>n=20</td>
<td>n=19</td>
<td>n=20</td>
<td>n=20</td>
<td></td>
</tr>
<tr>
<td>Subject 1</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 2</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 3</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 4</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 5</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 6</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 7</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 8</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 9</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 10</td>
<td>0 0</td>
<td>1 0</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 11</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 12</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 13</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 14</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 15</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 16</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 17</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 18</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 19</td>
<td>0 1</td>
<td>0 1</td>
<td>1 0</td>
<td>0 1</td>
<td>0 1</td>
<td>0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject 20</td>
<td>1 0</td>
<td>0 1</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td>1 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Total | 9 | 10 | 8 | 12 | 7 | 13 |
Table 2: Results for the trained listeners

<table>
<thead>
<tr>
<th></th>
<th>Male 1 O</th>
<th>Male 1 U</th>
<th>Male 2 O</th>
<th>Male 2 U</th>
<th>Female 1 O</th>
<th>Female 1 U</th>
<th>Female 2 O</th>
<th>Female 2 U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
<td>n=13</td>
</tr>
<tr>
<td>Subject 1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 9</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 11</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 14</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 16</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 17</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 18</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 19</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

7 6 7 6 6 7 3 10

Table 3: Results for the untrained listeners, the red cells show the disqualified subject

<table>
<thead>
<tr>
<th></th>
<th>Male 1 O</th>
<th>Male 1 U</th>
<th>Male 2 O</th>
<th>Male 2 U</th>
<th>Female 1 O</th>
<th>Female 1 U</th>
<th>Female 2 O</th>
<th>Female 2 U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=6</td>
<td>N=6</td>
<td>n=7</td>
<td>n=7</td>
<td>n=7</td>
<td>n=7</td>
<td>n=7</td>
<td>n=7</td>
</tr>
<tr>
<td>Subject 7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 14</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Subject 15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 20</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

2 4 2 4 2 5 4 3
The reason why “Male 1” and “Male 2” have one less subject in the tables for all subjects and the untrained listeners is because of a misunderstanding in the instructions. The subjects were instructed to pick one preference, but one subject for “male 1” and one subject for “male 2” failed to do so and therefore their results were disqualified for these two trials.

In an A/B-test the results required for a confidence level of 95% depend on the number of subjects. The required results needed can be found in a binomial distribution table.

Probability of Success, $p = 0.5$

\begin{align*}
n &= 20 \\
\text{Required answers: } &15 \\
1 - 0.020695 &= 0.979305 \end{align*}

\begin{align*}
n &= 19 \\
\text{Required answers: } &14 \\
1 - 0.031784 &= 0.968216 \end{align*}

\begin{align*}
n &= 13 \\
\text{Required answers: } &10 \\
1 - 0.046143 &= 0.953857 \end{align*}

\begin{align*}
n &= 7 \\
\text{Required answers: } &7 \\
1 - 0.007813 &= 0.992187 \end{align*}

\begin{align*}
n &= 6 \\
\text{Required answers: } &6 \\
1 - 0.015625 &= 0.984375 \end{align*}

This means that out of e.g. 20 subjects at least 15 have to pick the same answer to make it statistically significant. When reviewing the tables it can be seen that only one of the trials meets this criteria. Out of 13 trained listeners 10 chose the underneath microphone technique for the trial called “Female 2”. More on this in the analysis of the qualitative data.

Concerning the other trials it could be said that according to this test there is no clear preference between the techniques. To get a better understanding of the preferences chosen by the subjects the answers will be reviewed and analyzed. The purpose of this is to investigate if certain audio quality attributes can be tied to the preferences. Or if there are other factors in the stimuli that could have had an impact on the preferences chosen by the subjects.
3.2. The answers from the open question

The question given to the subjects was *describe why you chose A over B or vice versa*. To get a better understanding on what made the subjects chose their preferences, the open question has been analyzed to investigate if there are certain attributes that have affected the subjects answers. Since the A/B results was not statistically significant for most of the conditions, it was decided to look at the answers for the open question in a more general way. However since the underneath technique for the second female stimulus was statistically significant, the answers will be analyzed more specifically for this trial to investigate if there is a similar pattern between the trained listeners who chose the underneath technique as their preference. A third analysis will also be done to investigate what the difference between the overhead and the underneath technique is.

3.2.1. Attributes in general

The following list shows the most common attributes found in the answers from the subjects, as well as how many of the subjects that mentioned these attributes in their answers.

- Intelligibility, 7
- Background noise, 6
- Clarity, 5
- Naturalness, 5
- Bass, 5
- Presence, 2
- High mid, 2
- Sharpness, 2
- Dynamics, 2
- Muffled, 2
- Low mid, 1
- Treble, 1
- Brightness, 1

To get a better understanding of what these attributes mean, it is necessary to break them down. Some of these attributes are easy to understand but for some of the other attributes are more complex. Mapp (2016) wrote that many factors can affect the intelligibility of movie or TV sound. These factors relate to the main processes involved. These processes are sound capture, processing and reproduction. Apart from this technical chain it is also necessary to consider two other factors. The articulation and clarity
of the person speaking and the hearing acuity of the listeners. It is possible that these two latter factors are the ones affecting the intelligibility most since they are not linear.

Clarity and its counterpart muffled was also mentioned by the subjects. Clear speech is easily intelligible as opposed to speech with muffled phonemes, reducing its intelligibility (Bech & Zacharov, 2006), which means that clarity of speech is an important factor when it comes to intelligibility. Other factors regarding the speech is modulation, speech rate, accent, familiarity/fluency of language and voice level (Mapp, 2016).

Naturalness can be hard to define since there are different factors that can be tied to naturalness. Dynamics is an attribute that was mentioned by a few subjects. The reason why dynamics get tied to naturalness is because the subjects who mentioned dynamics in their answers thought that it made the speech more unnatural. Bech & Zacharov (2006) defines two attributes that relates to the general naturalness of speech. These attributes are mechanic, which means that the speech has mechanic-like characteristics that makes it sound like it is produced by a synthesiser, and metallic which make the speech sound cold as having a harsh resonance in the voice. These attributes were however not mentioned by any of the subjects in this study. Some of the subjects even chose their preference based on the statement “it felt more natural”, to get something out of an answer like that it would be necessary with a follow up question.

Sharpness was mentioned and perceived as a positive attribute by the subjects. As it contributed to clearer and more intelligible speech. Bech & Zacharov (2006) defines tense or sharp speech as opposed to relaxed speech. This could be interpreted as both positive and negative depending on the context. In different situations it is better with more sharp speech while in some situations more relaxed speech is preferred. However in this test the sharpness was perceived as a positive attribute.

Presence was also perceived as a positive attribute. This could be tied to open or distant speech which is defined by Bech & Zacharov (2006) as: “The speech is distant or thin, and sounds like has been occluded. In contrast, when speech is open, there is an impression of a speaker near the listener.”

The remaining attributes bass, high mid, low mid, treble and brightness are all frequency spectrum attributes. Attributes regarding the lower frequencies, bass and low mid, contributed to less intelligible speech according to the subjects. While attributes regarding the higher frequencies, high mid and treble, contributed to more intelligible speech. Brightness does have connection to the higher frequencies as well. Dark or bright speech, if the speech is dark its low frequency components dominate, and becomes gradually brighter by the inclusion of its high frequency components and the attenuation of its low frequency components (Bech & Zacharov, 2006).
3.2.2. Other factors in choice of preference

The background noise was mentioned by some of the subjects as something negative. The background noise in the stimuli being the sound design that was added to the stimuli to create a living sound environment for the film clips. Some of the subjects chose their preference and motivated the choice by answering that they chose the film clip where the background noise wasn’t as present. An interesting factor to mention here is that the subjects got instructions to concentrate on the dialogue in the film clip. But even with these instructions some of the subjects gave these answers. These were all untrained listeners. Which is understandable because the untrained listeners aren’t as used to listen and isolate different elements in a film soundtrack as trained listeners are. Another interesting factor to relate to this is that all the background sounds were level matched between the stimuli. The only thing that varied between the stimuli regarding the sound design is spot effects. Spot effects being sound effects that accompanies the ambient sounds, e.g. footsteps, telephone signal, a door that’s being closed etc. all of these were off screen sound effects. One of the subjects e.g. answered that he/she chose the preference because the sound of the door closing was disturbing according to the subject. This shows that the subject did not concentrate on the sound of the dialogue.

Another factors that were brought up from one of the subjects were that the sound and the picture were out of sync, thus making the subject pick the preference were the sound and picture were in sync. The stimuli was reviewed and checked for sync problems. Therefore it is safe to say that there was not a problem with the sync during the listening test. What made the subject perceive the sync problems in the stimuli is hard to determine. The subjects itself also wrote in the answer that it was a possibility that the subject’s imagination made the sound and picture go out of sync. Because of this it is possible that the subject thought that the purpose of the test was to investigate difference between audio in sync and audio out of sync, thus creating the illusion of the audio being out of sync for some of the stimuli.

3.3. The answers from the statistically significant trial

As mentioned earlier a preference for the second female stimulus recorded with the underneath microphone technique was statistically significant, shown by the preferences submitted by the trained listeners. To investigate this further the answers to the open question will be compared between the trained listeners who chose this as their preference. As for the attributes in general, the following list will display the attributes identified by the trained listeners, and how many, for this particular stimulus.
• Treble, 1
• Naturalness, 2
• Sharpness, 1
• Presence, 1
• Clarity, 2
• Bass, 1

All of these attributes are familiar from the section with the attributes in general, thus not going into any deeper analysis of this list of attributes. All of these attributes were perceived as positive attributes by the subjects. Which means that these attributes are preferred for this stimulus. However the answers are not that similar to each other. Some answers were also difficult to interpret, thus it would be necessary with follow up questions to get the subjects to further explain the meaning of their answers.

One particular answer for this stimulus stood out. This subject answered that the overhead technique for this stimulus sounded like too much low-cut had been applied. Which means that the underneath technique for this subject didn’t sound as if too much low-cut had been applied. With this answer it could be said that it is more low frequency content is preferred for this particular subject.

Another interesting factor regarding these answers, and the fact that this trial got a statistically significant result, is that both of the female stimuli are technically identical. The only differences between these stimuli are one spot sound effect and that the line the actress say is different. But the other female stimulus did not get a statistically significant result. It is possible that the way the actress articulated or expressed herself differed from the second stimuli, thus generating these different results.

3.4. Comparison between the overhead and underneath microphone technique

This section will focus on attributes that have been mentioned by the subjects for the different microphone techniques. The purpose of this is to investigate if there is a possibility to draw any conclusions on what the difference between the microphone techniques may be. In table 4 the answers from the different trials have been merged to one category for the microphone techniques, meaning that e.g. both trials from the male overhead technique is shown as “Male Overhead” etc.
Table 4: Attributes for the different microphone techniques

<table>
<thead>
<tr>
<th>Male Overhead</th>
<th>Male Underneath</th>
<th>Female Overhead</th>
<th>Female Underneath</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open</td>
<td>Clearer</td>
<td>More treble/high mid which gives more intelligible speech</td>
<td></td>
</tr>
<tr>
<td>Clearer</td>
<td>More natural</td>
<td>More presence, more natural timbre</td>
<td></td>
</tr>
<tr>
<td>Less noise</td>
<td>More</td>
<td>More natural</td>
<td></td>
</tr>
<tr>
<td>更高 resolution</td>
<td>Less dynamic</td>
<td>Too much low cut</td>
<td></td>
</tr>
<tr>
<td>Muddy</td>
<td>Irritating</td>
<td>Sharp mid</td>
<td></td>
</tr>
<tr>
<td>More natural</td>
<td>bass/Low mid</td>
<td>Less noise</td>
<td></td>
</tr>
<tr>
<td>Clear mid</td>
<td>More bass</td>
<td>Clearer</td>
<td></td>
</tr>
<tr>
<td>Less natural</td>
<td></td>
<td>Too sharp</td>
<td></td>
</tr>
</tbody>
</table>

It can be seen that both techniques have been described with the same attributes. The highlighted attributes are attributes that were unique in their specific group. The male overhead technique has additionally been described as more natural as well as less natural. This shows that the subjects perceive the sound of the same technique differently. The overhead technique was also described as having a higher resolution. The reason for this is difficult to determine but one explanation could be that the increase in low mid makes the underneath technique sound like it has a lower resolution. The underneath technique was perceived as having more presence. The increase in bass and low mid can be a contributing factor to this. The same goes for the female voice. The overhead technique has been described as sharp. An explanation to this is that the increase in bass and low mid makes the underneath technique less sharp and more present.

Regarding the techniques for the male stimuli, they have been described with similar attributes by the subjects, however one factor stands out. Some of the subjects explained in their answers that the low frequency content in the underneath technique for male voice was too dominant and made the speech unclear and unnatural. While the overhead technique had a more pleasant, clear and natural timbre which made the subjects prefer the overhead technique. An explanation to this is that the male voice has, in general, more low frequency content than a female voice.

The opposite could be found for the female stimuli. Some of the subjects described that the overhead technique sounded thin in comparison with the underneath technique, therefore they preferred the underneath technique. This is also explained by the fact that the low frequency content of the female
voice is of lower level than that in the male voice, and in this particular test this was prefered by some of the subjects.

As mentioned earlier it is hard to draw any conclusion on the differences between the techniques from the answers submitted by the test subjects. Even though this tendency with the low frequency content was found, it is still only a few subjects who mentioned it. All of the techniques have been described with similar attributes from the subjects.

Some of the subjects’ answers were also excluded from this section. The explanation to this is that these answers did not contribute to the descriptions of audio quality. Examples of these answers, from untrained listeners, are answers were it was clear that the subjects could not focus and concentrate on the sound of the dialogue. Examples of these answers are were the subjects chose their preference based on the amount of background noise, sound effects or music. Also answers with the motivation “it felt better” was excluded.
4. Discussion

4.1. The method

Concerning the method in this study, it could be said that it was successful. It was possible to execute within the given time frame, and it generated a result. However there are aspects in the method that should be reflected upon and discussed.

One aspect being the actors and the scene. The generated results only applies on these particular actors and this particular scene. Which means that it is not possible to generalize and state that the results in this study will apply to all dialogue recording. The reason for this is that different actors have different voices, which will affect the sound of the recorded dialogue. The sound design is another aspect that possibly will affect the perceived audio quality of the speech. In this study the scene was accompanied by a subtle sound design due to the nature of the environment. In a more complex scene the sound design would also be more complex.

The listening environment is another interesting factor when it comes to a listening test. It is possible that the listening environment could affect the results because different rooms have different sound characteristics. The screening room was however a good environment for this listening test, because of its cinema like design. Another alternative would have been to conduct the listening test in a control room. This listening environment would have been a controlled environment optimized for sound playback. However, it is not an environment where viewers normally would watch a movie.

4.2. The results

The results gathered from the A/B-test were not statistically significant, with the exception of one trial for the trained listeners, the second female stimulus. This is interesting because, as mentioned before, the only difference between the female stimuli was the line being said as well as one sound effect. In the analysis of the answers from the open question regarding this trial it was difficult to find a consistent pattern, thus making it difficult to come to a conclusion on how this trial could generate a result that was statistically significant. It is unlikely that it was due to the order the stimuli was played back to the subjects, because all of the subjects listened to the stimuli in a different order. When reviewing the stimuli it was also noted that the actress’ expression, pitch etc. was the same in both female stimuli.

It was also difficult to draw any conclusions from the answers on why the subjects chose their preference. One conclusion drawn is that it varies from subjects to subject on how they explained the differences in the sound quality attributes. This is because all of the different stimuli were described with similar attributes. It is however important to mention that a few differences between the techniques could be found in the descriptions. Meaning that these attributes were perceived as differences between the microphone techniques. Regarding the second female stimulus an explanation
is that trained listeners have a common perception on what qualities a good sound needs, and that the differences between the microphone techniques were clearest in the second female stimulus.

Some of the answers were also useless regarding audio quality attributes. It was however interesting how subjects perceived differences in e.g. background noise or spot sound effects, since the level for the sound design was the same for all stimuli. It is still important to mention this group. Clearly the background noise and the spot effects have had an impact on the choice of preference for some of the untrained listeners. It was also mentioned that the untrained listeners could not focus on the dialogue track to the same extent as the trained listeners. An explanation for this could be the differences in the spot effects that drew the listeners attention away from the dialogue track.

It was also stated in the method section that a comparison between the trained and the untrained listeners would be investigated. This was not done. The reason for this was that the group of untrained listeners was much smaller than the group of trained listeners. This was a mistake that should have been solved before the experiment began. Another reason is that some of the data gathered from the untrained listeners were discarded, thus making it unfair to compare the listener groups.

4.3. Conclusion

This study suggest that there should not be a difference in preference between the microphones techniques according to the A/B-test. The reason for this, as discussed earlier, is more difficult to state. The only conclusion that can be stated is what type of audio quality attributes that affects the choice of preference. A difference in attributes between the techniques could be found, but all the different stimuli were also described with similar attributes depending on the subjects.

It was however showed that there was an increase in the lower mid frequencies for the underneath microphone technique, although it seems that this does not affect the choice of preference. This will also depend on the characteristics of the actors’ voice.

As a boom operator, working on a film set, it can be worth thinking about this. This study shows which attributes that affect the choice of preference, however the A/B-test did not, in most of the trials, show that one technique is preferred over the other. The underneath technique shows an increase in the bass/low mid area of the spectrum, which could potentially be beneficial to some voices.

4.4. Further research

The further investigate this matter an expanded version of this study could be done. An experiment with more actors, would give a greater variation of voices. A complex scene could be done with more acting, both interior and exterior, as well as a more complex sound design to accompany the scenes. The listening environment could also be changed to a bigger movie theatre, to create a more standard film experience.
Acknowledgements

I would like to thank Arne Nykänen and Nyssim Lefford for their guidance and support. Thank you also to Tony Österholm at LjudBang AB, Oskar Hansson, Lisa Selander, Henri Vaenerberg and of course all the subjects that participated in the listening test.
5. References


Appendix A: The questionnaire for the listening test

Par 1

*Preferens:*

*Beskriv varför du valde A framför B eller tvärtom:*

Par 2

*Preferens:*

*Beskriv varför du valde A framför B eller tvärtom:*

Par 3

*Preferens:*

*Beskriv varför du valde A framför B eller tvärtom:*

Par 4

*Preferens:*

*Beskriv varför du valde A framför B eller tvärtom:*
Appendix B: The playback order for the subjects

<table>
<thead>
<tr>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
<th>Subject 5</th>
<th>Subject 6</th>
<th>Subject 7</th>
<th>Subject 8</th>
<th>Subject 9</th>
<th>Subject 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>untrained</td>
<td>trained</td>
<td>trained</td>
<td>untrained</td>
</tr>
<tr>
<td>F2 O/U</td>
<td>F2 O/U</td>
<td>F2 O/U</td>
<td>M2 O/U</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>M1 O/U</td>
<td>M1 O/U</td>
<td>M2 U/O</td>
<td>M2 U/O</td>
</tr>
<tr>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>M2 U/O</td>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
</tr>
<tr>
<td>M1 O/U</td>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
</tr>
<tr>
<td>M2 U/O</td>
<td>M1 O/U</td>
<td>F1 U/O</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
<td>M1 O/U</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject 11</th>
<th>Subject 12</th>
<th>Subject 13</th>
<th>Subject 14</th>
<th>Subject 15</th>
<th>Subject 16</th>
<th>Subject 17</th>
<th>Subject 18</th>
<th>Subject 19</th>
<th>Subject 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>trained</td>
<td>untrained</td>
<td>untrained</td>
<td>untrained</td>
<td>untrained</td>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>trained</td>
<td>untrained</td>
</tr>
<tr>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
<td>M1 O/U</td>
<td>F1 U/O</td>
<td>M1 O/U</td>
<td>F1 U/O</td>
<td>F1 U/O</td>
<td>F2 O/U</td>
</tr>
<tr>
<td>F2 O/U</td>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>F2 O/U</td>
<td>M2 U/O</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
<td>F2 O/U</td>
<td>M1 O/U</td>
<td>M1 O/U</td>
</tr>
<tr>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>M2 U/O</td>
<td>F1 U/O</td>
<td>M2 U/O</td>
<td>M2 U/O</td>
<td>M1 O/U</td>
<td>M2 U/O</td>
<td>M2 U/O</td>
</tr>
<tr>
<td>F1 U/O</td>
<td>M1 O/U</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>M1 O/U</td>
<td>F2 O/U</td>
<td>M2 U/O</td>
<td>F2 O/U</td>
<td>F1 U/O</td>
<td>M1 U/O</td>
</tr>
</tbody>
</table>

M1 = Male stimulus 1  
M2 = Male stimulus 2  
F1 = Female stimulus 1  
F2 = Female stimulus 2  

O/U or U/O = in which order the techniques were played back, with the left being the first
### Appendix C: Table of Cumulative Binomial Probability

<table>
<thead>
<tr>
<th>n</th>
<th>x</th>
<th>0.00</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.90</th>
<th>0.95</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.03</td>
<td>0.06</td>
<td>0.10</td>
<td>0.14</td>
<td>0.20</td>
<td>0.29</td>
<td>0.42</td>
<td>0.55</td>
<td>0.64</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.15</td>
<td>0.28</td>
<td>0.42</td>
<td>0.54</td>
<td>0.68</td>
<td>0.80</td>
<td>0.91</td>
<td>0.96</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.25</td>
<td>0.46</td>
<td>0.64</td>
<td>0.79</td>
<td>0.89</td>
<td>0.97</td>
<td>1.00</td>
<td>0.99</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.18</td>
<td>0.29</td>
<td>0.42</td>
<td>0.53</td>
<td>0.67</td>
<td>0.77</td>
<td>0.86</td>
<td>0.92</td>
<td>0.95</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.12</td>
<td>0.19</td>
<td>0.31</td>
<td>0.41</td>
<td>0.52</td>
<td>0.60</td>
<td>0.69</td>
<td>0.76</td>
<td>0.80</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.12</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.59</td>
<td>0.67</td>
<td>0.74</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.12</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.59</td>
<td>0.67</td>
<td>0.74</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.12</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.59</td>
<td>0.67</td>
<td>0.74</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.12</td>
<td>0.20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.59</td>
<td>0.67</td>
<td>0.74</td>
<td>0.77</td>
<td>0.80</td>
</tr>
</tbody>
</table>