DIRECTIVITY OF HUMAN AND ARTIFICIAL SPEECH

Teemu Halkosaari

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Otakaari 5 A, 02150 Espoo, Finland
thalkosa@cc.hut.fi

Markus Vaalgamaa

Nokia Technology
Audio Entity
Itämerenkatu 11-13, 00180 Helsinki, Finland
markus.vaalgamaa@nokia.com

ABSTRACT

This paper is a study on how directivity characteristics of artificial mouths correspond to the directivity of a real speaker. The motivation for the research is the measurement methods that are applied in the telecommunication industry for the microphones of phones and their accessories.

Responses of a mouth of an artificial torso simulator, B&K HATS 4128 were measured to several positions [1]. The same measurements were repeated for a group of test subjects. The measurement positions corresponded to the same positions were the microphones of phones and their accessories, so called headsets, would lay.

The basic mechanisms that produce the directivity pattern are discussed and also the contribution of the speech content is shown. The main contributor to the directivity is the mouth aperture size. The acoustical characteristics of the upper body are also a significant factor if the position is not directly in front of the mouth.

More than 10 dB differences were found between HATS and an averaged test subject. The key finding is that HATS is too directional at high frequencies. To enhance the correspondence of the telephonometry to real speakers, a simple equalization scheme and two structural improvement proposals to HATS are introduced.

1. INTRODUCTION

When a microphone module is designed for a new phone model the frequency response of the microphone is typically measured with an artificial mouth. The distance of the microphone to the mouth depends on the dimensions of the phone. The trend is that the mobile phones are getting smaller and smaller and thus the microphones in the phones are substantially far from the mouth compared to old landline phones.

The so-called headset equipment are also setting new requirements for the telephonometry. In these accessory gadgets the microphones lay on the chest or somewhere else not in a familiar phone microphone position. Generally speaking the directivity of the artificial mouths should cover nowadays accurately a larger set of positions than they were originally designed for. Still it is not well known what are the frequency response and sensitivity characteristics of current artificial mouths used in acoustical measurements in off-axis positions.

There is no extensive literature on the subject. Some studies can be found where the directivity of the mouth as a sound source is measured with a group of test subjects [2, 3, 4, 5, 6, 7, 8, 9]. Also some publications, standardization, and product specifications can be found on the directivity of artificial mouths [10, 11, 12, 13, 1]. Nevertheless the directivities of artificial mouths are not systematically compared to test subjects by similar measurements in any of these studies. thus there is clearly a need for measurement-based information how accurately the artificial mouths simulate the real human mouth and should their design be improved.

There are standards that define guidelines for the directivity pattern of the artificial mouths. The most important ones are: ITU-T P.58, Head and torso simulator for telephonometry and ITU-T P.51, Artificial mouth [11, 12]. Artificial mouths, especially the so-called Head and Torso Simulator (HATS), will most probably be used in future for wide band measurements, even though the mouth subassembly and its standardization should be improved in any case. The standards do not define especially in near field the directivity pattern for the positions that the microphones of the headsets and small phones use. In the narrow frequency bandwidth (300 Hz - 3400 Hz) the directional features are not yet affecting and so the simulators are adequate. But in the wide frequency bandwidth (150 Hz - 7 kHz) the situation is significantly different, as we will see later in this paper.

This study concentrates on both directivity characteristics of the human and B&K HATS 4128. The most important part of the study is to determine if there is a difference between the directivities of HATS and an average person. As a result the study should give basic information how the HATS measurements for phone handsets should be designed and for example should the difference between human and HATS directivities be compensated in the measurements.

Directivity is also studied in general. The far and near field conditions are discussed. The sound radiation pattern affected by head and torso is studied both by measurements and by modeling. Also characteristics of speech content affecting to the directivity are discussed.

The structure of the paper is mainly leaning on the measurements and the outcome of those. The details of the measurements are listed in the next section. Two simple models for a mouth as sound source are presented in Section 3. The results are shown by various figures in Section 4. In Section 5 the improvement proposals for telephonometry are discussed and finally in the end of the document the outcome of the study is gathered in a short conclusion.

2. MEASUREMENTS

Some kind of key requirements have to be defined as the basis for the measurements. The measurement system and process had to be compact and specified accurately to ensure reliable results. To ensure the easy conduction of the measurements, a multi-channel recording system was designed for the test subject measurements. The same system with minimum changes was used throughout all the measurement activities. The B&K HATS 4128 was measured using two-channel impulse response measurement system.

There were several objectives in the measurements. First of all the directivity of B&K HATS 4128 was to be determined to several positions by transfer functions that are referred to a fixed reference position. Next the measurements were to be repeated for a group of test subjects using the same measurement positions. The speech material was selected so that it would represent on average the content of natural speech. Some additional measurements were conducted for the HATS to study the characteristics of the sound field in general.

In phones and headsets the microphones lay normally near the cheek or chest. Depending on the phone model there are several different positions for microphones between the ear and the mouth. If we count in also the positions for other accessories, the amount of possible positions gets really large. All of these self-evidently cannot be covered in the study. Nevertheless the scope is to find some general features for directivity relying on an adequate amount of measurements. The positions are defined in Section 2.2.

2.1. Equipment

Brüel&Kjær HATS 4128 was measured in an anechoic chamber at Nokia facilities at Salo. The chamber ensures the free field conditions down to 90 Hz. B&K HATS was the main target on the study because it is most widely employed [1].

The MLS feature in the APWin measurement platform from Audio Precision was applied to acquire the impulse responses. The measurement itself was a set of consequent cases with outcome of impulses responses for two channels. 32 kHz sampling rate and 32767 sample MLS were the parameters for measuring. Two B&K measurement condenser microphones 1/8 inch in diameter were the pick-ups.

The directivity of the test subjects had to be studied by recording the speech. Measurements took place in an anechoic chamber at Helsinki University of Technology. The data acquisition system was build up from IOTech Wavebook/516 measurement system (IOTech, WaveBook/516, 16-bit 1-MHz Data Acquisition System). IOTech system has 8 input channels. One of the channels was reserved as the reference. So the system enabled directivity measurement by 7 transfer function estimates in parallel.

The IOTech is a separate DSP unit that can be connected over LPT-cable to a laptop computer. IOTech does the A/D conversion and laptop acts as storage for the data. The IOTech Wave book has 16-bit A/D conversion for each 8 channels. 32 kHz sample rate was selected also for these measurements.

Small electret microphones (Sennheiser KE 4-211-2) seemed to be the best choice for the test subject measurements because they could be easily attached to the test subjects. The diameter of the microphones is 4.75 mm. The responses of the microphones were calibrated by comparing each one to a B&K condenser microphone.

The measurement acquisition was controlled with dedicated IOTech user interface software. Signal levels were monitored and amplified with microphone pre-amplifiers (E.A.A. Professional Stereo Preamplifier, PSP-2) so that adequate SNR was achieved. The acquisition system was wired outside the anechoic chamber where the measurement operator could steer the session. Beside the acquisition system a test subject monitoring and guidance system was built up. Both a video and an audio link was between the test subject inside the chamber and the operator.

2.2. Target positions

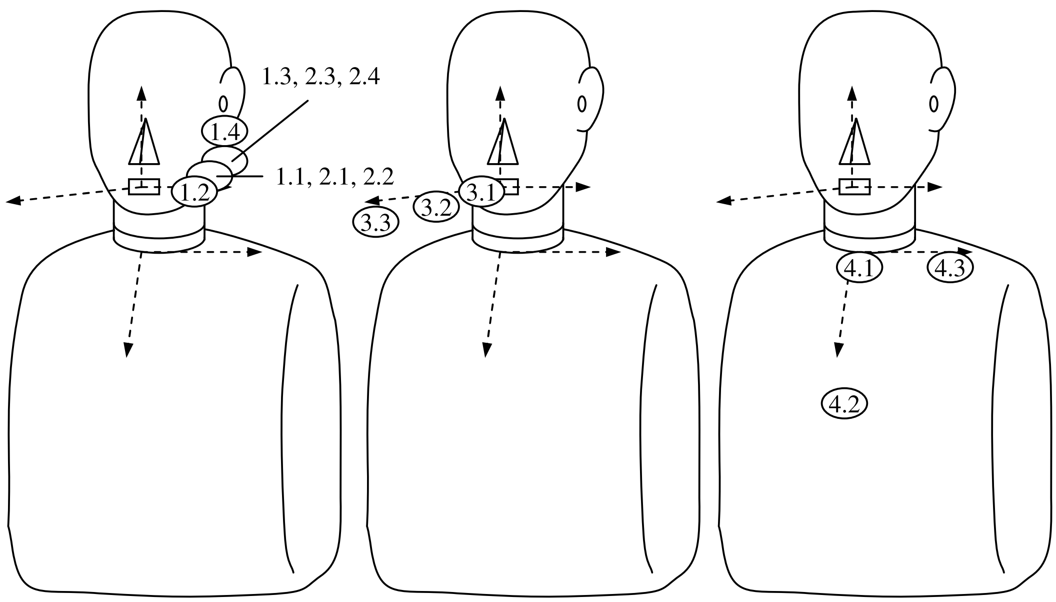


Figure 1: *Measurement positions.*

The measurement positions are illustrated in Figure 1. The B&K HATS 4128 was measured in all positions in parallel with the Mouth Reference Position (MRP) i.e. position 3.1.

The recording system for test subjects had eight channels. So the same eight positions were used throughout the activity: positions near the cheek 1.1 - 1.4, on the chest 4.1 - 4.3, and as reference 3.4. The far field reference was chosen because the positioning of a microphone to MRP with test subjects was seen difficult as well as the noise caused by the airflow from the mouth.

A sophisticated measurement helmet was designed and built to attach the measurement microphones as exact as possible to the specified positions in test subject measurements. Also the microphones in the close to chest were attached to each other in an array. The positioning equipment can be seen in Figure 2.

Table 1: *Measurements positions defined by coordinates. The coordinate axes are referred to the lip plane and its perpendicular (see Figure 1). On chest the positions are directly on the chest and the axes run downwards and sideward.*

Position	On axis	Off axis horizontal	Off axis vertical	Description
1.1	0	60	0	Cheek, large phone, normal angle
1.2	-10	70	-40	Cheek, very large phone, normal angle
1.3	-30	85	10	Cheek, small phone, normal angle
1.4	-70	100	30	Cheek, very small phone or boom HS
2.1	-10	85	-20	Cheek, large phone, pointing down
2.2	0	100	0	Cheek, large phone, far from cheek
2.3	-50	95	-15	Cheek, small phone, pointing down
2.4	-45	110	0	Cheek, small phone, far from cheek
3.1	25	0	0	MRP
3.2	100	0	0	Far field
3.3	250	0	0	Far field
3.4	500	0	0	Far field, reference point for test subjects
3.5	1000	0	0	Far field
	Downwards	Sideward		
4.1	50	0		On chest, near throat
4.2	200	0		On chest, middle of chest
4.3	0	100		On chest, near shoulder

The microphone positioning was referred to lip ring and its perpendicular. This is directly adopted from the ITU standardization [11, 12]. On the chest the positions lay freely and therefore the axes run from the throat downwards and sideward (see Figure 1).

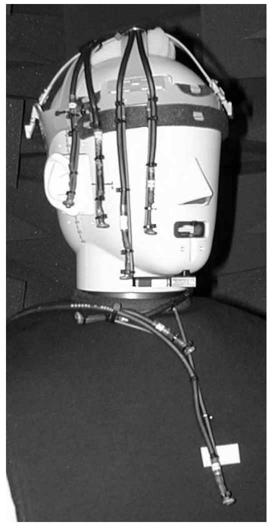


Figure 2: *The measurement helmet and the chest microphone array can be seen on the B&K HATS 4128. The microphones (Sennheiser KE 4-211-2) are the in the positions that were used throughout the test subject measurements.*

2.3. Measurement procedures

B&K HATS 4128 was measured in all of the positions three times: with and without the measurement vest (B&K 0600 shoulder damping fabrique) and also without the torso. The output of each measurement case was two 32767-sample impulse responses at 32 kHz sample rate for the reference point (MRP) and the point of interest. In addition some extra measurements were conducted. In two positions near the cheek the measurement was repeated with different mouth sizes: normal (30 mm × 11 mm), blocked to half of its width (15 mm × 11 mm), and without the mouth adaptor (42 mm × 16 mm).

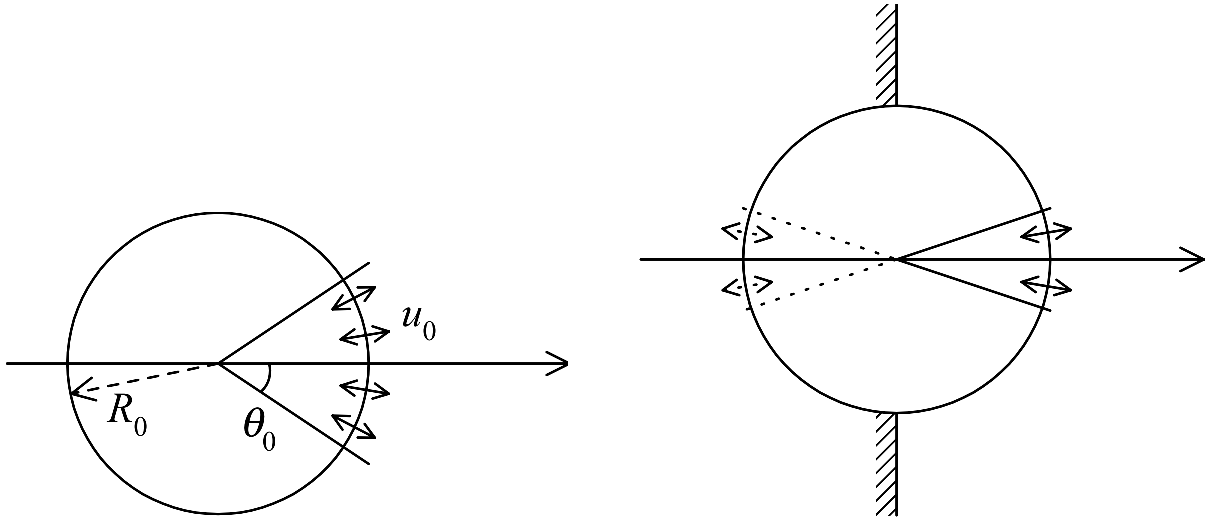
A sentence in Finnish was selected to test subject recordings because Finnish-speaking test subjects were the easiest to recruit. Every subject had to pronounce the sentence “Kaksi vuotta sitten kävimme Ravintola Gabrielissa Helsingissä, ja söimme siellä padallisen fasaania banaanilla höystettynä.” and it was recorded with the eight microphones at the same time. In the sentence there is at least one occurrence of all vowels and consonants of the Finnish language. There are altogether 95 separate short or long phonemes in the sentence.

The measurement session was repeated for 13 test subjects: 5 female and 8 male. Age of the subjects vary between 20 and 30 years. The group consisted of different body sizes from 160 cm tall to 190 cm. None of the test subjects had speech defects and all had Finnish as the mother tongue.

Test subjects were instructed to speak clearly and be still during the recording. A headrest was attached to the chair and two microphone stands was lined in front of the speaker, so that it would be easy to focus to keep the head still. The recordings were always repeated if some mistake in articulation or movement was monitored.

3. MODELING

The directivity of the voice production system can be estimated by various kinds of models. Later on the results are assessed comparing them to modeling. Also the reasons for the directivity characteristics can be predicted by altering the parameters in the models.



(a) *Piston and sphere model.*

(b) *A infinite baffle is added to the model. The mirror source method is illustrated.*

Figure 3: *Models that are applied in the study.*

The simplest model for the head is to model it with a sphere. The mouth is approximated with a round piston on the sphere. By these means the rotationally symmetrical coordinates can be acquired and therefore there is just two axes: the radius from the center of the sphere r and the angle from the piston perpendicular θ. The model is illustrated in Figure 3(a).

The particle velocity u_r on the surface of the sphere is

$$u_r = \begin{cases} u_0 & , \quad \theta \leq \theta_0 \\ 0 & , \quad \theta > \theta_0 \end{cases} \tag{1}$$

In other words a circular part of the sphere radiates and the rest of the surface is fixed and baffled. The displacement on the piston is radial and in this case it is an adequate approximation because the mouth aperture is very small compared to the head radius.

When the sound field is harmonic, the pressure field $p(r, \theta)$ for a general axisymmetric spherical sound source is

$$p(r, \theta) = i\rho_0 c_0 \sum_{n=0}^{\infty} \kappa_n \frac{h_n^{(2)}(kr)}{h_n^{(2)}(kR_0)} P_n(\cos\theta). \tag{2}$$

P_n is Legendre polynomial, $h_n^{(2)}$ Hankel function, k the wave number, ρ_0 the density of air, and c_0 the velocity of sound. Coefficients κ_0 are dependent of the displacement pattern on the sphere. In this case they are

$$\kappa_n = \frac{u_0}{2} \left[P_{n-1}(\cos\theta_0) - P_{n+1}(\cos\theta_0) \right]. \tag{3}$$

[14, 15].

The model has two parameters: the radius R_0 and the mouth aperture angel θ_0 . When the model is later on applied, these parameters have to be selected within some reasonable principles. The head dimensions cannot be applied directly to the model because the head and body dimensions are not axisymmetric. Nevertheless some kind of average value gives reasonable results. For the head radius 10 cm value was used. The mouth aperture angle was set so that the mouth cross-section area would be 3 cm^2 or close to it depending on the case.

The model represented above gives only a simplified view how the head interacts with the radiation from the mouth. Self-evidently the whole body affects to the sound field. At least shoulders and chest diffract, reflect, and absorb the sound.

If we consider the sound field close to the lip plane perpendicular and positions on the chest, the most significant effect of the body is the direct reflection. It causes a comb filtering shaped response to the field. One direct and simple way to model this reflection can be seen in Figure 3(b) where the space is divided with an infinite baffle. The head model itself is the same.

The main advantage in the presented baffle model is that it can be formulated using a mirror source method. The total field with the reflection can be expressed with the following equation

$$p(r,\theta) = p_{\text{direct}}(r,\theta) + p_{\text{mirror}}(r,-\theta). \tag{4}$$

4. RESULTS

4.1. Analysis overview

There were two approaches to analyze the measurement data because two different measurement systems were used. In both cases the aim was to obtain transfer function estimates from a reference position to other positions in the field.

The data from HATS measurements was directly impulse responses. Therefore the transfer functions were directly obtained by FFT. By supposing that the system is linear and the reference position identical in each case, the transfer functions from point to point can be calculated by dividing the FFTs.

Transfer function estimates were calculated from the test subject recordings by windowing the active signal in the recordings in 1024 sample frames (32ms at 32kHz sample rate) with a 50% overlap. The delay difference between channels was compensated. The delay was estimated by calculating cross correlation between channels. FFT was applied to each frame and by averaging the frames a transfer function estimate was acquired [16]. The transfer functions were averaged to 1/3 octave resolution.

4.1.1. Reliability consideration

The simplest means to consider the reliability of the measurement data is to examine the coherences and signal-to-noise ratios (SNR). Both SNR and coherence are obtained by cross spectrum estimates and auto spectrum estimates [16, 17]. The noise sample was taken from the end of each recording starting from the end of the last segmented active speech. The HATS measurements are omitted in the consideration because the reliability of those measurements was far better than test subject measurements in any case.

SNR and coherence were adequate on all frequency bands that were considered. The worst SNR in all channels was greater than 25 dB over the wide band. The coherence between the reference position and other positions were greater than 0.8 on narrow band and greater than 0.65 on wide band. The values imply that the transfer function estimates are reliable on wide band and very reliable on narrow band [16, 17].

If results are averaged from a large set of data, the confidential intervals are one way to see by statistical means how reliable the results are. The confidential intervals are included in Figures 12 for whole data and 9 for each vowel group. The distribution is calculated from the unoverlapped frames. The intervals stay mostly within a ± 1 dB range. This implies that the results are very representative. [16]

4.2. Directivity features of mouth

The directivity of the mouth is approached by considering what are the key aspects that play the most important role. Measurement data for HATS is more reliable and so those measurements are used for the consideration.

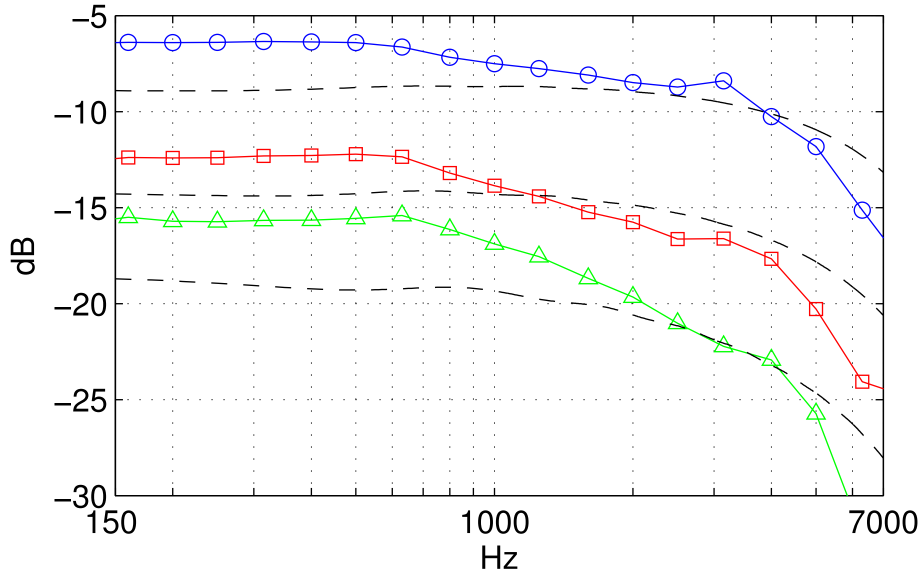


Figure 4: *Transfer functions of HATS from position 3.1 (MRP) to positions 1.1, 1.3, and 1.4 (○, □, and △). Corresponding modeled cases are included for the same positions.*

The sound field is produced by the radiation from the mouth aperture and modified by the head and torso depending on their dimensions and position. Generally speaking there are three different aspects that affect the sound field:

1. Mouth cross-section area.
2. Far field ↔ Near field.
3. Reflections from body.

Next there is some explanation on each of these items on the list.

The cross-section area and shape of the mouth aperture are connected to how much the bare mouth directs. If the size of the mouth is substantially large compared to the wavelength the mouth starts to direct. On low frequencies the mouth is more or less a simple omnidirectional point source.

Second, near field and far field differ because of the head dimensions. The study is only concentrated to the field near the mouth perpendicular. In far field on high frequencies, the head enhances the directivity the same way as the mouth if assembled to an infinite wall. On low frequencies the head does not affect. On the other hand if the position is shifted off the perpendicular, for example behind the head, the head shadows the sound on high frequencies.

Third, reflections, diffraction, and absorption from the upper body modify remarkably the sound field. Near the mouth, for example close to the cheek, these fluctuations are not strong enough so they can be omitted. On the other hand in far field and especially near the chest significant fluctuations can be seen.

Generally speaking the directivity of the mouth is seen as reduction of high frequencies when the position shifted towards the ear (see Figure 4). The same kind of effect is seen when the position goes downwards the chest (see Figure 5). The reduction of high frequencies is more or less monotonic and when the position is closer to the ear the slide is steeper.

It is advisable to use an about 2 cm thick measurement vest covering the torso when B&K HATS 4128 is used in acoustic measurements. It is supposed that this vest enhances the correspondence with the human upper body. Nevertheless the case is not that simple when the phone equipment is considered. For example some of the headsets are placed close to the chest and therefore the covering of the chest is essential for the responses.

In Figures 5 and 6 we see that the chest reflections cause peaks and dips to the responses. Close to 1 kHz there is a strong dip in the transfer functions for the position in 0.5 m distance (Figure 6). The dip around 1 kHz implies that there is about 15 cm difference in the path of reflected and direct sound. This is in correspondence with the dimensions. The mouth in the model is in the center of the head and therefore the path difference from mouth to chest is longer and the dip is on a lower frequency. If the vest is taken off the dip shifts to higher frequency because the reflection comes faster i.e. the sound velocity in vest is less. In the curve for bare head we see only the 3 dB level difference between high and low frequencies.

In the Figure 5 we see a dip near 5 kHz that is caused by the reflection through the vest. If the vest is taken off the position is directly on the hard surface and so only duplication of pressure field occurs. The dips in both Figures 5 and 6 are repeated on higher frequencies in comb filtering shape but the frequency smoothing omits it.

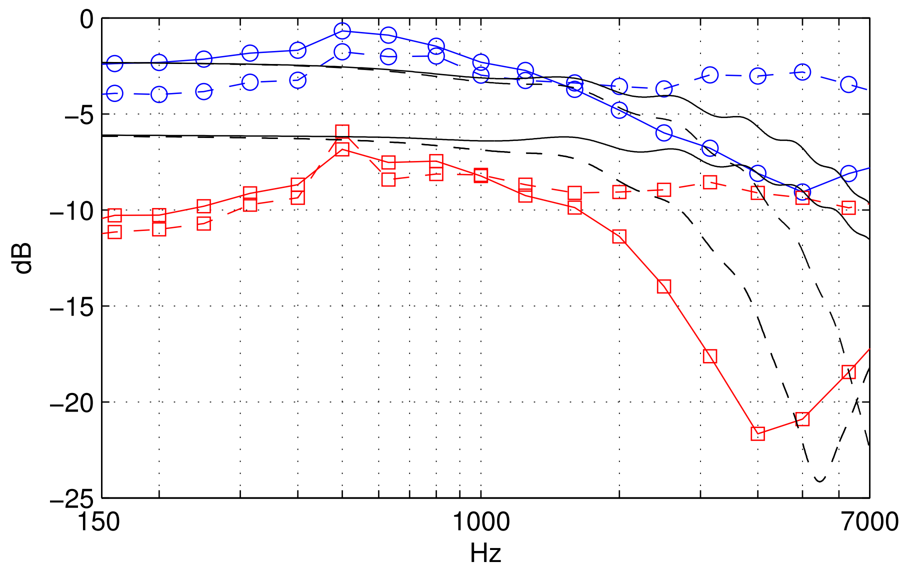


Figure 5: *Transfer functions of HATS from MRP to positions 4.1 and 4.2 (○ and □) and the same measurements without the vest (dashed ○ and □). Corresponding modeled cases with the infinite baffle are included for the same positions. Both positions are modeled with to different distances from the infinite baffle: 1 cm (solid) and 8 cm (dashed)*

4.3. Articial mouth compared to test subjects

The differences of the averaged test subjects and artificial mouths are seen in Figures 7 and 8 in five positions. The curves are calculated by dividing transfer function for averaged person with the transfer function for an artificial mouth to same position. So if the curve has positive values the artificial mouth is more directional and vice versa.

In the test subject measurements the reference position was in 500mm distance because MRP would have been difficult to implement to the measurements. So the reference position in far field causes same fluctuation to the curves because there is also difference in reference between test subjects and artificial mouths.

The difference curves near the cheek are substantially flat on narrow band but on wide band the directivity difference is significant. Interesting remark is that the differences are more or less independent of the position. There is a little more difference if the position is closer to the ear if the overall level difference is omitted.

The two positions near the chest show that there is significant difference in reflection from chest between HATS and test subjects. This is seen near 800 Hz for the reference and near 5 kHz for the close to chest reflection.

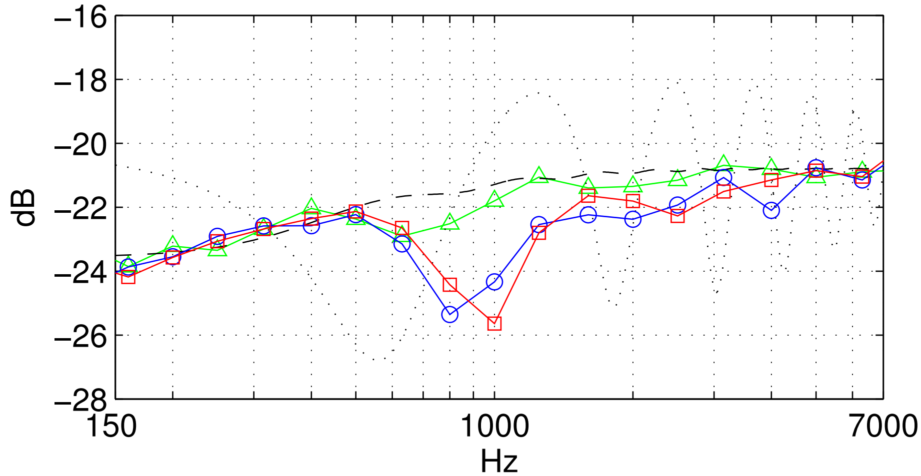


Figure 6: *Transfer functions of HATS from MRP to positions 3.4 (0.5 m in front of mouth) with the vest, without the vest, and without the torso (○, □, and △). Corresponding modeled cases with and without the infinite baffle are included.*

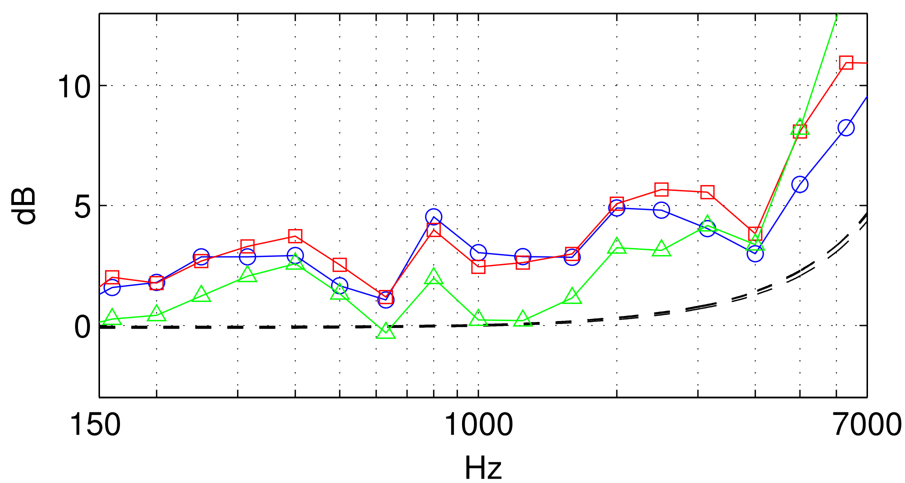


Figure 7: *Difference of averaged test subject and HATS. The comparison is done for transfer functions from microphone position 3.4 to 1.1, 1.3, and 1.4 (○, □, and △). A positive dB value implies that the HATS is more directional on that frequency. Three corresponding models are included where the mouth aperture radii for corresponding HATS and averaged person in the direct models are 1.5 cm and 0.5 cm.*

4.4. Speech content

As we saw within the results where test subjects and HATS were compared, the size of the mouth seems to be an essential factor in the difference. The curves in Figures 7 and 8 were modeled using two different mouth aperture sizes for corresponding HATS and average test subject. As it is seen the general trend of the measurement results follows the modeled curves. Because mouth aperture size seems to be one key factor we go further on with the analysis. One important aspect is to see if the mouth aperture size during speech can be predicted using directivity features. The traditional mouth and vocal tract size measurements using some kind of imaging method can be found for example in [18, 19, 20, 21].

It is important to emphasize that the mouth aperture size is effective relative value in acoustical sense. The

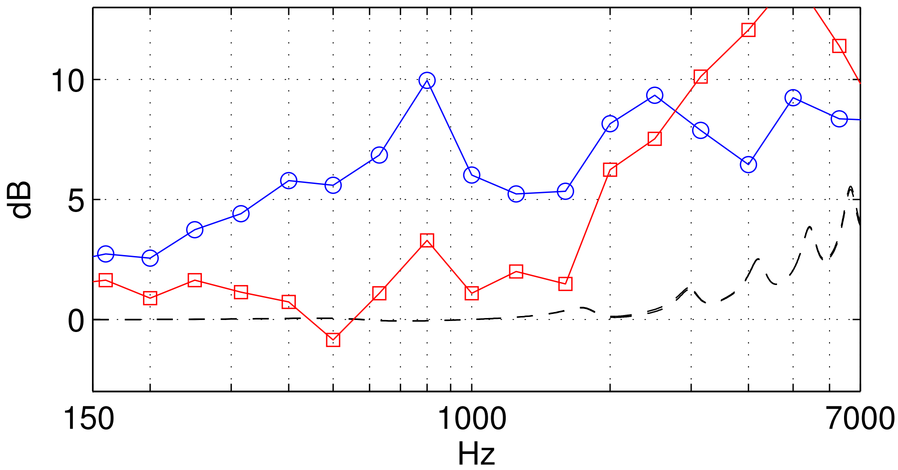


Figure 8: *Difference of averaged test subjects and HATS. The comparison is done for transfer functions from microphone position 3.4 to 4.1 and 4.2 (○ and □). A positive dB value implies that the HATS is more directional on that frequency. Two corresponding models are included where the mouth aperture radii for corresponding HATS and averaged person in the baffled models are 1.5 cm and 0.5 cm. The positions in the models are in a 1 cm distance from the baffle.*

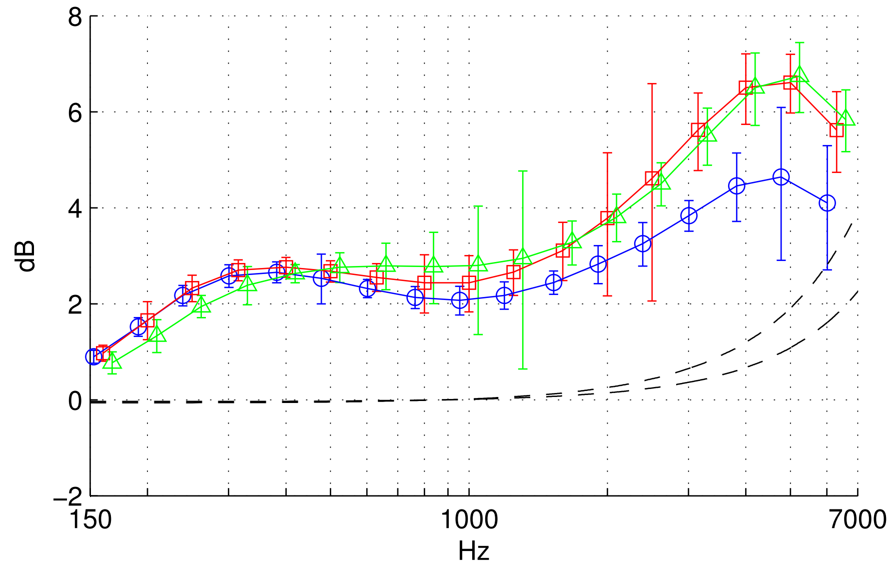


Figure 9: *Difference of averaged test subjects and HATS between different vowels. The comparison is done for transfer functions from microphone position 3.4 to 1.3 for open (○), close-mid (□), and close (△) articulated within the sentence. Two modeled cases are included where the mouth aperture radii for corresponding HATS and averaged person in the direct models are 1.5 cm and 0.5 cm and for downmost curve 1.5 cm and 1.5 cm. 95% confidential intervals can be seen for each frequency band. Curves are smoothed in in frequency.*

physical absolute size is difficult to directly measure or yield from the acoustical measurements.

There are several factors that cause variance in speech data: the test subject, phoneme, speech volume, etc. Same of these factors should directly be linked to the mouth aperture size.

The recordings were labeled so that the data could be divided into phoneme groups and the transfer function estimates for the phone groups could be obtained. In Figure 9 is shown differences between averaged test subjects and HATS by using three different phonemes as an input data. Basically if the mouth is larger the curve should be on high frequencies closer to the 0 dB axis. In the figure we see that the open vowels (a, ä) are more closer the directivity of the HATS than the close-mid (e, o, ö) and close (i, y, u) vowels [22].

5. IMPROVEMENTS PROPOSALS FOR TELEPHONOMETRY

5.1. Mouth aperture size in HATS

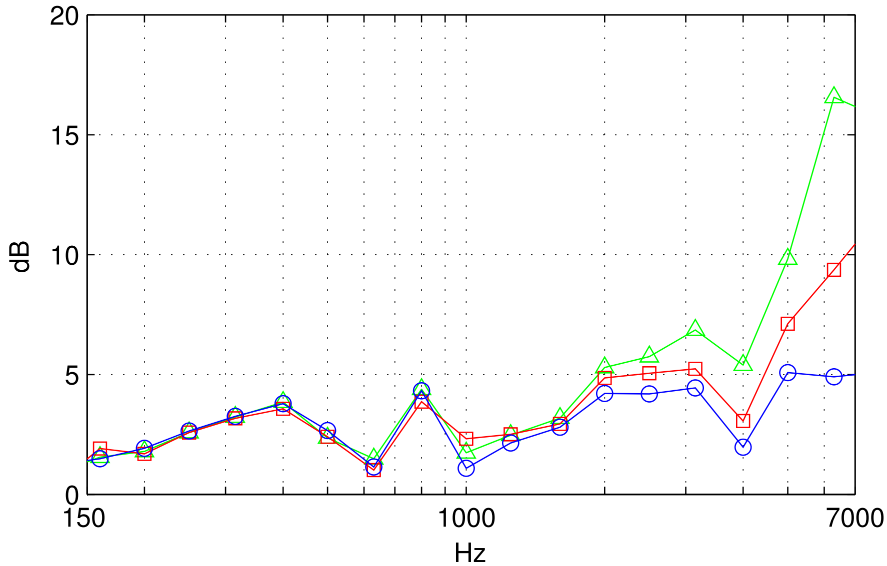


Figure 10: *Comparison between averaged human and B&K HATS 4128 in position 1.3 with three different mouth sizes: 30 mm × 11 mm (□), 15 mm × 11 mm (○), and 42 mm × 16 mm (△).*

We saw that the mouth aperture size is the most important parameter, which affects the directivity especially close to the mouth. On the other hand the measurements showed that the directivity of B&K HATS 4128 does not comply with the averaged test subjects. So could the difference be somehow corrected by tuning the mouth aperture size in HATS?

In Figure 10 the averaged test subject and HATS with three different mouth sizes are compared. We see that the reduced mouth size gives even with full wide bandwidth almost flat correspondence if the small fluctuation is omitted.

5.2. B&K specific measurement vest

A dedicated 2 cm thick vest (B&K model DS 0900) is recommended to be used on B&K HATS 4128 in acoustical measurements. The covering should reduce the reflections from the torso so that it resembles better human body. The official name for the vest "shoulder damping fabric" describes best its original purpose. The vest is especially designed for the binaural measurements with the artificial ears and the close-to-mouth telephonometry. Strong reflections from the shoulder are in those measurements an undesired side effect.

Although the correspondence of the torso in HATS to real human upper body is not known, it is widely used in headset measurements. The microphone positions 4.1 - 4.3 follow the same positions as the microphones in these phone accessories. By considering the measurement results for the position 4.1 and 4.2 in the Figure 11 we see significant differences between bare torso and the torso with vest.

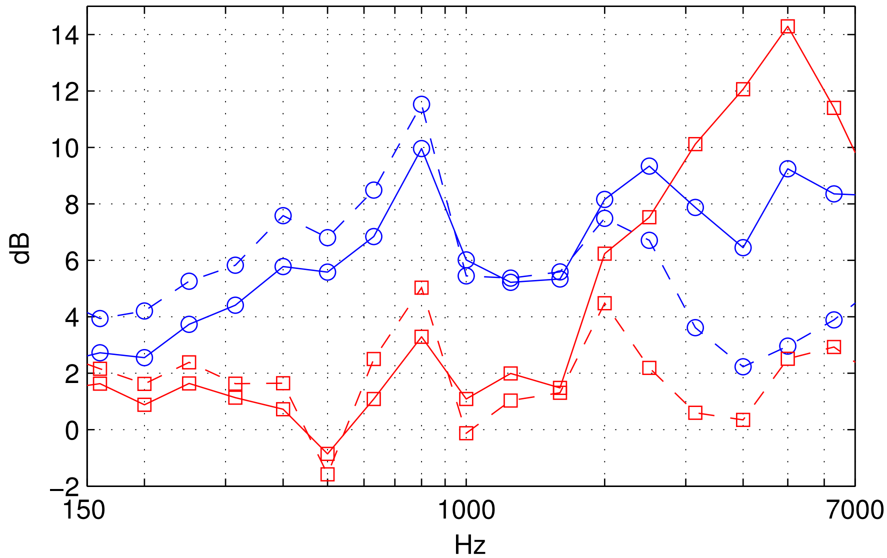


Figure 11: *Comparison of HATS to averaged test subject in positions 4.1 (○) and 4.2 (□) with (solid) and without (dashed) the vest on chest.*

If the general trend is considered in the curves (Figure 11) we see that the difference of the near chest reflection is omitted if the vest is taken off. The difference of the reference point still causes fluctuation near 800 Hz. The HATS is slightly more directional than average person in high frequencies because of the difference of the mouth aperture size.

The final conclusion is that the usage of the measurement vest should always be considered carefully. It seems that close to torso it should not be applied especially when the microphone rests freely on the chest as in several headsets.

5.3. Equalization consideration

When the directivities of the B&K HATS and the averaged test subject were compared significant differences were found (see Figure 7). The starting point for the study was the frequency response measurement for microphones in phones and headsets. If HATS does not correspond to the averaged directivity of a person some kind of compensation should be applied. Next a simple equalization scheme is introduced.

There were two areas that were covered in the measurements: positions near cheek and near chest. The measurements of the chest positions were heavily affected by the reflections from the chest and the difference seems to be substantially linked to the torso and its covering (see Figure 8). For these reasons it is meaningless to yield equalization scheme for these positions and therefore they are omitted in this approach.

The curves for the positions near cheek in Figure 7 have some fluctuation, which is mainly caused by the difference of the reference points in far field in other words the difference of the chest reflections between HATS and test subjects. The curves are therefore first smoothed and then used as the target responses for the equalization. The smoothing is implemented averaging neighbor frequency pins weighting with Hanning window. Window size was nine so 4 frequency pins were used from both sides for averaging. This window size corresponds to three octaves. The smoothed curves can be seen in Figure 12.

There are three possibilities to approach the equalization of the handset measurements. A starting point is that we have a set of curves of difference for certain positions and the differences are to be compensated with an equalization curve or curves. The different approaches are listed below.

1. Average frequency correction curve.

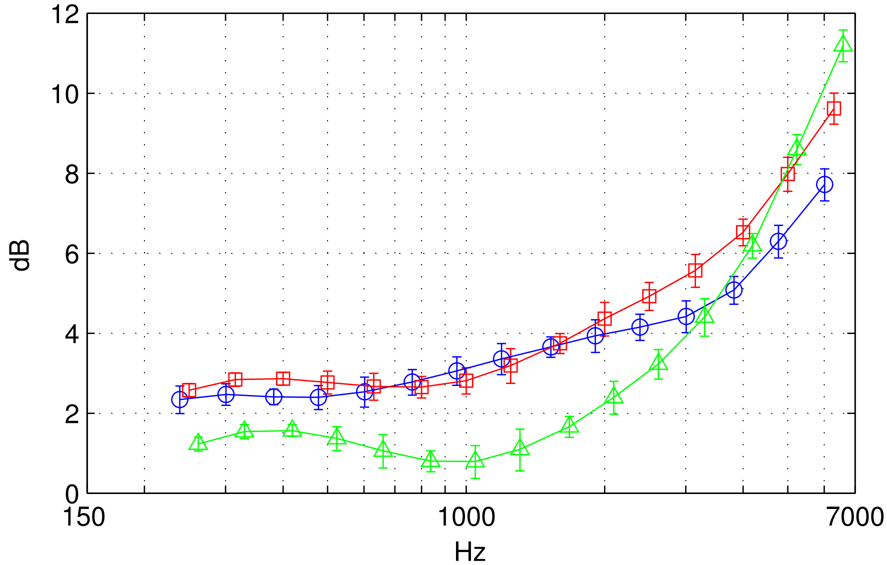


Figure 12: *Smoothed differences of HATS and averaged test subjects from Figure 7 between microphone position 3.4 and positions: 1.1, 1.3, and 1.4 (○, □, and △). 95% confidential intervals are included.*

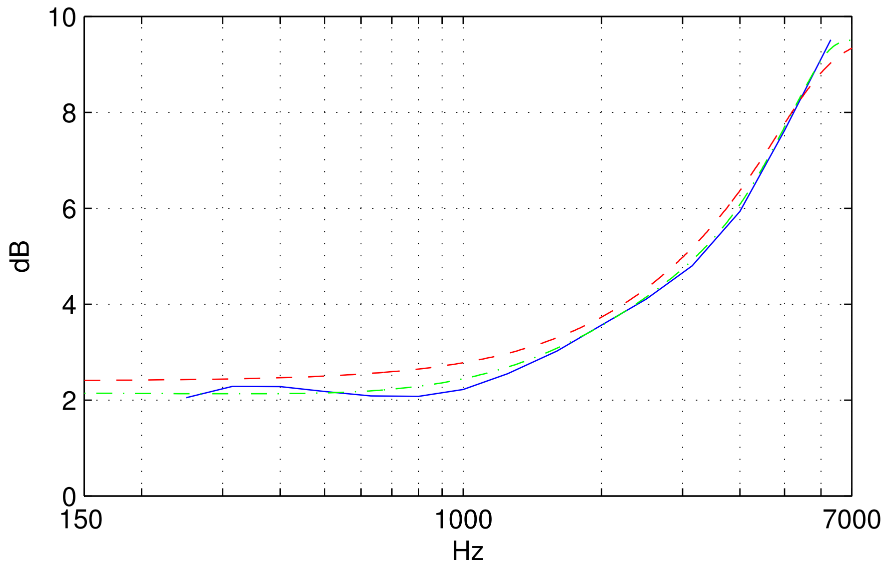


Figure 13: *Target curve for the average equalization filter (solid blue). Frequency responses for the filters where order is 3 (dashed red) and 7 (dash dotted green). Sample rate was 32kHz.*

2. Separate frequency correction curves for each position.
3. A model based position dependent frequency correction.

A filter design scheme is next presented. The motivation is to see what kind of filter design fits to the target curves discussed above. Only the average of the curves in Figure 12 is covered as an example.

The averaged target curve for the equalization can be seen in Figure 13. As we can see the curve has no dips or peeks so it can be modeled with a low order digital filter. An infinite impulse response (IIR) filter structure was selected for this consideration. It follows more accurately the target curve on same order than finite impulse response (FIR) filter.

In Matlab the IIR filters were recursively designed using a least-squares method (Yule-Walk function in signal processing toolbox). The target curve and two IIR filter responses are shown in Figure 13. The order three was the lowest design that stayed within 1dB from the target response.

6. CONCLUSIONS

The objective of the study was to assess the directivity of the artificial mouths from telephonometry point of view. The B&K HATS 4128 is widely used in telecom industry so the study concentrated on it. The sound field around the head and upper body was studied to see the basic characteristics. The most important part of the study was to see how accurately the directivity of the HATS would correspond to an averaged person. The speech content was also considered in a sense if it had an effect to the directivity.

In positions near the cheek the attenuation of high frequencies was found. If the position was closer to the ear there were more attenuation on high frequencies. Near the chest the reflection from the chest causes fluctuation but the high frequency attenuation is not that visible as near the cheek. In far field the chest reflection originated fluctuation is seen.

The speech content as well as the test subjects caused significant amount of variation to the directivity pattern. Nevertheless it was found that the directivity pattern could be linked to the speech content if the mouth aperture size during the articulation is known. For example comparing vowels and different kind of articulation the effect of the mouth aperture size was seen. The body size differences were omitted in the consideration because they self-evidently affect the sound field.

The test subject data was averaged and compared to HATS in the same positions. Significant differences were found in high frequencies. It seems that the mouth size difference between HATS and on average for test subjects is the key to the difference. The mouth aperture size of the speakers was during the articulation less than the size of the mouth in B&K 4128 HATS. Near the chest also the physical dimensions and features of the upper body caused differences in directivity.

Three improvement proposals were introduced in Section 5 to enhance the correspondence of the HATS measurements to an averaged person. An equalization scheme for the handset measurements was discussed. There are several concepts to implement a compensation for the directivity error. In this study it is shown that a low order IIR filter meets the equalization requirements. Therefore the equalization scheme could be easily implemented to telephonometry.

The measurement vest usage in the measurements was also questioned in a case where the microphone is laying on the chest. This discussion is focused on the measurement of so-called headset accessories of mobile phones. The measurement vest shifts the position of distance by its thickness from the hard surface of the torso of HATS. This distance causes a delay between direct sound and reflection from the torso and the outcome of it does not comply with an averaged person. The bare torso seems to be a better simulation for the positions on the chest.

Because B&K HATS 4128 was measured with three different mouth sizes it was easy to see what mouth size gives best correspondence to averaged test subject. It was found that if half the width of the mouth was blocked the difference curves were almost flat on wide band.

As we saw there are open questions on the field of the directivity of artificial mouths. The product that was studied in this paper did not correspond to the averaged test subjects. The future work should concentrate on improvement of the products as well as improvement of the standardization. Some temporary solutions to improve the telephonometry were introduced.

7. REFERENCES

[1] Brüel & Kjær, DK-2850 Nærum, Denmark, *Product data: Head and Torso Simulator - Type 4128 C*.

[2] Bruce A. Bartlett, "Tonal effects of close microphone placement," *J. of the Audio Engineering Society*, vol. 29, no. 10, Oct. 1981.

[3] Eddy B. Brixen, "Spectral degradation of speech captured by miniature microphones mounted on persons' head and chest," in *AES 100th Convention*, May 1996.

[4] Eddy B. Brixen, "Near field registration of the human voice: Spectral changes due to positions," in *AES 104th Convention*, May 1998.

[5] H. K. Dunn and D. W. Farnsworth, "Exploration of pressure field around the human head during speech," *J. of the Acoustical Society of America*, vol. 10, pp. 184–199, Jan. 1939.

[6] James L. Flanagan, "Analog measurements of sound radiation from the mouth," *J. of the Acoustical Society of America*, vol. 32, no. 12, pp. 1613–1620, Dec. 1960.

[7] T. Leino and A-M. Laukkanen, "Äänietäisyyden vaikutus puheäänen keskiarvospektriin," in *Papers from the 17th Meeting of Finnish Phoneticians*, Helsinki, 1992, pp. 117–129, Department of Phonetics, University of Helsinki.

[8] F. S. McKendree, "Directivity indices of human talkers in English speech," in *Proceedings - 1986 International Conference on Noise Control Engineering*. Noise Control Foundation, Jul. 1986, pp. 911–916.

[9] K. Sugiyama and H. Irii, "Comparison of the sound pressure radiation from a prolate spheroid and the human mouth," *Acustica*, vol. 73, pp. 271–276, 1991.

[10] Jyri Huopaniemi, Kaisa Kettunen, and Jussi Rahkonen, "Measurements and modeling techniques for directional sound radiation from the mouth," in *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1999, pp. 183–186.

[11] ITU-T, "Artificial mouth," Series P: Telephone Transmission Quality, Objective Measuring Apparatus P.51, International Telecommunication Union, 1996.

[12] ITU-T, "Head and torso simulator for telephonometry," Series P: Telephone Transmission Quality, Objective Measuring Apparatus P.58, International Telecommunication Union, 1996.

[13] Eero Lampio, "Pallokeinopään kehittäminen ja sen käyttö puhelimen akustisissa mittauksissa," M.S. thesis, Teknillinen Korkeakoulu, 1961.

[14] Philip M. Morse and Uno K. Ingard, *Theoretical Acoustics*, Princeton University Press, Princeton, New Jersey, 1968.

[15] Earl G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, 1999.

[16] Julius S. Bendat, *Random Data Analysis and Measurement Procedures*, John Wiley & Sons, 2nd edition, 1986.

[17] Julius S. Bendat and Allan G. Allan, *Engineering Applications of Correlation and Spectral Analysis*, John Wiley & Sons, 1980.

[18] J. Dang, K. Honda, and H. Suzuki, "Morphological and acoustical analysis of the nasal and the paranasal cavities," *J. of the Acoustical Society of America*, vol. 96, pp. 2088–2100, 1994.

[19] T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J. of the Acoustical Society of America*, vol. 90, pp. 799–828, 1991.

[20] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. of the Acoustical Society of America*, vol. 100, pp. 537–554, 1996.

[21] Gunnar Fant, *Acoustic theory of speech production with calculations based on X-ray studies of Russian articulations*, Description and analysis of contemporary standard Russian ; 2. Mouton, The Hague Paris, 2nd edition, 1970.

[22] The International Phonetic Association, "The international phonetic alphabet," Revised to 1993, Updated 1996, http://www.arts.gla.ac.uk/ipa/ipa.html.