Directivity of Artificial and Human Speech*

TEEMU HALKOSAARI,[1] MARKUS VAALGAMAA,[2] *AES Member,* **AND MATTI KARJALAINEN,[1]** *AES Fellow*
(teemu.halkosaari@hut.fi) (markus.vaalgamaa@nokia.com) (matti.karjalainen@hut.fi)

[1]*Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 02150 Espoo, Finland*

[2]*Nokia Technology, Audio Entity, 00180 Helsinki, Finland*

A study of how the directivity characteristics of artificial mouths correspond to the directivity of a real speaker is presented. The primary motivation for the research was the measurement methods applied in the telecommunications industry for the microphones used in telephones and their accessories. The responses of three artificial mouth simulators were measured in several positions. The same measurements were repeated for a group of test subjects. The measurement positions corresponded to the same positions where the microphones of telephones and their accessories, so-called headsets, would be situated. The basic mechanisms that produce the directivity patterns are discussed, and the contribution of the speech content is shown. The main contributor to the directivity is the aperture size of the mouth. The acoustical characteristics of the upper body are also a significant factor if the microphone position is not directly in front of the mouth. A greater than 10-dB difference with wide-band speech was found between artificial mouths and test subjects. It appears that the directivities of the artificial mouths are too narrow at high frequencies. To improve the correspondence of telephonometry and real speakers, a simple equalization procedure and two structural improvements are proposed.

0 INTRODUCTION

The directivity of the sound radiating from the human mouth is of interest in all communications by voice. In speech and audio technology, such as telephone conversations or recordings of singing, the position of the microphone affects the timbre of the sound.

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

A wider use of accessories such as headset equipment set also new requirements for telephonometry. In these accessory gadgets the microphones lie on the chest, hang on a wire, or are situated somewhere else, not in a traditional telephone microphone position.

Generally speaking, nowadays the directivity of artificial mouths should cover accurately a larger set of positions than for which they were originally designed. Nevertheless the frequency responses and the sensitivity characteristics of current artificial mouths used in acoustical measurements in off-axis positions are not well known.

There is no extensive literature available on the subject. A few essential classical papers discuss the basic aspects of the directivity of the mouth. In these studies a very large number of measurements had been conducted with test subjects. These studies give comprehensive information on the radiation pattern in the far field, typically at a distance of more than 500 mm from the mouth, and also in a couple of positions in the near field [1]–[3].

More detailed and specialized approaches to the subject can be found in a couple of sources. These papers cover typically a large set of measurements and modeling of the mouth and head system. Sugiyama and Irii present both a model-based approach and a wide amount of measurements [4]. Both far-field and near-field positions are measured for a group of test subjects. The measurement positions as well as the reference point are partly the same as in the present study. Brixen wrote two papers that concentrate only on direct measurements [5], [6]. Here a number of measurements were conducted with a group of test subjects using several measurement points around the head.

The motivation behind [7], [8] was to improve our ability to make predictions of the speech transmission index

*Manuscript received 2004 August 24; revised 2005 May 15.

and the articulation index. In [7] McKendree conducted the measurements only with a group of test subjects, and in [8] Bozzoli and Farina repeated the measurements with two artificial mouths. The measurement positions in the studies were on a horizontal plane around the head at about 1-m distance from the subject. The same type of measurements for the head and torso simulator (HATS) were published by Huopaniemi et al. [9], where the scope was to evaluate the reciprocity theorem with the artificial mouth. A more musician-oriented approach to directivity can be found in Bartlett [10], where the miking of several musical instruments and speech is discussed.

Nevertheless, in any of these studies the directivities of artificial mouths are not compared systematically to those of test subjects using similar measurements. Thus there is clearly a need for measurement-based information on how accurately artificial mouths simulate the real human mouth and whether their design should be improved.

There are standards that define guidelines for the directivity patterns of artificial mouths. The most important ones are ITU-T P.58, "Head and Torso Simulator for Telephonometry" [11], and ITU-T P.51, "Artificial Mouth" [12]. The problem with these standards is that they do not define specifically the directivity patterns in the near field for the positions of the microphones of headsets and small phones.

In the currently used narrow-hand telephone bandwidth (300–3400 Hz) the directional features are not yet considerably prominent. In wide-band speech (150 Hz to 7 kHz) the directional features are more prominent, as we will see later. Although the artificial mouths, especially the so-called HATS, were designed for narrow-band applications, they will most probably be used for wide-band measurements in the future.

This study concentrates on the directivity characteristics of both human and artificial mouths. The comparison is mainly concerned with human and Brüel & Kjær (B&K) HATS 4128 artificial mouths. Two other artificial mouths were also studied: Head Acoustic (HA) HMS II.3 and B&K 4227. The most important part of the study was to determine whether there is a difference between the directivities of simulators and an average person. As a result the study should give basic information on how the artificial mouth measurements for phone handsets should be designed and how the potential difference between the directivities could be compensated in the results.

The behavior of mouth directivity is also studied in general. The far- and near-field conditions are discussed. The sound radiation pattern affected by the head and torso is studied by both measurements and modeling. Also the characteristics of the speech content as it affects the directivity are discussed.

The structure of this Engineering Report focuses on the measurements and their outcomes. Two simple models for a mouth as a sound source are presented in the next section. The details of the measurements are listed in Section 2, with the results shown in Section 3. Section 4 discusses proposals for improvements with regard to telephonometry and Section 5 concludes by describing the outcome of the study.

1 THEORY AND MODELING

The directivity of the human voice production system can be approximated by various kinds of models. Here we present some simple models, which subsequently will be compared to measurements. Also the directivity characteristics can be predicted by altering the parameters in the models.

One of the simplest models for the head is a sphere. The mouth is approximated by a round piston in the sphere. In this way rotationally symmetrical coordinates can be obtained and therefore there are just two axes: the radius r from the center of the sphere and the angle θ from the axis. The model is illustrated in Fig. 1.

The particle velocity u_r on the surface of the sphere is given by

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

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

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

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

where P_n is the Legendre polynomial, $h_n^{(2)}$ the Hankel function, k the wavenumber, ρ_0 the density of air, and c_0 the velocity of sound. The coefficients κ_n are dependent on the displacement pattern on the sphere [13, 14]. In this case they are

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

This model gives a simplified view of how the head interacts with the radiation from the mouth. Self-evidently the whole body affects the sound field. In particular the shoulders and chest diffract, reflect, and absorb the sound.

If we consider the sound field close to a line perpendicular to the lip plane and positions on the chest, the most significant effect of the body is direct reflection. It causes a comb-filter-shaped response to the field. One direct and simple way to model this reflection is shown in Fig. 2, where the space is divided by an infinite baffle. The head model itself is the same as before.

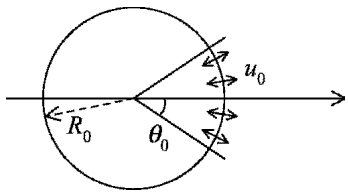


Fig. 1. Piston and sphere model.

The main advantage in this baffle model is that it can be formulated using the mirror source method. The total field can be expressed by

$$p_{\text{baffled}}(r,\theta) = p(r,\theta) + p(r, -\theta). \quad (4)$$

The models are used to assess the measurement results, and so the parameters of the models have to be selected using some reasonable principles. These parameters are the radius of the head R_0 and the mouth aperture angle θ_0. The ITU standard [11] specifics the dimensions of the head and torso simulators. 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, a value of 100 mm was used, and the mouth aperture angle was derived from the product specification [15], where the mouth cross sectional area is 300 mm^2. Depending on the case, other aperture sizes were also used in the models.

2 MEASUREMENTS

The measurements had several objectives. First the directivities of the artificial mouths were to be determined for several positions using transfer functions 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.

To ensure easy completion of the measurements, a multichannel recording system was designed for the test subjects. The same system was used with minimum changes throughout all the measurements. The artificial mouths were measured using a two-channel impulse response measurement system.

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

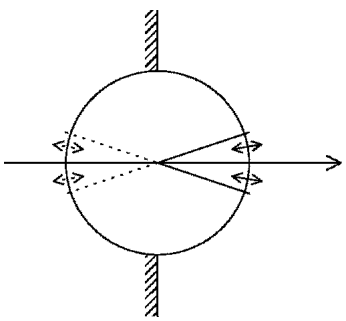


Fig. 2. Mirror source method. Infinite baffle is added to model.

2.1 Equipment

The artificial mouths were measured in an anechoic chamber in the Nokia facilities in Salo, Finland. The chamber ensures free-field conditions down to 90 Hz. The B&K HATS was the main target in the study because it is the most widely used device in telecommunication measurements [15].

The maximum-length-sequence (MLS) method on the Audio Precision measurement platform was applied to generate the impulse responses from the artificial mouth to the microphones [16]. The measurement itself consisted of a set of sequential cases with the resulting impulse responses for two channels. A sampling rate of 32 kHz and a 32 767-sample MLS were the parameters for the measurements. Two B&K pressure-type condenser microphones 1/4 in in diameter, were used.

The directivity of the test subjects had to be studied by recording their speech. Measurements took place in the anechoic chamber at the Helsinki University of Technology. The data acquisition system was built up from an IOTech WaveBook/516, 16-bit 1-MHz data acquisition system. It has eight input channels. One channel was reserved as the reference. Thus the system permitted directivity measurements by seven transfer function estimates in parallel. The system has 16-bit analog-to-digital conversion that is shared among the eight channels. The sampling rate of 32 kHz was selected for these measurements as well.

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

The signal levels were monitored and amplified microphone preamplifiers (E.A.A. professional stereo preamplifier, PSP-2) so that an adequate signal-to-noise ratio (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. Both a video link and an audio link were set up between the test subject inside the chamber and the operator.

2.2 Target Positions

The measurement positions are illustrated in Fig. 3 and defined in Table 1. The B&K HATS 4128 was measured in all positions in parallel with the mouth reference position (MRP), that is, position 2.1. The B&K 4227 and the HA HMS II.3 HATS were measured in positions next to the cheek (1.1 to 1.4) and in front of the mouth (3.1 and 3.3).

The recording system for the test subjects had eight channels. Therefore the same eight positions were used throughout the measurements: positions near the cheek (1.1 to 1.4), on the chest (3.1 to 3.3), and as a reference (2.2). The far-field reference was chosen as positioning a microphone at MRP proved difficult for the test subjects because of the noise caused by the airflow from the mouth.

A sophisticated measurement helmet was designed and built to permit attaching the measurement microphones as accurately as possible to the specified positions for the test subject measurements. Also the microphones close to the chest were attached to each other in an array. The positioning of the equipment is shown in Fig. 4.

The microphone positioning was referred to the lip plane and the line perpendicular to it containing the center of the lip ring. This is adopted directly from the ITU standards [11], [12]. On the chest the positions lie freely, and therefore the axes run from the throat downward and sideways (see Fig. 3).

2.3 Measurement Procedures

The B&K HATS 4128 was measured three times in every position: with and without a measurement vest (B&K 0600 shoulder damping fabrique) and without torso. The output for each measurement case was two 32 767-sample impulse responses at a 32-kHz sampling rate for the reference point (MRP) and the point of interest. In addition some further measurements were conducted. In two positions near the cheek the measurement was repeated with different mouth sizes: normal (30 × 11 mm), blocked to half its width (15 × 11 mm), and without the mouth adaptor (42 × 16 mm). The other artificial mouths were measured in a similar way as the B&K HATS, but without vest or mouth size modifications and with a smaller set of measurement positions.

A sentence in Finnish was selected for the 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 siella padallisen fasaania banaanilla höystettynä" (Two years ago we went to the restaurant Gabriel in Helsinki and we ate there a pot of pheasant larded with banana). This was recorded using the eight microphones simultaneously. In this sentence there is at least one occurrence of all the vowels and consonants of the Finnish language. There are altogether 95 separate short or long phonemes.

The measurement session was repeated for 13 test subjects, five female and eight males. The ages of the subjects varied between 20 and 30 years. The group consisted of different body sizes, from 1.60 to 1.90 m tall. The test subjects were sitting on a small chair and were wearing casual summer clothes. None of the test subjects had any speech defect, and all had Finnish as their mother tongue.

The test subjects were instructed to speak clearly and stay still during the recording. A head rest was attached to the chair and two microphone stands were aligned in front of the speaker so that it would be easy to keep the head still. The recordings were always repeated when a mistake

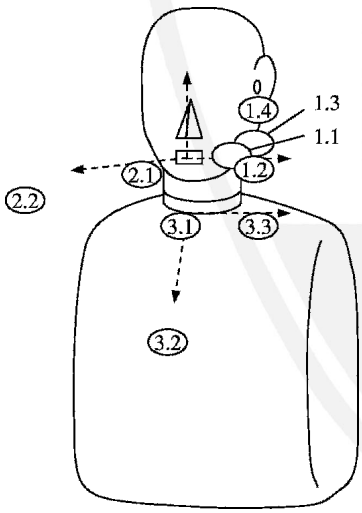


Fig. 3. Measurement positions.

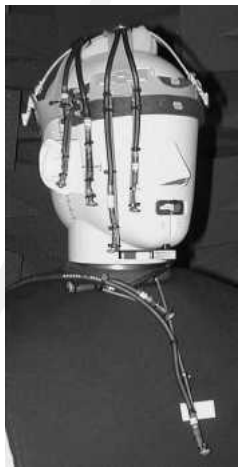


Fig. 4. Measurement helmet and chest microphone array on B&K HATS 4128. Microphones (Sennheiser DE 4-211-2) are in positions used throughout test subject measurements.

Table 1. Measurement positions defined by coordinates in mm.*

Position	On Axis	Sideways	Upward	Downward	Description
1.1	0	60	0		Large phone, normal angle
1.2	−10	70	−40		Large phone, lower angle
1.3	−30	85	10		Small phone, normal angle
1.4	−70	100	30		Boom headset
2.1	25	0	0		MRP
2.2	500	0	0		Far field, also reference
3.1		0		50	On chest, near throat
3.2		0		200	On chest, middle of chest
3.3		100		0	On chest, near shoulder

* The coordinate axes are referred to the lip plane and the line perpendicular to it (see Fig. 3). On the chest the positions are directly on the chest and the axes run downward and sideways.

was heard in the articulation or a movement was noticed in the video monitoring.

3 RESULTS

3.1 Analysis Overview

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

The data obtained from the artificial mouth measurements were direct impulse responses. Therefore the transfer functions were obtained by fast Fourier transform (FFT). By assuming that the system is linear and the reference positions are identical in every case, the transfer functions from point to point can be calculated by dividing the FFTs.

The transfer function estimates $T_{xy}(f)$ were calculated from the test subject recordings by windowing the active signal in the recordings in 1024-sample frames (32 ms at a 32-kHz sampling rate) with a 50% overlap. The transfer function estimate is defined as

$$T_{xy}(f) = \frac{G_{xy}(f)}{G_{xx}(f)} = \frac{G_{yy}(f)}{G_{yx}(f)} \tag{5}$$

where $G_{xy}(f)$ is the cross-spectrum or autospectrum estimate if it is calculated for the signal itself. The variables x and y refer to the measurement positions.

The delay difference between the channels was compensated for. The delay was estimated by calculating the cross correlation between channels. The FFT was applied to each frame, and by averaging the frames a transfer function estimate was obtained [17]. The transfer functions were averaged to one-third-octave resolution.

3.1.1 Reliability Considerations

The simplest way to consider the reliability of the measurement data is to examine the coherence and the SNR. Both SNR and coherence are obtained by cross-spectrum and autospectrum estimates [17], [18]. The noise sample was taken from the end of each recording, starting from the end of the last segmented active speech. The artificial mouth measurements can be omitted in this consideration because in all cases the reliability of those measurements was far better than that of the test subject measurements.

SNR and coherence were adequate in all frequency bands that were considered. The worst SNR in all channels was better than 25 dB over the wide band. The coherence between the reference position and other positions was better than 0.8 on narrow band and better than 0.65 on wide hand. The values imply that the transfer function estimates are reliable on wide band and very reliable on narrow band [17], [18].

If the results from a large set of data are averaged, the confidence intervals are one way to see by statistical means how reliable the results are. The confidence intervals are included (see Figs. 14, 16, and 17 for whole data and Fig. 12 for each vowel group). The distribution is calculated from the nonoverlapping frames. The intervals stay mostly within a ±1-dB range. This implies that the results are representative [17].

3.2 Directivity Features of the Mouth

The directivity of the mouth is approached by considering which are the key aspects that play the most important role. Measurement data for artificial mouths are more reproducible than human data and the B&K HATS measurements are used for the consideration.

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

1) Mouth cross-sectional area
2) Distance from mouth—far or near field
3) Reflections from body.

Some explanations of these items follow.

The cross-sectional area and the shape of the mouth aperture determine how much the bare mouth directs. If the size of the mouth is not small compared to the wavelength, the mouth starts to be directive. At low frequencies the mouth can be considered a simple omnidirectional point source.

Second the near field and the far field differ depending on the dimensions of the head. The study concentrates only on the field perpendicular to the lip plane. In the far field at high frequencies, the head enhances the directivity, the same way as the mouth, if attached to an infinite wall. At low frequencies the head has no effect. On the other hand, if the position is shifted off the perpendicular, such as behind the head, the head shadows the sound at high frequencies.

Third, reflections and diffraction from the upper body modify the sound field considerably. Near the mouth, for example close to the cheek, the direct sound from the mouth is so dominant that the influence of the upper body can be disregarded. On the other hand, in the far field, and especially near the chest, reflections from the body cause significant fluctuations to the field.

Generally speaking the directivity of the mouth is seen as a reduction in the level of high frequencies when the measurement position is shifted toward the ear (see Fig. 5).

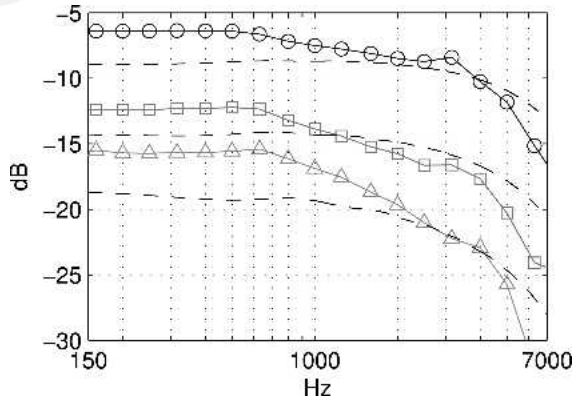


Fig. 5. Transfer functions of B&K HATS from position 2.1 (MRP) to positions 1.1 (○), 1.3 (□), and 1.4 (△). Modeled cases for same positions are included (- - -).

The same kind of effect is seen when the position moves down the chest (see Fig. 6). The reduction of high frequencies is more or less monotonic, and when the position is closer to the ear, the rolloff is steeper.

The HATS 4128 can be worn with a dedicated 20-mm thick measurement vest. The vest is intended to enhance the correspondence with the human upper body in binaural recordings. Applying the vest when taking measurements of the telephone and certain accessories is not so straightforward.

Figs. 6 and 7 show that the chest reflections cause peaks and dips in the responses. Close to 1 kHz there is a strong dip in the transfer functions for the position at 0.5-m distance (Fig. 7). The dip around 1 kHz implies that there is about a 15-mm difference between the paths 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 clip is at a lower frequency. If the vest is taken off, the dip shifts to a higher frequency because the reflection arrives faster, that is, the sound velocity in the vest is lower. In the curve for the bare-head model we see only a 3-dB level difference between high and low frequencies.

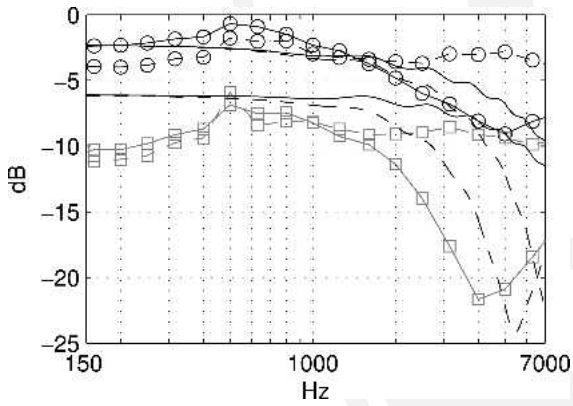


Fig. 6. Transfer functions of B&K HATS from position 2.1 (MRP) to positions 3.1 (○) and 3.2 (□) and same measurements without vest (---). Corresponding modeled cases with infinite baffle are included for the same positions. Both positions are modeled with two different distances from the infinite baffle: 10 mm (—) and 80 mm (-—-).

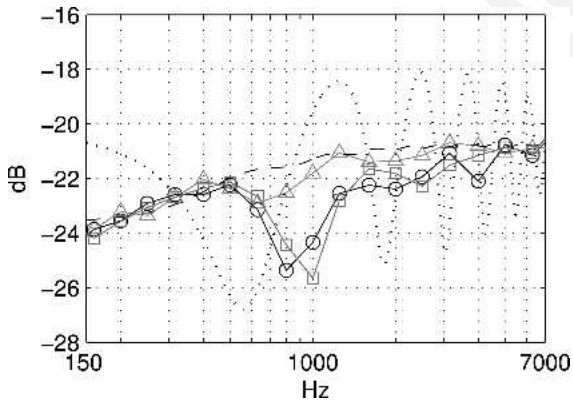


Fig. 7. Transfer functions of B&K HATS from position 2.1 (MRP) to positions 2.2 (0.5 m in front of mouth) with vest (○), without vest (□), and without torso (△). Two models, bare head (---) and head with infinite baffle (· · ·), are included using the same positions.

In Fig. 6 there is a dip near 5 kHz which is caused by a reflection through the vest. If the vest is taken off, the test position is directly on the hard surface and so only a duplication of the pressure field occurs. The dips in both Figs. 6 and 7 recur at higher frequencies in a comb-filtering shape, but the third frequency resolution makes it invisible.

3.3 Comparison of Artificial Mouths and Test Subjects

The differences between average test subjects and artificial mouths are shown in Figs. 8–11 for five test positions. The curves are calculated by dividing the transfer function for an averaged person by the transfer function for an artificial mouth to the same position. A positive value in the curve indicates that the artificial mouth is more directional than the human mouth, that is, there is a higher sound pressure level in the front as compared to the cheek or chest positions in the artificial mouth than in a human mouth. If the value is negative, the human mouth is more directional.[1]

In Figs. 8–11 fluctuations can be seen in the curves around 600–800 Hz and 2 kHz. These fluctuations are caused by the measurement setup, where the reference position was at a 500-mm distance in front of the mouth and the chest reflections create dips and boosts at slightly different frequencies for artificial and human measurements. Use of the MRP very close to the mouth would have attenuated this difference, but because of difficulties with airflow from the mouth and accurate positioning, MRP was not used during the measurements.

Differences between an average test subject and the B&K HATS near cheek positions are shown in Fig. 8. In

[1]In this context the term "directional" refers to the measured roll off of the high frequencies in off-axis positions.

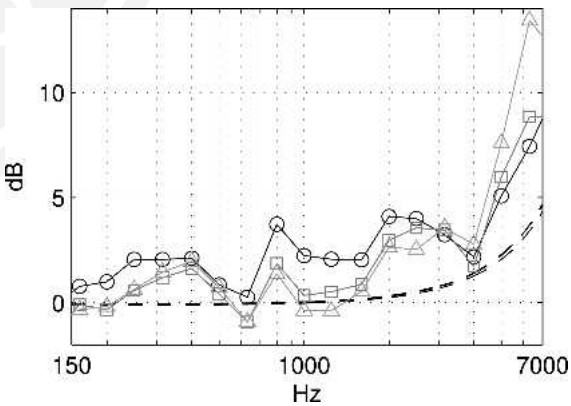


Fig. 8. Difference between average test subject and B&K HATS. Comparison is made for transfer functions from position 2.2 to positions 1.1 (○), 1.3 (□), and 1.4 (△). A positive dB value implies that the B&K HATS is more directional on that frequency. Three models (---) are included where B&K HATS and average person are compared in the same three positions. The bare-head model was applied for all models for B&K HATS and average person with the same head radius but with different mouth aperture radii (15 mm for HATS and 5 mm for average person).

the narrow frequency band the differences are between −1 and 4 dB, mostly so that the HATS is 2–3 dB more directional than an average subject. For frequencies above 4 kHz the differences increase; this indicates that the HATS is much more directive than the average human mouth. Above 5 kHz the difference is greater than −5 dB, and the maximum difference between the transfer functions of positions 2.2 and 1.4, that is, the position closes to the ear, is 14 dB. It is interesting to note that the general shapes of the difference curves near the cheek are very similar to each other.

The positions near the chest shown in Fig. 9 indicate that there are significance differences between test subjects and the B&K HATS. The main contributor to the difference is the chest reflection, which can be seen clearly around 800 Hz and 5 kHz. The HATS can be worn with a dedicated 20-mm-thick vest to reduce the shoulder reflections in binaural recordings (see Section 4.2). The results are shown with and without the vest. If the vest is used, the difference above 2 kHz grows considerably, especially between positions 2.2 and 3.2. While the difference above 2 kHz is smaller for the HATS without the vest, use of the vest below 2 kHz gives slightly more undisturbed curves.

The B&K 4227 is an artificial mouth in a small enclosure, without head or torso simulators. The lack of a torso can be seen in Fig. 10 to cause large differences between the mouths of test subjects and the B&K 4227, especially between 500 Hz and 1 kHz and around 2 kHz. In the narrow frequency band the differences are between −1 and 8 dB. At high frequencies the difference is not as noticeable as with the B&K HATS.

The third measured artificial mouth was the HA HATS, which consists of a head and shoulder part, but there is no lower part of the torso. Thus in the narrow frequency band the curves in Fig. 11 are very similar to those obtained with the B&K 4227. The differences are between −3 and 6 dB, and the longest differences occur around 800 Hz and 2 kHz. An interesting detail is a sharp dip around 4 kHz, which indicates that the sound level at that frequency created by the HA mouth is considerably lower in the near field off axis than in the front of the mouth.

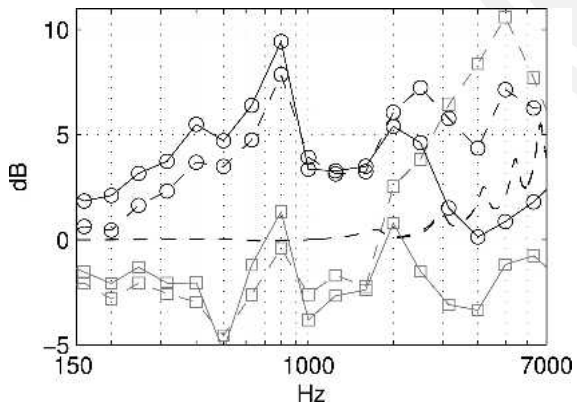


Fig. 9. Difference between average test subject and B&K HATS. Comparison is made for transfer functions from position 2.2 to postions 3.1 (○) and 3.2 (□) with (---) and without (——) vest on chest. Two corresponding models (---) are included where mouth aperture radii for B&K HATS and average person in baffled models are 15 mm and 5 mm. Positions in models are 10 mm from baffle.

3.4 Effect of Speech Content

An interesting trend can be seen especially in Fig. 8. The difference between the average test subject and the HATS increases toward high frequencies. Figs. 8 and 9 also include modeled curves, where the corresponding models for the HATS and an average person were compared. In modelling a larger mouth size was used in the HATS model. As can be seen, the general trend of the measurement results follows the modeled curves.

Since the aperture size of the mouth seems to be one key factor, the topic was analyzed further. One important aspect was to see whether the aperture size of the mouth of the during speech can be predicted using directivity features. Traditional mouth and vocal tract size measurements using imaging methods can, for example, be found in [19]–[22].

It is important to emphasize that the aperture size of the mouth is an effective relative value in the acoustical sense. The physical absolute size is difficult to measure directly or to obtain from the acoustical measurements.

There are several factors that cause a variance in speech data, such as test subject, phoneme, and speech volume.

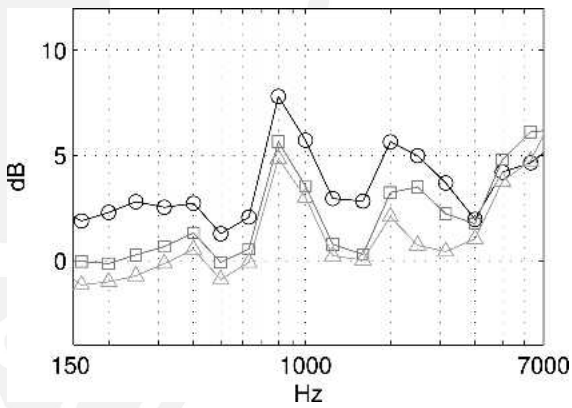


Fig. 10. Difference between average test subject and B&K 4227 artificial mouth. Comparison is made for transfer functions from position 2.2 to positions 1.1 (○), 1.3 (□), and 1.4 (△).

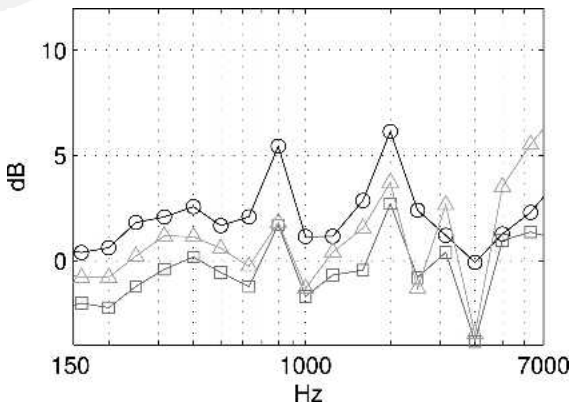


Fig. 11. Difference between average test subject and HA HMS II.3 HATS. Comparison is made for transfer functions from positions 2.2 to positions 1.1 (○), 1.3 (□), and 1.4 (△).

Some of these factors should be linked directly to the aperture size of the mouth.

The recordings of test subjects were labeled so that the data could be divided into phoneme groups, and the transfer function estimates for the phoneme groups could be obtained. Fig. 12 shows the differences between average test subjects and the B&K HATS when three different phonemes were used as input data. Theoretically if the human mouth is larger, the curve at high frequencies should be closer to the 0-dB axis. In the figure we see that the open vowels (/a/, /æ/) are closer to the directivity of the B&K HATS than the close-mid (/e/, /o/, /ø/) and close (/i/, /y/, /u/) vowels [23]. The reason is a larger opening of the human mouth with open vowels than with close-mid or close vowels.

4 IMPROVEMENT PROPOSALS FOR TELEPHONOMETRY

4.1 Mouth Aperture Size in B&K HATS

We have shown that the aperture size of the mouth is the most important parameter affecting directivity, especially close to the mouth. On the other hand the measurements showed that the directivity of the B&K HATS 4128 does not correspond to that of the average test subject. So could the difference be somehow corrected by fine tuning the aperture size of the mouth in the HATS?

Fig. 13 compares the average test subject and the B&K HATS for three different mouth sizes. We can see that even with a wide bandwidth, a reduced mouth size gives a relatively flat correspondence if the small fluctuations are disregarded.

4.2 B&K Specific Measurement Vest

A dedicated 20-mm-thick vest (B&K model DS 0900) can be used on the B&K HATS 4128 in acoustical measurements. The covering should reduce the reflections from the torso so that it better resembles the human body.

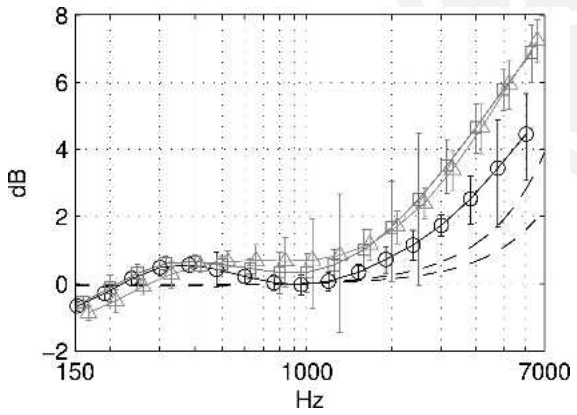


Fig. 12. Difference between average test subject and B&K HATS for various vowels. Comparison is made for transfer functions from position 2.2 to positions 1.3 for open (○), close-mid (□), and close (△) vowels articulated within a sentence. Two corresponding models (---) are included where mouth aperture radii for B&K HATS and average person in direct models are 15 mm and 5 mm and, for lowest curve, 15 mm and 10 mm. 95% confidential intervals can be seen for each frequency band. Curves are smoothed in the frequency domain.

The official name for the vest, "shoulder damping fabric," describes its original purpose best. The vest is specifically designed for binaural measurements with artificial ears and for close-to-mouth telephonometry. In those measurements strong reflections from the shoulders are an undesired side effect.

Although the correspondence between the torso in the B&K HATS and a real human upper body is not known, it is widely used in headset measurements. Microphone positions 3.1 to 3.3 hold the same positions as the microphones in these phone accessories. When considering the measurement results for positions 3.1 and 3.2 in Fig. 9, significant differences can be seen between the bare torso and the torso with vest.

If the general trend of the curves in Fig. 9 is observed, we see that the difference in the near chest reflection disappears if the vest is taken off. A difference in the reference point still causes fluctuations near 800 Hz. The B&K HATS is slightly more directional than an average person at high frequencies because of the difference in the aperture size of the mouth.

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

4.3 Considerations of Equalization

When the directivities of the artificial mouths and the average test subjects were compared, significant differences were found (see Figs. 8–11). The primary motivation for the study was the measurement of frequency responses for microphones in the telephones and headsets. If artificial mouths do not correspond to the average directivity of a person, some kind of compensation should be applied. Next a simple equalization scheme is introduced for the B&K HATS.

Two areas were covered in the B&K HATS measurements—the positions near the cheek and near the chest. The measurements of the chest positions were affected heavily by the reflections from the chest and the difference seems to be linked predominantly to the torso and its covering (see Fig. 9). For these reasons it is meaningless to

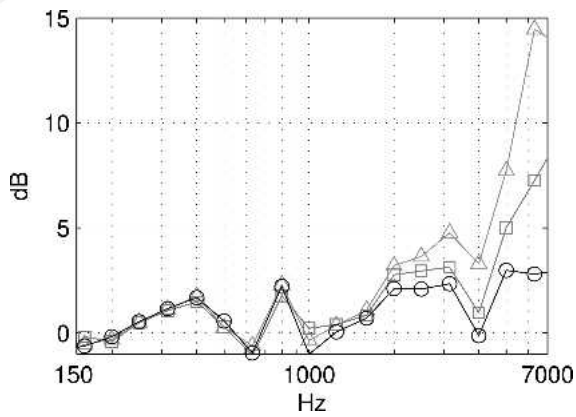


Fig. 13. Difference between average test subject and B&K HATS 4128 in position 1.3 with three different mouth sizes: 30 × 11 mm (□), 15 × 11 mm (○), and 42 × 16 mm (△).

present an equalization scheme for these positions, and therefore they are omitted in this approach.

The curves for the positions near the cheek in Fig. 8 show some fluctuations, which are caused mainly by the difference in the reference points in the far field, that is, the difference of the chest reflections between the B&K HATS and the test subjects. The curves are therefore first smoothed and then used as target responses for the equalization. The smoothing is implemented through averaging neighboring frequency bins by weighting them with a Hanning window. The window size was nine points, so that four frequency bins were used from both sides for averaging. This window size corresponds to three octaves. The smoothed curves for the B&K HATS are shown in Fig. 14.

There are three possibilities to approach the equalization of the handset measurements. The starting point is a set of difference curves 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
2) Separate frequency correction curves for each position
3) Position-dependent frequency correction.

A filter design scheme is presented next. The motivation is to see what kind of equalization filter design fits the target curves discussed. Only the average of the curves in Fig. 14 is used as an example.

The averaged target curve for the equalization is shown in Fig. 15. As we can see, the curve has no dips or peaks, 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 of the same order than a finite impulse response (FITR) filter.

The IIR filters were designed recursively in Matlab using a least-squares method (Yule–Walk function in the signal processing toolbox). The target curve and two IIR filter responses are shown in Fig. 15. The order of 3 was the lowest design that stayed within 1 dB from the target response.

The smoothed difference curves for the B&K 4227 and the HA HMS II.3 HATS are shown in Figs. 16 and 17. A high-frequency difference similar to that in the B&K HATS 4128 does not exist with other mouths, but differences of a few deciBel can still be seen.

5 DISCUSSION AND CONCLUSIONS

The objective of this study was to assess the directivity of artificial mouths versus human mouths from the telephonometry point of view. The B&K HATS 4128 is widely used in the telecom industry, so the study concen-

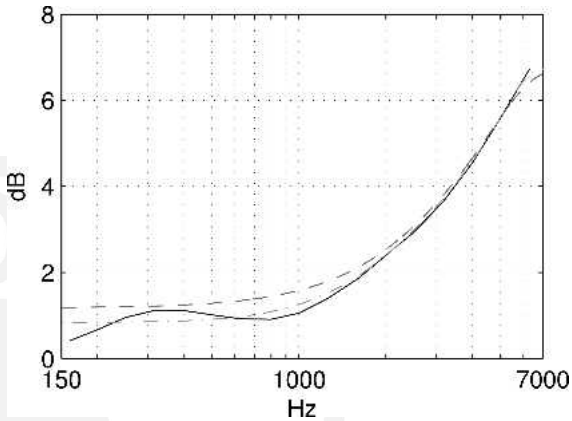


Fig. 15. Target curve for average equalization filter for B&K HATS 4128 (——). Frequency responses for filters were of orders 3 (---) and 7 (- · -); sampling rate was 32 kHz.

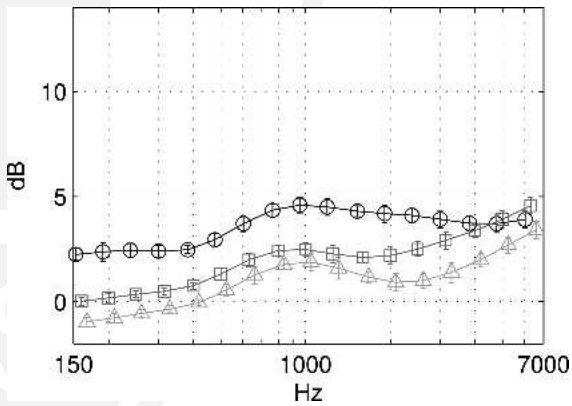


Fig. 16. Smoothed differences between B&K 4227 artificial mouth and average test subject for position 2.2 and positions 1.1 (○), 1.3 (□), and 1.4 (△). 95% confidence intervals are included.

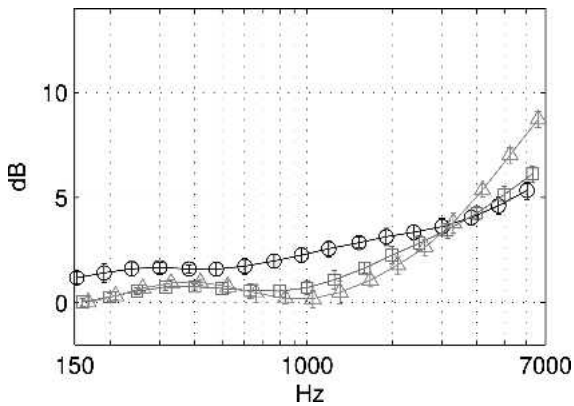


Fig. 14. Smoothed differences between B&K HATS 4128 and average test subject from Fig. 8 for position 2.2 and positions 1.1 (○), 1.3 (□), and 1.4 (△). 95% confidence intervals are included.

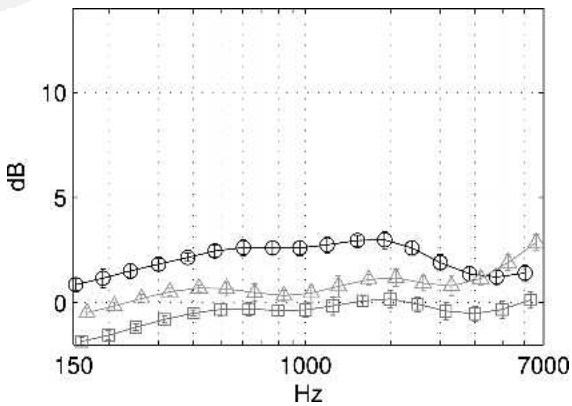


Fig. 17. Smoothed differences between HA HMS II.3 HATS and average test subject for position 2.2 and position 1.1 (○), 1.3 (□), and 1.4 (△). 95% confidence intervals are included.

trated mainly on it, but the B&K 4227 and the HA HMS II.3 HATS were also measured. The sound field around the head and upper body was studied to determine the basic characteristics. The most important part of the study was to see how accurately the directivity of the artificial mouths would correspond to an average person. The speech content was also considered to see whether it had an effect on the directivity.

In positions near the cheek an attenuation of the high frequencies was found. If the position is closer to the ear, the attenuation increases at high frequencies. Near the chest the reflection from the chest causes fluctuations in the frequency domain, but high frequencies are not as attenuated as near the cheek. In the far field comb-filter-type fluctuations due to chest reflections are clearly visible.

The speech content as well as the test subjects caused significant variations in the directivity pattern. Nevertheless it was found that the directivity pattern could be linked to the speech content if the aperture size of the mouth during the articulation is known. This was found by comparing the directivity patterns of different vowel groups.

The test subject data were averaged and compared to those for artificial mouths in the same positions. Significant differences were found in the comparison. The difference between the human mouth and the B&K 4227 or HA HATS near the cheek was on average between 1 and 6 dB, the HA HATS showing a slightly smaller difference compared to the average human.

However, the most significant difference between the average human and the B&K HATS near the cheek was found at high frequencies. It seems that the size difference in mouth between the B&K HATS and the average test subjects is the key to the difference. The aperture size of the mouth of a speaker during articulation is less than the size of the mouth of the B&K HATS. Near the chest the physical dimensions and features of the upper body also cause differences in directivity.

Three improvement proposals were introduced in Section 4 to enhance the correspondence of the artificial mouth measurements and those of an average person. An equalization scheme for the handset measurements was discussed. There are several concepts for implementing 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 for telephonometry.

The potential use of a measurement vest with the headset was also questioned in the case where the microphone is lying on the chest. The measurement vest shifts a position by its thickness from the hard surface of the torso of the B&K HATS. This distance causes a delay between direct sound and reflections from the torso, and the result does not correspond to that for an average person. The bare torso seems to represent a better simulation for positions on the chest.

Because the B&K HATS was measured with three different mouth sizes it was easy to see which mouth size corresponds best to an average test subject. It was found that if half the width of the mouth was blocked, the difference curves were almost flat on a wide band.

There remain open questions concerning the directivity of artificial mouths. The artificial mouths studied in this engineering report did not correspond fully to the mouth of the average test subject. Future work should concentrate on the improvement of the products as well as the standardization. Some temporary solutions to improve telephonometry were introduced.

6. REFERENCES

[1] H. K. Dunn and D. W. Farnsworth, "Exploration of Pressure Field around the Human Head during Speech," *J. Acoust. Soc. Am.,* vol. 10, pp. 184–199 (1939 Jan.).

[2] J. L. Flanagan, *Speech Analysis Synthesis and Perception,* 2nd expanded ed. (Springer, New York, 1972).

[3] J. L. Flanagan, "Analog Measurements of Sound Radiation from the Mouth," *J. Acoust. Soc. Am.,* vol. 32, pp. 1613–1620 (1960 Dec.).

[4] 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).

[5] E. B. Brixen, "Spectral Degradation of Speech Captured by Miniature Microphones Mounted on Persons' Heads and Chests," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 44, p. 641 (1996 June/July), preprint 4784.

[6] E. B. Brixen, "Near-Field Registration of the Human Voice: Spectral Changes Due to Positions," presented at the 104th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 46, p. 569 (1998 June), preprint 4728.

[7] F. S. McKendree, "Directivity Indices of Human Talkers in English Speech," in *Proc. 1986 Int. Conf. on Noise Control Engineering* (Noise Control Foundation, 1986 July), pp. 911–916.

[8] F. Bozzoli and A. Farina, "Directivity Balloons of Real and Artificial Mouth Simulators for Measurement of the Speech Transmission Index," presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 51, p. 1247 (2003 Dec.), convention paper 5953.

[9] J. Huopaniemi, K. Kettunen, and J. 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* (1999 Oct.), pp. 183–186.

[10] B. A. Bartlett, "Tonal Effects of Close Microphone Placement," *J. Audio Eng. Soc.* (*Engineering Reports*), vol. 29, pp. 726–738 (1981 Oct.).

[11] ITU-T P.58, "Head and Torso Simulator for Telephonometry," Series P: Telephone Transmission Quality, Objective Measuring Apparatus, International Telecommunications Union, Geneva, Switzerland (1996).

[12] ITU-T P.51, "Artificial Mouth," Series P: Telephone Transmission Quality, Objective Measuring Apparatus, International Telecommunications Union, Geneva, Switzerland (1996).

[13] P. M. Morse and U. K. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).

[14] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, San Diego, CA, 1999).

[15] Brüel & Kjær, "Product Data: Head and Torso Simulator—Type 4128," Nærum, Denmark (2001).

[16] Audio Precision, "High-Performance Testing with the Audio Precision 2700 Series" (2004), http://www.audioprecision.com/.

[17] J. S. Bendat, *Random Data Analysis and Measurement Procedures,* 2nd ed. (Wiley, New York, 1986).

[18] J. S. Bendat and A. G. Allan, *Engineering Applications of Correlation and Spectral Analysis* (Wiley, New York, 1980).

[19] J. Dang, K. Honda, and H. Suzuki, "Morphological and Acoustical Analysis of the Nasal and the Paranasal Cavities," *J. Acoust. Soc. Am.,* vol. 96, pp. 2088–2100 (1994).

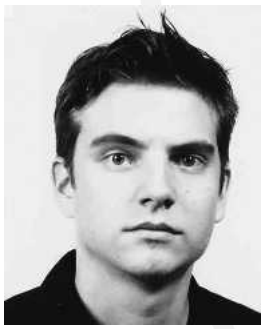
[20] T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye, "Analysis of Vocal Tract Shape and Dimensions Using Magnetic Resonance Imaging: Vowels," *J. Acoust. Soc. Am.,* vol. 90, pp. 799–828 (1991).

[21] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal Tract Area Functions from Magnetic Resonance Imaging," *J. Acoust. Soc. of Am.,* vol. 100, pp. 537–554 (1996).

[22] G. Fant, *Acoustic Theory of Speech Production with Calculations Based on X-Ray Studies of Russian Articulations—Description and Analysis of Contemporary Standard Russian,* 2nd ed. (Mouton, The Hague, Paris, 1970).

[23] International Phonetic Assoc., "The International Phonetic Alphabet," rev. to 1993; updated 1996. http://www.arts.gla.ac.uk/ipa/ipa.html/.

THE AUTHORS



T. Halkosaari

M. Vaalgamaa

M. Karjalainen

Teemu Halkosaari was born in Turku, Finland, in 1977. He received an M.Sc. degree in electrical engineering with a major in acoustics and audio signal processing from the Helsinki University of Technology (TKK), Finland, in 2004.

He has been working in the research area for this study since 2002 at TKK as well as at Nokia. At present he is with Iqua in Espoo, Finland, which produces high-quality wireless accessories for mobile phones, such as headsets, car kits, and portable hands-free devices. He is positioned as the project manager for an upcoming product, and in parallel his main task is to steer the audio and acoustics R&D in the company, including loudspeaker and microphone design, verification measurements, quality control, and related electronics and DSP issues in all products. His main interests in sound are audio measurements, audio reproduction, microphone and loudspeaker design, noise acoustics, and related hardware, which he has pursued since childhood. He has published several conference papers.

●

Markus Vaalgamaa has studied at the Helsinki University of Technology, Finland, majoring in acoustics and audio signal processing. He received an M.S. degree with distinction in 1999. His M.S. thesis "Moving Average Vector Quantization in Speech Coding" was done at the Nokia Research Center.

He worked in the Laboratory of Acoustics and Audio Signal Processing at HUT in 1998 and 1999 on the topics of audio and speech coding. He joined the audio research and technology team at Nokia as a research engineer in 2000. His work there encompasses audio quality and testing, psychoacoustics and audio DSP related research and development, the steering of several research projects and subcontracting projects and managing a team responsible for Nokia global audio quality and testing development. Currently he is a project manager and specialist.

His research interests are concerned with understanding human hearing and preferences, setting those to measurable audio requirements and audio DSP processing, especially, lately, dynamic range compressors and audio enhancement algorithms. He enjoys music in different flavors: listening to music, playing guitar, and singing in a band. He is a member of the Audio Engineering Society, a participant in technical committees on loudspeakers and headphones and on audio for telecommunication. He is also a member of the Acoustic Society of Finland.

●

Matti Karjalainen was born in Hankasalmi, Finland, in 1946. He received M.Sc. and Dr.Tech. degrees in electrical engineering from the Tampere University of Technology, Finland, in 1970 and 1978, respectively. Since 1980 he has been a professor in acoustics and audio signal processing at the Helsinki University of Technology, Finland, and is on the faculty of electrical and communications engineering.

In audio technology his interest is in audio signal processing, such as DSP for sound reproduction, perceptually based

signal processing, and music DSP and sound synthesis. In addition to audio DSP his research activities cover speech processing, perceptual auditory modeling and spatial hearing, DSP hardware, software, and programming environments, as well as various branches of acoustics, including musical acoustics and modeling of musical instruments. He has written 350 scientific or engineering papers and contributed to organizing several conferences and workshops, serving as the papers chair of the AES 16th Conference and technical chair of the AES 22nd Conference.

Dr. Karjalainen is a fellow of the Audio Engineering Society, and a member of the Institute of Electrical and Electronics Engineers, the Acoustical Society of America, the European Acoustics Association, the European Speech Communication Association, and several Finnish scientific and engineering societies.