HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Teemu Halkosaari

Radiation Directivity of Human and Artificial Speech

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, Apr 13, 2004

Supervisor:	Professor Matti Karjalainen
Instructors:	Markus Vaalgamaa

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Teemu Halkosaari		
Name of the thesis:	Radiation Directivity of Human and Artificial speech		
Date:	Apr 13, 2004	**Number of pages:**	57
Department:	Electrical and Communications Engineering		
Professorship:	S-89		
Supervisor:	Prof. Matti Karjalainen		
Instructors:	Markus Vaalgamaa, M.Sc.		

The directivity features of the artificial mouth are not well known on off-axis positions. In addition literature cannot be found for the correspondence of the artificial mouths to real speakers. The standardization (for example ITU) is more or less inadequate.

The frequency responses of the microphones in phones are measured with artificial mouths in telecommunication industry. When the size of the phones is getting smaller and at the same time the frequency bandwidth is extending, it is important that the directivities of the artificial mouths are correct or at least tested. Wearable mobile phone accessories, so called headsets, are also measured with the same methodology although their microphones are substantially further from the mouth.

The main scope of the study was to measure the directivity of an artificial mouth to a set of measurement positions and repeat the measurements for a group of test subjects. The positions were measured with measurement microphones but also with a few phones and headsets. The results were in practice a set of transfer functions referred to a fixed reference position. One additional goal was to see if the speech content has an influence on the directivity pattern.

For the measurements an eight-channel measurement system was set up in an anechoic chamber. About 15 test subjects and B&K HATS 4128 were measured. The measurement positions lay near the chest and the cheek. The transfer function estimates were calculated and compared between the artificial mouth and averaged test subjects.

It was found that the mouth of the B&K HATS 4128 was too directional on high frequencies. Also the torso simulator in the artificial mouth did not correspond to the real case. In the end of the thesis two schemes are proposed to improve the phone measurements.

Keywords: Head and Torso Simulator, Directivity, Mouth, Speech

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Teemu Halkosaari		
Työn nimi:	Radiation Directivity of Human and Artificial Speech		
Päivämäärä:	13.4.2004	**Sivuja:**	57
Osasto:	Sähkö- ja tietoliikennetekniikka		
Professuuri:	S-89		
Työn valvoja:	Prof. Matti Karjalainen		
Työn ohjaajat:	DI Markus Vaalgamaa		

Tämä diplomityö keskittyy tarkastelemaan ihmisen puheentuottojärjestelmän synnyttämän äänikentän suuntaavuutta. Tarkastelun tärkeimpänä motivaationa on puhelimien ja matkapuhelimien ns. handfree-lisälaitteiden mikrofoneille tehtävät taajuusvastemittaukset. Mittaukset suoritetaan ihmisen yläkehoa muistuttavilla keinopääsimulaattoreilla.

Julkaistua kirjallisuutta ei kuitenkaan löydy siitä mikä on oikean ihmisen suun suuntaavuuden ja keinopään suuntaavuuden erot ja onko mahdollisilla eroilla merkitystä matkapuhelinmittausten pätevyyteen. Lisäksi luonnollisen puheen suuntaavuutta tarkastellaan yleisesti.

Lähtökohtana tutkimukselle suoritettiin noin 15 hengen ryhmälle mittauksia kaiuttomassa huoneessa. Mittauksissa nauhoitettiin pään ympäriltä useammista eri pisteistä sekä yksittäisiä äänteitä että kokonaisia lauseita koehenkilöiden puhumana.

Mittauksista saadusta aineistosta laskettiin siirtofunktiot referenssimittapisteestä seitsemään mittapisteeseen. Mittauspisteitä oli suusta korvaan kulkevalla akselilla posken lähellä ja rinnan päällä. Mittauspisteet oli valittu siten, että ne vastaavat kohtuullisen hyvin oikeiden puhelimien ja handfree-laitteiden mikrofonien paikkoja.

Pään ja rinnan lähellä suuntaavuuden vaikutus näkyy pääasiassa korkeiden äänien vaimenemisena siirryttäessä suun akselilta sivulle. Käytetty keinopää ja keskiarvoistettu koehenkilöiden puhe erovat suuntaavuudeltaan merkittävästi tarkastelluissa mittapisteissä. Käytönnössa ero näkyi keinopään suurempana suuntaavuutena. Lisäksi puheen sisältö osoittautui vaikuttavan suuntaavuuteen.

Työn lopussa on esitelty kaksi ehdotusta matkapuhelinmittausten parantamiseksi.

Avainsanat: Keinopää, Suuntaavuus, Suu, Puhe

Acknowledgements

This thesis has been prepared at the Laboratory of Acoustics and Audio Signal Prosessing of the Helsinki University of Technology. The project was launced and funded by Nokia Mobile Phones.

I wish to thank my supervisor Prof. Matti Karjalainen and istructor Juha Backman. Especially I want to express my kindest gratitude for my other instructor Markus Vaalgamaa who made all the essential arrangements and guided me through the long process. A warm thank goes also to the crew at the Audio laboratory of Nokia in Salo for helping me with the measurements.

Otaniemi, April 13, 2004

Teemu Halkosaari

Contents

Abbreviations

ANOVA	Analysis of Variance
B&K	Brüel & Kjær
DC	Direct Current
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
IPA	International Phonetic Association
ITU	International Telecommunication Union
HATS	Head and Torso Simulator
MRP	Mouth Reference Point
PC	Personal Computer
SNR	Signal-to-Noise Ratio
SSD	Sample Standard Deviation

List of Figures

List of Tables

Chapter 1

Introduction

The new multimedia and internet related features have apparently overtook in importance the original purpose of the mobile phone. Nevertheless a good sound quality is a feature that the end user usually notices first, at least when the sound quality is reduced. The conversation chain in a mobile phone system consists of the microphone, the transmission path, and the loudspeaker. The motivation for this study is linked to the design of the microphone. The directivity of the mouth should somehow be compensated in the microphone system. The position of the microphone is far from the position where we are used to listen to a conversation.

When a microphone module is designed for a new phone model the response of the microphone is measured with some kind of an artificial mouth. The distance of the microphone to the mouth depends on the dimensions of the phone. The current trend is that mobile phones are getting smaller and smaller. Therefore 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. So 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 on off axis-positions.

There are standards that define guidelines for the directivity pattern of artificial mouths [20, 19]. Artificial mouths, especially the so called Head and Torso Simulator (HATS), will most probably be used also 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. On the narrow band (300 Hz - 3400 Hz) the directional features are not yet affecting and so the simulator is adequate, but on the wide band (150 Hz - 7 kHz) the situation is totally different as we will see later in this thesis.

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 [4, 7, 8, 12, 14, 24, 25, 30]. Also some publications, standardization, and product specifications can be found on the directivity of the artificial mouths [17, 19, 20, 22, 9]. Nevertheless the directivities of the artificial mouths are not systematically compared to test subjects by similar measurements in any of these studies. So clearly there is a need for measurement-based information how accurately the artificial mouths simulate the real human mouth and should their design be improved.

This study concentrates on both directivity characteristics of the human and the HATS mouths. 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 of human and HATS directivities be compensated in the measurements.

Directivity is also studied in general level. The far and near field conditions are discussed. The sound radiation pattern affected by head and torso is studied both by measurements and by modelling. Also characteristics of speech affecting to the directivity, such as pronunciation of different phonemes and articulation, are discussed.

The plan for the empirical part of the study was to measure responses for the mouth of the HATS in near field close to the head and the torso. Measurement positions were selected so that they would follow the positions of the microphones in real phone equipment. One far-field position directly on 0.5 m distance in front of the mouth was also measured as a reference point. The second part was to measure a couple of phones and headset accessories. Equivalent measurements were repeated for about 15 test subjects.

A multi-channel recording system was build up in an anechoic chamber. The audio recordings were analyzed in the Matlab software environment. In practice the results were a set of transfer function estimates between different microphone positions. The reliability of the measurements was ensured beforehand by careful measurement procedure design and afterwards by data analysis, such as signal to noise ratios and coherences.

The thesis is a part of a project that was launched and financed by the Nokia Mobile Phones. The contribution of the author in this thesis consists of requirements definition, measurement design, measurements, and analysis. The facilities and equipment were mainly supplied by Nokia and also by the Laboratory of Acoustics and Audio Signal Processing at the Helsinki University of Technology.

This thesis consists of six chapters. After introduction in Chapter 2 the background information on the subject is presented. A short literature overview is included in the end of the chapter. Conducted measurements are discussed in Chapter 3 which tries to give a complete view why and what was measured. In Chapter 4 there is a discussion how the measurement data was processed. The results, mainly transfer function estimates, are also included in the chapter. Some guidelines are discussed in Chapter 5 how the phone equipment measurements could be improved. Overall conclusions are included in the end of the thesis.

Chapter 2

Background

2.1 Human voice production system

The voice production system generates the sound which radiates from the mouth generating a sound field around the head and body. Nevertheless the directional features of the sound field are mainly dependent of the outer dimensions of the head and body. Still it is important to first discuss some basic properties of the voice production system. This subject is well presented in the literature. Therefore only a short summary is included below and the details can be found for example in [15, 10, 32, 23].

A scheme of the human voice production system can be seen in Figure 2.1. The vocal cords in the throat modulate an air flow from the lungs producing a periodic sound signal. Throat, mouth and nasal cavities filter the signal. This whole path is called the vocal tract. The main part of the produced sound radiates from the mouth opening. A part of the sound radiates also through the nose and directly through the chest, throat and cheeks.

Natural speech is a combination of various segments called phonemes. Articulation of some of the phonemes includes discontinuities in the air flow, vocal tract constriction generated noise or such. In the next section there are more details on phonemes and their classification. Also prosodic features in the speech affect how we use the voice production system.

What are the key dependences between the directivity and the structure of the voice production system? The sound mostly radiates from the mouth opening. Other paths for the radiation such as through the nasal cavity or directly through skin do have importance only in articulation of certain phonemes. So the main physical interface between sound production system and induced sound field around head and body is the mouth cross-section. The cross-section area or the size of the lip ring alters within the content of the speech, in other words, depending on the articulated phoneme. As a conclusion we can say that knowing

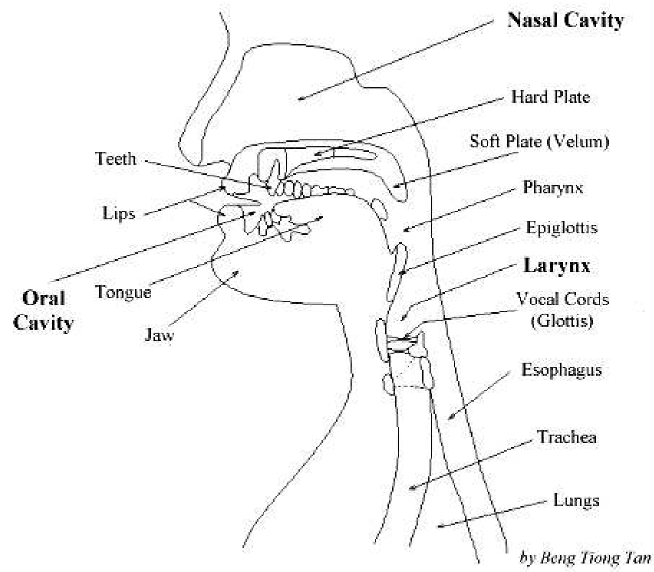


Figure 2.1: Scheme of the human sound production system [http://murray.newcastle.-edu.au/users/staff/speech/home_pages/tutorial_acoustic.html].

the size of the lip ring we know everything we need for sound radiation point of view.

The vocal tract shape during speech is presented in [11, 3, 29, 13]. In these studies the shape of the vocal tract is measured by magnetic resonance or X-ray imaging. Only separately uttered phonemes were included in the measurements. The results are presented as vocal tract cross-section area in 30 slices from the glottis to the lips. The slices most close to the mouth opening can be used as reference for the mouth cross-section area during speech. Later on we will see that the mouth cross-section area can also be estimated from the directional characteristics of the sound field.

2.1.1 Phoneme classification

If we start to split the speech in smaller and smaller parts, we finally end up with segments that are either close to stationary sound signals or some individual transients. These segments are classified in phonetics as phonemes. Phonemes are a means to transcribe the speech content. In this study the phoneme transcription is needed for defining the speech material used in the measurements.

There are several transcription systems for the phonemes. More or less they are language-dependent. For example the Finno-Ugric transcription is used in Finland [32]. The world-wide system to transcribe all languages is defined by the International Phonetic Association (IPA) [2].

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Figure 2.2: Pulmonic consonants according to the IPA classification [2].

Phonemes are classified in groups more or less using the physical means how they are generated in the vocal cords and the vocal tract. Pulmonic consonants according to the IPA can be seen in Figure 2.2 and vowels in Figure 2.3. There are more of these classification tables by IPA but these two tables include all the phonemes used for example in English and Finnish.

The classification implies mostly how the vocal tract is used to articulate the phoneme. In this study the interest is mostly on the mouth cross-section area during the articulation. The IPA classification cannot be directly used for this purposes. Nevertheless there are some phoneme groups that imply directly this feature and therefore they are also used in the analysis later on. For example in nasals the mouth is mostly closed and on the other hand the mouth is substantially open in open vowels. More detailed explanation on the articulation of the phonemes can be found from [32, 15, 23].

In this study the Finnish language was chosen as the language for the measurements. This made it easier to recruite the test subjects. Phonemes in the Finnish language according to the IPA are

Vowels: /ɑ, e, i, o, u, y, ø, æ/.
Consonants: /f, h, j, k, l, m, n, ŋ, p, b, t, d, r, ʋ, s, g/.

In Section 4.4.3 there is a discussion how the phonemes are grouped and used in this study.

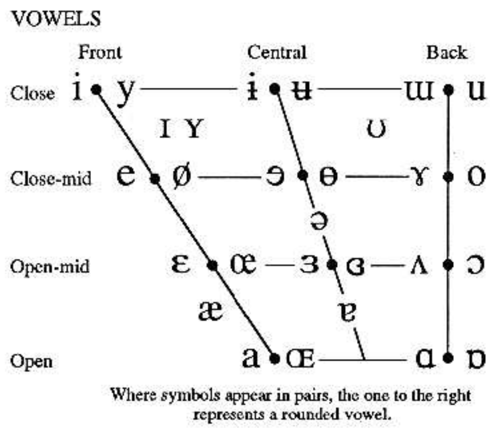


Figure 2.3: Vowels according to the IPA classification [2].

2.2 The directivity concept

The directivity as it is defined in classical acoustics for sound sources can be defined as the fraction of the field induced by the studied source and the field induced by similar omni-directional source. Similarity in this case means that the sources have the same total sound power. Directivity index DI is a logarithmic presentation of the fraction

$$DI = 10 \log_{10} \frac{p^2}{p_o^2} \tag{2.1}$$

where p is the pressure field of the original source and p_o is the pressure field of an omni-directional source.

There is also standardization on the concept. For example the ANSI standard [1] defines how the directivity is measured for a sound source. The standard is more or less prepared for noise source measurement purposes. Mainly it gives guidelines how to measure the sound directivity of a sound source to all directions in the far field.

This study focuses just on directivity to few positions in the sound field. The directivity is here considered by transfer functions for discrete field positions, referred to a fixed reference point. This method was directly adopted from the standardization of artificial mouth simulators [19, 20].

2.3 Theoretical modeling

The directivity of the voice production system can be estimated by various kinds of models. These models can be used in many situations. Later on the measurement results are assessed

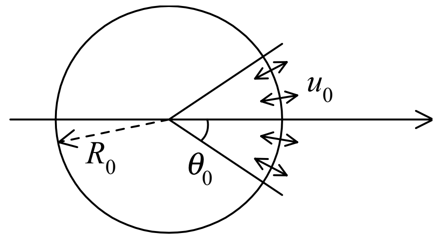


Figure 2.4: Head model which consist of sphere and a round radial vibrating piston.

comparing them to modeling. Also the reasons for the directivity characteristics can be predicted by altering the parameters in the models. Next, few models are discussed that are later on used in the analysis.

2.3.1 Sphere with a piston source

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 2.4.

The particle velocity u_r on the surface of the sphere is

$$u_r = \begin{cases} u_0 & , \quad \theta \leq \theta_0 \\ 0 & , \quad \theta > \theta_0 \end{cases} \tag{2.2}$$

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. The radial displacement is an adequate approximation because the mouth aperture is very small compared to the head radius. This approximation also simplifies the formulas.

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

$$p(r, \theta) = i\rho_0 c_0 \sum_{n=0}^{\infty} \kappa_n \frac{h_n^{(2)}(kr)}{h_n^{(2)}(kR_0)} P_n(\cos\theta). \tag{2.3}$$

P_n is Legendre polynomial, $h_n^{(2)}$ Hankel function, k the wave number

$$k = \frac{2\pi f}{c_0}, \tag{2.4}$$

ρ_0 the density of air, c_0 the velocity of sound, and f the frequency [26, 31, 33].

Coefficients κ_0 are dependent of the displacement pattern on the sphere. In this case they are

$$\begin{aligned} \kappa_n &= \left(n+\tfrac{1}{2}\right) u_0 \int\limits_{\cos\theta_0}^{1} P_n(x)dx \\ &= \tfrac{u_0}{2}\left[P_{n-1}(\cos\theta_0) - P_{n+1}(\cos\theta_0)\right] \end{aligned} . \tag{2.5}$$

The directivity is considered by transfer functions between two points in the sound field. Therefore also the model is used comparing two field points giving the transfer function T_{12}

$$T_{xy} = \frac{p(r_y, \theta_y)}{p(r_x, \theta_x)}. \tag{2.6}$$

If we substitute the pressure field terms with the Equation 2.3 we get

$$T_{xy} = \frac{\sum\limits_{n=0}^{\infty} \left[P_{n-1}(\cos\theta_0) - P_{n+1}(\cos\theta_0)\right] \frac{h_n^{(2)}(kr_y)}{h_n^{(2)}(kR_0)} P_n(\cos\theta_y)}{\sum\limits_{n=0}^{\infty} \left[P_{n-1}(\cos\theta_0) - P_{n+1}(\cos\theta_0)\right] \frac{h_n^{(2)}(kr_x)}{h_n^{(2)}(kR_0)} P_n(\cos\theta_x)}. \tag{2.7}$$

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 standard [20] specifies the dimensions for 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 10 cm value was used. The mouth aperture angle was derived from the product specification [9] where the mouth cross-section area is 3 cm^2.

2.3.2 Model for the body

The model represented in the previous section 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. These phenomena are in addition very much frequency dependent. On high frequencies the body absorbs the incident sound and on the other hand on low frequencies the body is too small an object compared to the wavelength to affect on the sound field.

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 very direct and simple way to model this reflection can be seen in Figure 2.5. The space is divided with an infinite baffle. The head model itself is the same as presented in the previous section.

The main advantage in the presented baffle model is that it can be formulated using a mirror source model. The effect of the infinite baffle is same as there would be a similar

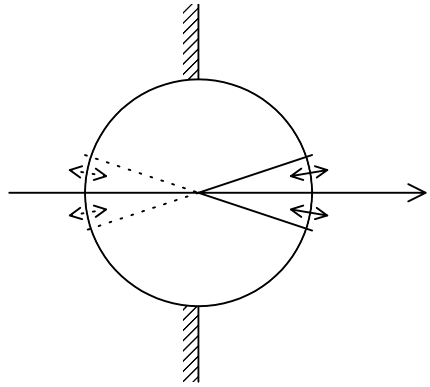


Figure 2.5: Same head model as in Figure 2.4 with an infinite baffle. Mirror source model is illustrated with dashed line.

sound source on the other side of the head as seen in Figure 2.5. The total field with the reflection can be expressed with the following equation

$$p(r,\theta) = p_{\text{direct}}(r,\theta) + k(f) * p_{\text{mirror}}(r,-\theta). \tag{2.8}$$

The mirror-source is weighted with the factor $k(f)$. This factor can be used to model the frequency dependent characteristics of the body reflection.

2.3.3 Other modified models

There are several other variations on the models presented above. The main rough approximations in the models were the head and mouth dimensions, the mouth placement in the head, the mouth radiation model, and the body model. Of course a question rises how much the model gets better if these approximations are made more realistic. There is some literature on these modified models, but it seems that the basic model gives adequate qualitative results for the study.

An comprehensive analytical formulation of spherical sound sources is found from [26]. There are different models presented for the direction of the mouth radiation displacement. Basically these changes affect on the coefficients κ_n in Equation 2.5. In [30] the spherical head is replaced with a spherical prolate. This does not seem to change the results much.

2.4 Artificial head and torso simulators

The objective measurements in telephonometry are not practical without an artificial mouth [22]. The response of an artificial mouth is usually tuned so that it corresponds to an average speech spectrum when an excitation signal of flat frequency response is used. The directivity pattern should also correspond at least in some parts of the field to the average person. Nowadays the artificial ear simulator is normally merged in the same package as well as some kind of shoulder or upper body simulator. The artificial ears are used for example in phone loudspeaker and earpiece measurements and for binaural sound recording. These applications have been at least as important drivers for the design of the head and torso simulators as the mouth.

2.4.1 Standards

The most important standardization for artificial mouths are included in two standards P.58 and P.51 from Telecommunication standardization sector of International Telecommunication Union (ITU-T) [20, 19]. The standard P.51 is more or less a subset of P.58. P.51 defines only the characteristics for the artificial mouth where P.58 is also a standard for the ear simulators and dimensions for the whole head and torso simulators.

The standard P.58 is titled as ”Head and Torso simulator for telephonometry” and it contains an extensive amount of requirements for the head and torso simulators. First in the standard the essential definitions are defined. For example, standard defines MRP (Mouth Reference Point, 25 mm in front of mouth) which is used later on in the analysis. Physical dimensions are presented with several measures for head, torso and pinnas. Acoustic characteristics are defined for both sound pick-ups, i.e. ears, and sound generation, i.e. for the mouth. The mouth directivity pattern is defined with 23 normalized free-field responses to different positions. In addition normalized obstacle diffraction, maximum deliverable sound pressure level, distortion and linearity characteristics are defined as well as composite characteristics: free-field plane wave diffraction at MRP, diffuse-field diffraction at MRP and mouth to ear cross-talk. Finally there are also some requirements for the surface materials for the simulator.

2.4.2 Products

Two main players on the head and torso simulator markets are Brüel & Kjær and Head Acoustics [16, 27]. Their main products can be seen in Figures 3.1 and 2.6(a). Both simulators are equipped with the same features. Both simulators have also ear simulators and a wide range of other accessories available for measurement purposes. The most notable difference is that in the Head Acoustics HMS II.3 there is no upper body in the simulator. Both products comply with the ITU standards [20, 19].

There are also small mouth simulators for simple handset microphone response measurements. These devices do not try to simulate the dimension of the head or so. The motivation is just to generate the standardized sound field. This technique for the measurements has

(a) HEAD Acoustics: HMS II.3 HEAD Measurement System.

(b) Brüel & Kjær: Mouth Simulator - Type 4227

Figure 2.6: Artificial mouth products.

a long history [22]. One example model from Brüel & Kjær can be seen in Figure 2.6(b). The product complies with the standard [19].

There are some important accessories available for the simulators. Especially these are necessary for the phone equipment measurements. For example a dedicated handset positioner and measurement vest can be seen in Figure 3.1 [9]. The vest should enhance the reflection properties of the torso to simulate better an average human upper body.

2.5 Literature overview on similar studies

There is not much literature on speech directivity in general. Few essential old articles discuss the basic aspects of the mouth directivity. In these studies a very large amount of measurements have been done. These studies give very comprehensive information on the radiation pattern in the far field and also in a couple of positions in the near field [12, 15, 14].

More detailed and specialized approaches to the subject can be found in some sources. Both simple measurement and model-based articles and combination of those can be found. A short overview is presented below.

A very interesting paper on the subject is the article by Sugiyama and Irii [30]. The paper has both a model-based approach and wide amount of measurements. The models that were used are already presented in section 2.3.3. Both far field and near field microphone positions were used in the measurements for a group of test subjects. The measurement positions as well as the reference point are partly almost the same as in this study. Therefore the results of the article are used later as a benchmark to assess the results.

Brixen from Delta Acoustics & Vibration has written two articles which are concentrated only on direct measurements [7, 8]. In both papers a lot of measurements were conducted for a group of test subjects using several measurement points around the head. Most interesting measurements were done by attaching microphones directly to the cheek, forehead, and chest. There is a huge amount of figures of measurement result presented in the papers. Unfortunately adequate conclusions are not included and also the reference point is chosen for the measurements so that the results are not easy to interpret neither compare.

In [17] the same head and torso simulator has been measured as in this study. The main focus in the study was to test the resiprocity theorem with the artificial mouth. All measurement points were in the far field at 2 meter distance and 10° azimuth angular intervals. Measurement results are compared to a spherical head approximation equal to the model presented in section 2.3.1.

McKendree from Westinghouse R&D Center recorded with 8 microphones 17 test subjects in an anechoic chamber. Microphones were in 1 metre radius from the head in 30° horizontal intervals. One microphone was positioned on 45° elevation from the horizontal plane. The motivation behind the paper was to improve our ability to make predictions of articulation index and speech privacy in open-plan offices [25].

More musician-oriented approach to the directivity can be found by Bartlett [4]. The scope in the paper is to discuss the miking for several music instruments. For the speech Bartlett has made few measurements with different distances in front of the mouth. In the paper there are the same basic equalization proposals how the difference between the near field and the far field should be compensated.

One other more or less phonetically oriented study on the difference of the far field and the near field can be found in [24]. The paper is based on measurements done in reverberant conditions using distances from 10 cm to 100 cm with 10 cm steps. The results are presented in coarse frequency resolution and cannot therefore be applied here.

The most important remark when considering the published literature above is that there is nothing directly to the field of this thesis. The directivities of the artificial and human mouth have not been compared in any of the references. The same measurement positions cannot also be found in literature that were used in this study. Also for the special aspects such as how the speech content affects the directivity or the handset-based point of view on the subject cannot be found in the literature.

Chapter 3

Measurements

3.1 Requirements

The starting point for this study was the measurements of handset and headset microphones used in mobile phone industry. Handsets and headsets are measured according to either standardized methodology or more often according to some company or product dependent specification. In Figure 3.1 a staged measurement setup is seen in parallel for mobile phone and headset. The headset is laying on the chest of the head and torso simulator (HATS). The handset is positioned so that the small loudspeaker, so called earpiece, is centered to the ear. The angle of the handset, for example, in this setup is a feature of the handset positioner. This kind of measurements are used widely in industry to design the microphones for phones and their accessories.

In these phone and headset measurements the microphones lay normally near the cheek or chest. Depending on the phone model there are several positions between the ear and the mouth which are used. 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 this study. Nevertheless the scope is to find some general features for directivity relying to adequate amount of measurements.

Some kind of key requirements have to be defined as the basis for the measurements. The requirements more or less are the answers to two questions: What to measure and how? In this study the first question is mainly answered by selecting representative microphone positions. The second question is covered also quickly because the measurements were to be pure audio recordings.

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. The same system with minimum changes was used through-

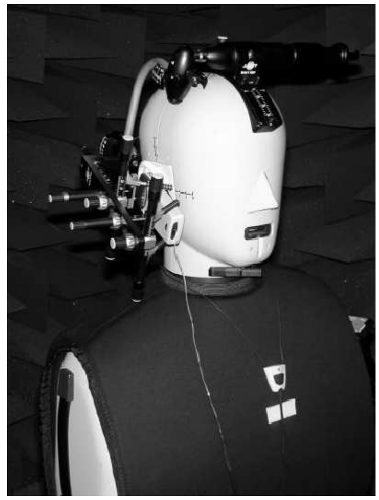


Figure 3.1: Staged measurement setup for mobile phone and headset. Head and torso simulator (HATS) is model B&K 4128. Measurement targets here are Nokia 8310 mobile phone and Nokia HDS-3 headset. A handset positioner (B&K 4606) is attached to the head and it is holding the mobile phone.

out all the measurement activities so that the measurements would be comparable to each other. The measurements were to be repeated for HATS and test subjects. The measurement system is discussed in the next section and the measurement processes are discussed in the end of this chapter.

3.2 Setup

All the measurements where conducted in anechoic chambers to ensure free field conditions for the measurements. Two different chambers in two locations were used: Helsinki University of Technology in Espoo and Nokia Mobile Phones in Salo. Both chambers meet the anechoic and the background noise requirements over the considered frequency range.

The measurement system was built inside and partly outside an anechoic chamber. An overview of the whole system is seen in Figure 3.2. The measurement chain begins with the microphones, following by biasing for microphones, preamplification, multi-channel A/D-converter, and finally the data is stored to a laptop computer. Some parts of the acquisition system were located outside the chamber so that minimum noise would be induced in the anechoic chamber.

The core in the acquisition system was IOTech Wavebook/516. It can read up to eight input channels in parallel. It converts the input signals directly to 16-bit samples using

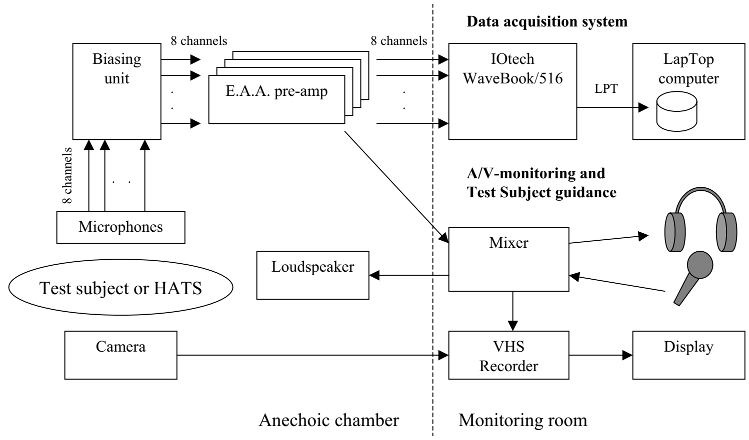


Figure 3.2: Measurement setup.

1MHz sampler shared between all inputs. Wavebook is connected to laptop PC over LPT-port. Wavebook is controlled with WaveView software that runs on Microsoft Windows operating system. WaveView enables an online monitoring of the waveforms of the input data, setting the parameters for the data acquisition (for example sample rate and sensitivity) and data storing to the hard drive. Wavebook offers also many other features, for example signal level triggering for the start of the recording [18].

The sample rate of 32 kHz per channel was chosen for all measurements because it seemed to use the maximum performance of the data link between Wavebook and laptop computer when all eight inputs were used. Acquired raw audio data was stored to files in 16 bits per sample format. 16 bits gives about 96 dB dynamic range which was adequate. The maximum dynamics for the data was ensured by adjusting maximum sensitivity for the sampler and at the same time monitoring the signal waveform from clipping. By preamplification the signal was amplified beforehand to ensure better signal to noise ratio and also adequate signal level for the sampler.

Two types of preamplifiers were attached to the system depending on the microphones: E.A.A. Professional Stereo Preamplifier PSP-2 and L&H Lernout & Hauspie Mouse AMP 31, Rev 3.3 (mono). Two L&H preamplifiers were used for phone handset and headset measurements. The electret microphones were connected to a separate biasing unit and four E.A.A. preamplifiers.

The audio and video monitoring system was built alongside with the acquisition system. Monitoring especially made easier to conduct the measurements for the test subjects. Conversation channels to both directions from operator to test subject and other way around was essential. A visual monitoring was built to track unwanted movements of the test

Table 3.1: Microphone positions near cheek and far field. The coordinate axes are referred to the lip plane and its perpendicular. H is horizontal displacement from the perpendicular and V vertical. The directions of the axes are illustrated in Figure 3.3.

Nro	On-axis (mm)	Off-axis (mm)
1	500	0
2	0	H 60
3	-10	H 70 V -40
4	-30	H 85 V 10
5	-70	H 100 V 30

Table 3.2: Microphone positions on chest. The coordinate axes are referred to the throat running down and sideward on the chest. The distance from the chest is not exactly defined although it is supposed to be zero.

Nro	Downward (mm)	Sideward (mm)
6	50	0
7	200	0
8	0	100

subjects during data acquisition. All the test subject sessions were also recorded with a VHS-recorder.

3.2.1 Measurement microphones and positioning

In the measurements both for test subjects and HATS there were two kinds of probes used: small measurement microphones and real mobile phones and their accessories. Separate microphones were used to measure accurately the directivity, and real products were used just for a point of comparison.

The electret microphones seemed to be the best choice for the study. The microphones had to be small so that they could be easily attached to HATS and test subjects. Also they had to be comfortable for the test subject to wear. As already mentioned above the acquisition system had eight input channels. So eight Sennheiser KE 4-211-2 microphones were used, one for each numbered microphone position throughout all measurements. The diameter of these Sennheiser microphones is only 4.75 mm. They have also an adequately flat frequency response.

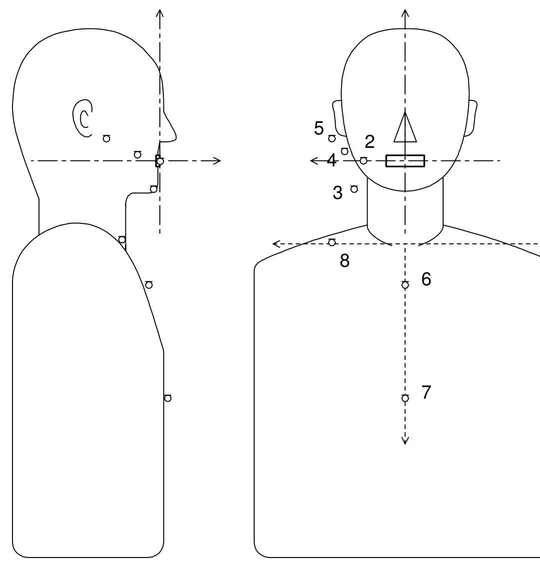


Figure 3.3: Microphone positions illustrated with HATS. Exact coordinates can be found in Tables 3.1 and 3.2

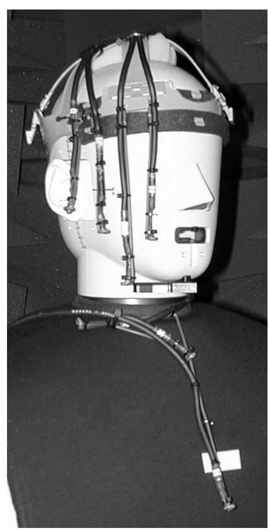


Figure 3.4: HATS wearing the measurement helmet and chest microphone array.

The locations for the measurement microphones are shown in Figure 3.3 and the exact coordinates are shown in Tables 3.1 and 3.2. Coordinate axes in Table 3.2 are referred to lip plane perpendicular. This axis runs from the center of the lips directly forward. The off-axis positions are defined in horizontal and vertical deviation from this axis. The positioning on the chest is referred to the throat where the axes run directly down and sideward. The coordinate system for chest positions is test subject dependent but is adequate because the starting point was the headset positioning.

The first microphone position is the reference point for all measurements. The reference is essential later on in analysis. The reference was selected so that it was easy to implement to the test subject measurements. Usually the reference in acoustical measurements is taken directly as close as possible from the sound source. In this case the mouth reference point (MRP, 0.025 m in front of mouth) would have been this point. MRP has however also lots of disadvantages. The stream of air from the mouth causes turbulence noise and the exact positioning close to mouth is also not that straightforward. Far field reference is corrupted by some reflection from the body and such. However as we see later on it does not cause problems when the original scope is considered.

A sophisticated measurement helmet was designed and built to attach the measurement microphones as exact as possible to the specified positions. Also the microphones close to chest were attached to each other in an array. The positioning equipment can be seen in Figure 3.4. The helmet and chest microphone array were used both for HATS and test subjects.

3.2.2 Handset and headset pick-ups

Measurements microphones give or should give an exact information on the directivity to studied positions. In addition, real phone equipment were also measured aside. Both cases are studied to to see if especially the positioning, i.e. how for example a test subjects hold the phones, causes variation to the directivity. Also the hand holding the handset has different shadowing effect on the sound field than the handset positioner.

Phones and headset that were used are listed in Table 3.3 and seen in Figures 3.5 and 3.6. As many different kinds of handset and headset models were selected to the measurements as can be seen from the figures. The selection was chosen so that it would represent the whole range of microphone positions used in phone equipment. The coarse correspondence to the measurement microphone positions is also included in the tables.

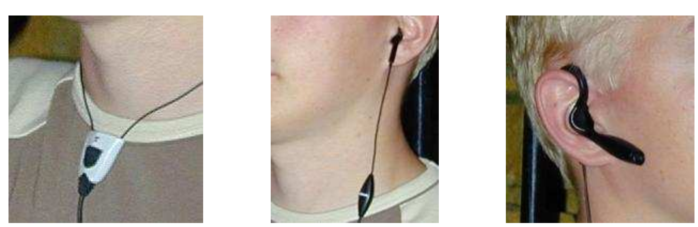
The microphones of the phones were wired directly out, biased, and preamplified with separate equipment as discussed in Section 3.2. Headsets were also connected to the same measurement chain directly.

Table 3.3: Mobile phones and headsets. Correspondences refer to the microphone position numbering in Tables 3.1 and 3.1

Product name	Length	Corresponds
Nokia, 9110	158mm	2 and 3
Nokia, 5110	132mm	2 and 3
Nokia, 8310	97mm	4
Nokia, Boom Headset HDB-4	-	5
Nokia, Stereo Headset HDS-3	-	6 and 7
Nokia, Headset HDC-5	-	8

(a) Nokia 9110 (b) Nokia 5110 (c) Nokia 8310

Figure 3.5: Pictures of mobile phones.



(a) Nokia HDS-3 (b) Nokia HDC-5 (c) Nokia HDB-4

Figure 3.6: Pictures of the headsets in use.

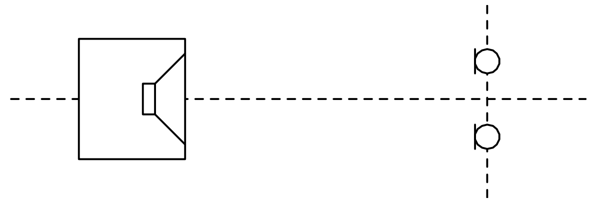


Figure 3.7: Calibration scheme. The reference microphone and the subject are symmetrically positioned aside the sound source perpendicular. The distance from the sound source is adequate for far field conditions.

3.2.3 Calibration

It is important to know if microphones itself corrupt the results in measurements. Electret microphones have usually significant sensitivity differences although they are of same model. Nevertheless their frequency responses are normally flat. The microphone assemblies in phones and headsets are tuned so that the narrow frequency band (300 Hz - 3400 Hz) is covered. In many cases this is done in expense of out of band frequencies. Therefore especially mobile phones and headsets had to be calibrated because a wider frequency band (150 Hz - 7 kHz) was covered in measurements.

All the measurement microphones, mobile phones and headsets were calibrated using the measurement setup shown in Figure 3.7. Calibrations were conducted in an anechoic chamber to ensure free field and minimum noise conditions. The axisymmetrically radiating sound source was driven with a sine sweep excitation. A B&K measurement microphone with flat response was positioned symmetrically with the calibration subject close to the axis of the source in far field. The sensitivity and frequency response of the subject were acquired comparing its response to the response of the B&K measurement microphone.

There are two key assumptions that cause error in this method. First the measurement microphone is though to have exactly flat response. Second the microphone positions are not exactly equal in the sound field. Nevertheless this method gives adequate results because these errors are minor compared to error sources in the actual directivity measurements.

The overall sensitivities of the Sennheiser microphones were within 4 dB difference. Responses were almost flat from 100 Hz to 10 kHz. In high frequencies over 1 kHz there was ± 1 dB fluctuation in responses. Microphones in mobile phones and headsets had huge up to 20 dB sensitivity differences as was expected. Also there was up to 5 dB boosts seen on frequencies above 3 kHz. On lower frequencies the responses were flat.

The most important scope in this study was a comparison of measurements for HATS and for test subjects. So the results are in practice comparisons of same microphone po-

Table 3.4: Measurement case flow for HATS.

Case	Pick ups (the amount of channels)	Remarks
1.1	All measurement microphones (8)	
1.2	All measurement microphones (8)	Torso uncovered
2.1	Mic. pos. 1, 9110, and HDC-5 (3)	
2.2	Mic. pos. 1, 8310, and HDS-3 adjusted up (3)	
2.3	Mic. pos. 1, 5110, and HDS-3 adjusted down (3)	
2.4	Mic. pos. 1, HDB-4 (2)	
3.1	Two B&K measurement mic. in mic. pos. 1 and MRP (2)	
3.2	Two B&K measurement mic. in mic. pos. 1 and MRP (2)	Torso unattached

sitions. Therefore the microphone response differences are automatically compensated. Nevertheless the directivity is also considered with single transfer functions. In these cases the overall level difference of two microphone positions are compensated. The frequency coloring was not compensated from any of the results because other error sources were more significant.

3.3 Measurements for HATS

All HATS measurements were carried out with B&K type 4128 HATS, because it is widely employed in telecommunication industry. It has also wide range of accessories and it is also equipped with a large upper body simulator i.e. torso. The torso was covered in all measurements with a 2cm thick foam plastic vest (B&K DS 0900 shoulder damping fabric). B&K 4606 handset positioner was used to position the mobile phones.[9]

The excitation signal to the HATS mouth was generated directly from a standard sound card of the laptop computer. Signal was amplified with the B&K power amplifier [9].

In HATS measurements a compensated white noise was used as excitation signal. The signal was frequency compensated so that the response was flat to the MRP within the frequency range from 200 Hz to 10 kHz where HATS mouth operates [9]. The excitation noise burst was adequately long to ensure saturated state in the system and enough data for analysis. The same microphone arrangements and case flow were used for the HATS and test subjects (see Tables 3.4 and 3.5).

To study the effect of the torso to far field responses, a special measurement cases were conducted for HATS (cases 3.1 and 3.2). Response to MRP (0.025 m in front of mouth) and to 0.5 m distance (microphone position 1) was measured in parallel using MLS (Max-

Table 3.5: Measurement session flow for each test subject.

Case	Pick ups (the amount of channels)	Sp. material	Sp. volume
1.1	All measurement microphones (8)	vowels	normal
1.2	All measurement microphones (8)	consonants	normal
1.3	All measurement microphones (8)	sentence	normal
1.4	All measurement microphones (8)	sentence	loud
1.5	All measurement microphones (8)	sentence	silent
2.1	Mic. pos. 1, 9110, and HDC-5 (3)	sentence	normal
2.2	Mic. pos. 1, 8310, and HDS-3 adjusted up (3)	sentence	normal
2.3	Mic. pos. 1, 5110, and HDS-3 adjusted down (3)	sentence	normal
2.4	Mic. pos. 1, HDB-4 (2)	sentence	normal

imum length sequence) measurement technique. The measurement was repeated for both the whole HATS and the head when torso was unattached from it. Also the effect of the covering vest was studied by one measurement case (1.2).

3.4 Measurements for test subjects

The measurement session was repeated for 13 test subjects: 5 female and 8 male. The test subjects were from 20 to 30 years old. 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 mother tongue.

The same measurement procedures were repeated for each subject. The session started with a short briefing how the speech material should be articulated. The subject was also instructed not to move during the recording. In anechoic chamber the subjects sat on a chair in which a thin rest for the head was attached. The far field microphone was attached to a separate microphone stand and positioned for each subject again. A paper was hanging in front of the subject from which the speech material could be read. An extra microphone stand was in line with the subject and the far field microphone so that the subject was able to focus better to keep the head still.

9 different recording cases were measured for each subject. The case flow is listed in Table 3.5. The speech material is discussed more detailed in the next section. The probes were always positioned as discussed above excluding the mobile phones. The test subject was only instructed to hold the mobile as naturally as possible.

Tasks of test subjects were monitored and guided with the system presented in Figure 3.2.

If unwanted clipping, articulation mistakes, movements or such occurred during recording the case was always repeated.

3.4.1 Speech material

The most important criteria to select the speech material for the measurements were to have a compact sentence that was easy to pronounce. Finnish was selected for the language because all the available test subjects had the Finnish as mother tongue. The sentence was also selected so that it contained all the phonemes in Finnish [32]. The distribution of the phonemes had to be also balanced. The sentence is presented below in IPA (International Phonetic Association) transcription and in Finnish [2]. A translation is also included.

Transcription:	/kɑksi ʋuotːɑ sitːen kæʋimːe rɑʋintolɑ gɑːbrielisːɑ helsiŋːisːæ jɑ søimːe sielːæ pɑdɑlːisen fɑsɑːniɑ bɑnɑːnilːɑ høystetːynæ/.
Finnish:	Kaksi vuotta sitten kävimme ravintola Gabrielissa Helsingissä ja söimme siellä padallisen fasaania banaanilla höystettynä.
English:	Two years ago we went to restaurant Gabriel in Helsinki and we ate there a pot of pheasant larded with banana.

Also pure phonemes were used to see the significance of continuous pronunciation. Unfortunately there are not many phonemes that can be articulated with continuous phonation [32]. For example plosives are articulated with a single constriction. Vowels and couple of suitable consonants were selected to the measurements and they are seen below. The test subjects were told to articulate the phonemes with steady state phonation for about 1s per each.

Vowels:	/ɑ e i o u y ø æ/.
Consonants:	/m n ŋ r s/.

Also different speech volumes were included in measurements as seen in Table 3.5. Although subjects were instructed beforehand what to say the loud and the silent speech volumes were more or less subject dependent. Loud and silent were defined for the subject in following way “Use your voice louder/silent but try to keep the articulation normal, i.e. don’t shout/whisper”. The cases were repeated if the articulation somehow did not meet the requirements.

3.5 Reliability and repeatability consideration

More or less the most important part in the design of the measurement system is to ensure that the system gives reliable results. Especially when test subjects are used these reliability issues play an important role. Moreover when a set of the same measurements is done consequently the repeatability has to be considered. A short summary on these aspects is included below although some discussion is found in the previous sections.

The repeatability was mainly ensured by making the decisions concerning the measurement processes beforehand and using the same processes all the way. The key repeatability aspects were the measurement chain similarity for each subject and the positioning of the probes. The positioning of the measurement microphones was ensured by attaching them to the helmet and chest rod frame seen in Figure 3.4. Microphones were attached to rods so they could be tuned for each person separately. More detailed discussion is found in Section 3.2.1.

The key feature in the measurement setup to check the quality of the data was the direct signal waveform plot in the acquisition software. Clipping and random noise bursts were easy to omit by repeating the case if such was monitored. Positioning of the probes was always checked before session and during measurement using video monitoring. Two microphone stands were lined in front of the subject so that the subject was able to focus not to move. Mistakes in the articulation and such were also omitted by direct audio monitoring of the takes. Noise inside the anechoic chamber was minimal in signal to noise point of view. The only noise source basically was the test subject itself.

Chapter 4

Analysis and Results

4.1 Analysis procedures

Extensive amount of measurement data was gathered during the measurement activity. The main objective was to study the directivity using transfer functions. Transfer functions would be referred to the same position in far field as discussed in the previous chapter. The reliability of the data has to be considered somehow. Also the natural speech measurement data was further analyzed to see if some of its features would have significant affect on results. As can be seen several kinds of data analysis were to be applied.

The multi-track measurement data was recorded in raw binary format with 16 bit accuracy. The data was analyzed in Matlab release 13 analysis environment. The analysis procedure is presented in the list below. A detailed discussion is following.

1. DC-rejection.
2. Data segmentation.
3. Inter-channel delay compensation.
4. Windowing and time to frequency domain transfer.
5. Transfer function calculation.
6. Coherence and SNR calculation.
7. Frequency smoothing and averaging.

The measurement chain should not have been induced any DC offset to the recordings. Nevertheless some channels were found to be corrupted. Therefore all the data was filtered with a simple elliptic infinite impulse response (IIR) filter. The filter was a high pass filter with a cutoff frequency at 20 Hz and a stop band ($<$1 Hz) rejection at least 80 dB.

A basic problem in speech analysis is the degree of divisibility of the speech wave. The most common approach has been to start with the linguistic criteria in terms of a phonemic transcription and to impose this as a basis for division. By a systematic comparison of the sound patterns of different contexts it is possible to make general statements as what sound features are typical for a particular phoneme [13].

One of the main scopes of this study was to see if the phonemes modulate the directivity pattern. Therefore phonemes and by the same means the active signal were labelled and extracted from the recordings.

In the study the recordings were segmented by listening and tracking the waveform of the signal. The speech material is discussed in Section 3.4.1. The HATS measurement data was simply segmented in one active signal part. A background noise reference for signal to noise (SNR) calculations was estimated from the leftovers of segmented data. Later on the segments are used separately in the transfer function estimate calculations for each channel.

There were significant delays between the measurement channels because the measurement positions were in different distances from the sound source or the mouth. The furthest distance differences were about 0.5 m (see Section 3.2.1) which corresponds 1.5 ms time delay difference. A delay between two signals can be seen as a phase shift in transfer function analysis. Therefore a small delay is not a problem in transfer function analysis as long as only amplitude is considered. However the segmented phonemes would have been interlaced if no delay compensation would not have been applied. So compensations were applied for all measurement cases and channels with an accuracy of one sample at the sampling frequency of 32 kHz. The delay estimates were obtained from the maximum of the cross-correlation between a considered channel and the reference channel. The microphone position 1 was selected as zero delay channel for all cases.

The data segments were windowed and for each window a Fast Fourier transform (FFT) was applied. FFT window was chosen so that more than one window would fit in each phoneme segment. 1024 sample Hanning window was used in all analysis. A 1024 samples with 32 kHz sample rate corresponds about 30 ms long window which is at least twice the length of the shortest phoneme segments. A 50% overlapping improved the overall FFT estimates.

Next the equations for the transfer function, SNR and coherence calculations are defined. The starting point is basically two signals $x(t)$ and $y(t)$. Later on we see that $x(t)$ is reference channel (microphone position 1) and $y(t)$ one of the microphone positions from 2 to 8 or a phone equipment.

Cross spectrum estimate G_{xy} for signals $x(t)$ and $y(t)$ is defined

$$G_{xy}(f) = \frac{2}{T} E\left[X^*(f,T)Y(f,T)\right] \tag{4.1}$$

where $X(f,T)$ and $Y(f,T)$ are the Fourier-transforms for the signals in time frame T. $X^*(f,T)$ is the complex conjugate of the term. If the cross spectrum estimate is calculated between signal and itself it is called the auto spectrum estimate or power spectrum.

Transfer function T_{xy} for two signals $x(t)$ and $y(t)$ is defined using auto and cross spectrum estimates. There are two formulations for the function

$$T_{xy}(f) = \frac{G_{xy}(f)}{G_{xx}(f)} = \frac{G_{yy}(f)}{G_{yx}(f)}. \tag{4.2}$$

The one with less noisy auto spectrum estimate should be selected always. Therefore also the SNRs have to be considered for both channels. The SNR is simply the ratio of auto spectra of active signal G_{xx} and background noise G_{xx}^{noise}

$$SNR(f) = \frac{G_{xx}(f)}{G_{xx}^{\text{noise}}(f)}. \tag{4.3}$$

Coherence $\gamma_{xy}^2(f)$ between two signals $x(t)$ and $y(t)$ is defined

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}. \tag{4.4}$$

Coherence can have values between one to zero. Coherence is one indicator for the similarity of the input signals and quality for the transfer function.[5]

The final transfer functions were frequency smoothed so that the results were more reliable and easier to interpret. All the results are smoothed in standard 1/3-octave bands by applying averaging for the bins within each band. At the same time the results were also limited in frequency range so that they would only cover the so called wide band. Wide band according to ITU-T recommendations is the range from 150 Hz to 7000 Hz. One reason for this was the fact that the B&K 4128 HATS, used in the measurement, operates only on the same frequency range.

For the natural speech data the transfer functions were smoothed to 1/3-octave bands using weighted averaging because the power spectra of different phonemes differ significantly. The frequency bins hitting to each band were averaged by weighting the bins with their coherence. This weighting makes the transfer function estimates more reliable.

$$T_{xy}(B_k) = \sqrt{\frac{\sum\limits_{f_i \in B_k} \left(|T_{xy}(f_i)|\, \gamma_{xy}^2(f_i)\right)^2}{\sum\limits_{f_i \in B_k} \left(\gamma_{xy}^2(f_i)\right)^2}}. \tag{4.5}$$

B_k is a set of bins f_i in each 1/3-octave band k. Other terms are the same as above.

4.1.1 Variance analysis

The measurement data was analyzed also with statistical methods. In case of HATS measurements, the analyses are straight forward, as the excitation signal is stationary random noise. On contrary, natural speech is not statistically a well-behaving random process. A variance analysis (ANOVA) is therefore introduced here to be later on applied to the measurement data. ANOVA is applied for example in [10] for same kind of extensive amount of speech data.

Analysis of variance (ANOVA) is a statistical concept that was selected to study the statistical structure of speech data. ANOVA is a method to test if there are relations between features in data and certain factors and moreover if the factors have interaction. So as a starting point a set of factors have to be defined which are assumed to affect on the data. The data sample is grouped under each factor forming an N-dimensional data mesh if N factors are defined. Each factor combination, in other words, a cell can have several replicates.

Basically N-way ANOVA can be used to determine if the means in a set of data differ when grouped by multiple factors as defined above. If they do differ, you can determine which factors or combinations of factors are associated with the difference. If a significant effect is found for some factor or an interaction of factors the analysis can be continued by comparing the sets of data within a factor or factors and trying to group them.

The linear 2-way ANOVA model for a replicate y_{ijk} is defined

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}, \tag{4.6}$$

where μ is the total average of whole data, the effect α_i is the deviation of the mean on level i from μ for factor A, the effect β_j the deviation of the mean on level j from μ for factor B, and e_{ijk} is the residual error. k represents the index for the replicates on same combination of levels i and j. The model can be expanded for N-way case by adding more effects α_i, β_j, etc.

As can be seen from Figure 4.8 the ANOVA analysis is based on the sum of squares (SS). For each factor the sum of square deviation from the μ is calculated as well as for the residual error. There are two types of sum of squares that can be used in the model. Type 1 is called fixed effects model and type 2 random effects model. The test values for the F-test analysis are calculated by dividing the SS of a factor with the residual error and scaling it with the degrees of freedom (DF).

In practice ANOVA analysis gives as result P-values from F-tests. The null hypothesis in the test is that the effects α_i, β_j, etc. are zero. The P-values represent the significance of each factor. If the P-value is smaller than 0.05 the considered factor is supposed to be significant. In other words it means that the factor causes significant variation to the data. If a considerable amount of the total sum of squares is residual error the model does not apply to the data. In other words the data does not behave in consequence of the factors.[21, 28]

If significant factors are found, the data within the factor can be grouped by various methods [21]. The idea is to reduce the levels in factor if statistical difference is not found.

This further analysis is not applied in the study.

ANOVA analysis was carried out in Matlab environment. There is a function for the N-way ANOVA analysis. The ANOVA function gives directly the results as a ANOVA table (see Figure 4.8).

4.2 Parameters for modeling

In Section 2.3 the theoretical models for the head as a sound source were introduced. In addition to the measurement results, the models help to understand the characteristics of the directivity. The models are rough simplifications of the real dimensions. Therefore the parameters and the positioning have to defined carefully by using some estimated values. Nevertheless the main scope is not the direct correspondence of the model and measurement but to get guidelines for the conclusion.

There are significantly different microphone positions that were measured (see Tables 3.1 and 3.2). The reference point is in far field directly in front of the mouth. Three positions were near the chest. Both the reference position and the chest positions are affected by the reflections from the body. The baffled model of the body is therefore applied for the chest positions. Four cheek positions are close to the head. These positions are not affected by the reflections from the body or shoulders significantly. The simple head model with no baffle is applied for these positions.

The two parameters for the models were the head radius and mouth aperture angle. 20 cm diameter for the head seemed to be adequately close to the average of head height, length, and breadth. Different mouth sizes were used depending on the case. The dimensions of B&K 4128 HATS mouth are 11 mm $\times$ 30 mm [9]. This area was used as a starting point. One important issue is that the mouth is not round as in the models. In other words it is acoustically seen larger close to horizontal axis and smaller on vertical axis. Therefore for positions close to cheek a larger mouth size usage is reasonable.

4.3 Directivity of HATS

The main scope of this study was to compare the transfer function measurement results between test subjects and HATS. These transfer functions are referred to microphone position 1 in the far field. Nevertheless first the directivity 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 consideration.

The positions that are interesting in mobile phone equipment point of view lay near the chest and close to the head near cheek. The reference point in this study was located in far field. The sound field is generated from the mouth aperture and modified by the head and torso depending on their size and the field position. These two objects in acoustical sense affect in three different ways the sound field. These aspects are listed below.

1. Mouth cross-section area.
2. Far field $\leftrightarrow$ Near field.
3. Body reflections.

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 wave length 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. Here the discussion 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. Again the head dimensions have to be much larger than the wavelength. On low frequencies the head does not affect.

Third, reflections, diffraction, and absorption from the upper body changes remarkably the sound field. Near the mouth, for example close to the cheek, these fluctuations are not strong enough so they can be omitted. In far field and especially near the chest strong reflection based fluctuations can be seen.

A far field position was selected as the reference point for all transfer functions. This point was selected because of the reasons explained in Section 3.2.1 although the MRP (Mouth Reference Point) would have been in general sense the best field position. In MRP the sound field is not corrupted by any reflection, diffraction or so. Nevertheless although MRP is not used as the general reference point it is next used to split the transfer functions in two parts. This way the directivity to different areas of the field is easier to consider.

If the system is linear the transfer function can be divided to two parts

$$T_{xy}(f) = T_{xz}(f) * T_{zy}(f). \tag{4.7}$$

The transfer function splitting is illustrated in figure 4.1. In other words the parts are eight transfer functions from MRP to microphone positions 1 to 8 (see Tables 3.1 and 3.2).

The direct transfer functions from the MRP to microphone positions were yielded from the measurements with two different procedures. The transfer function from microphone position 1 to MRP was measured with a separate measurement session as mentioned in Section 3.3.

The HATS mouth was driven by a compensated white noise excitation signal. The signal was equalized so that the frequency response in MRP was flat with ± 1 dB accuracy. The

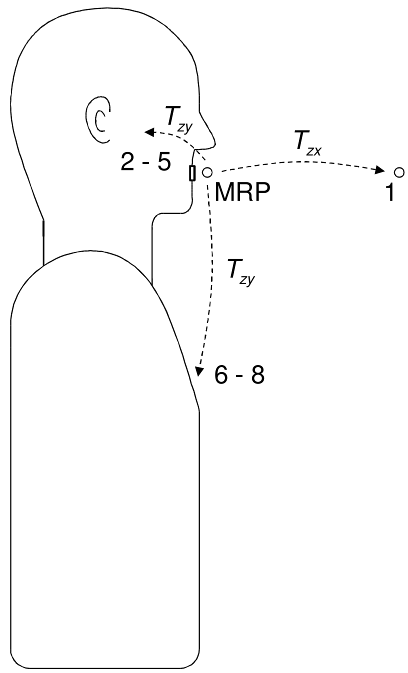


Figure 4.1: Illustration how the transfer functions are divided with MRP.

loudspeaker inside the mouth operates best around 200 Hz - 7 kHz which is also seen in the results. The compensation was done so that the output of the mouth operation was still linear and there was no distortion. Because the response to MRP was flat the power spectra can be used directly as the transfer function from MRP to the position (see Equation 4.2).

The results are shown in Figures 4.2 and 4.3. Modelled responses to part of the positions are also included. By these curves the results are easier to assess. Also they give guidelines to the reliability consideration.

From the curve for the position 1 we see that the head becomes significantly large comparing to the wavelength on frequencies above 500 Hz. This is the far field to near field difference effect. The reflection from chest is seen as fluctuation around 1 kHz. The body reflections and other aspects on that field are discussed more detailed in the next section.

Near the chest the microphone positions from 6 to 8 we see same sort of fluctuation but higher in frequency. It is caused by short reflection from the chest. The effect is similar to comb filtering. The dip is around 3 kHz which implies that there is about 5 cm difference in paths of the direct and reflected sound. It is important to notice that the torso is covered with a measurement vest. So the microphone positions are not directly on the hard surface of the torso simulator.

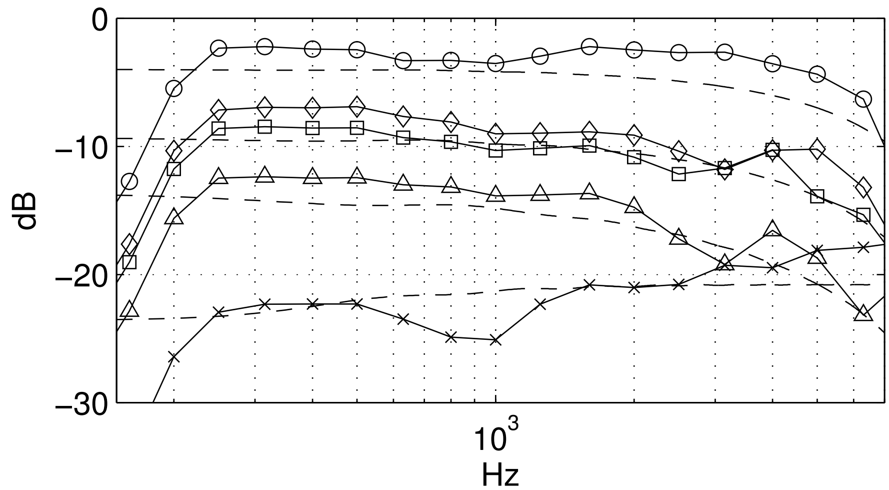


Figure 4.2: Power spectra for HATS to microphone positions 1-5 (×, ○, ◊, □, △). Excitation noise is flat in MRP so the curves correspond transfer functions from MRP to positions 1-5. Modelled cases are included for microphone positions 1, 2, 4, and 5 (- -).

The curves for close to cheek positions attenuate more on high frequencies when the position is closer to the ear. The overall level differences are the consequence of distance differences to the mouth. The same attenuations are seen also in chest positions but not that clearly. Also diffraction and reflections from shoulders can be seen as fluctuations in the transfer function for the closest positions to the ear.

If the results in Figures 4.2 and 4.3 are merged the transfer functions from the reference point to microphone positions 2-8 are got. These transfer functions are seen in Figures 4.4 and 4.5. They are later on compared to the averaged test subject measurements. The fluctuation caused by far field conditions on reference are seen in these curves and also later on when the HATS and test subjects are compared. Nevertheless if the general trend is considered the effect is omitted.

4.3.1 The upper body

It is advisable to use an about 2 cm thick measurement vest covering the torso when HATS is used in acoustic measurements. For example B&K has as an accessory for the HATS this kind of a measurement vest. 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 are considered. For example some of the headsets are placed close to the chest and therefore the sound field has near surface phenomena.

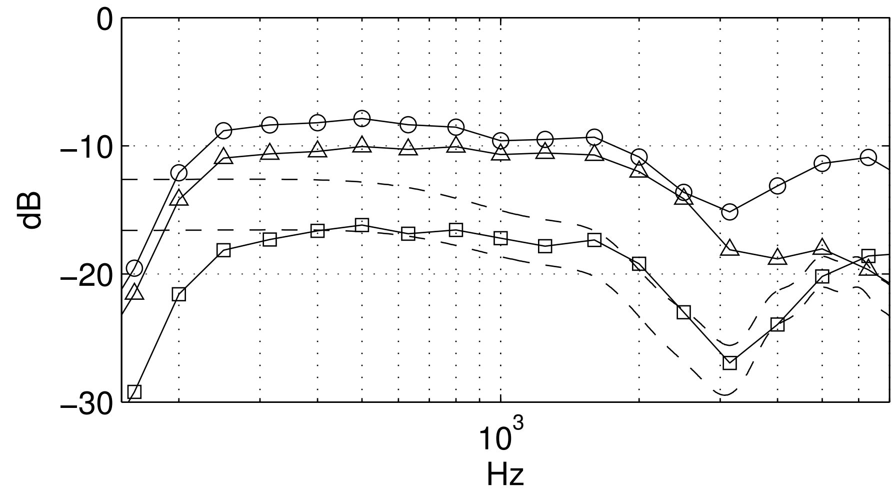


Figure 4.3: Power spectra for HATS to microphone positions 6-8 (∘, □, △). Excitation noise is flat in MRP so the curves correspond to transfer functions from MRP to positions 6-8. Modelled cases are included for microphone positions 6 and 7 (- -). In the model the distance from chest was 50 mm for position 6 and 90 mm for position 7.

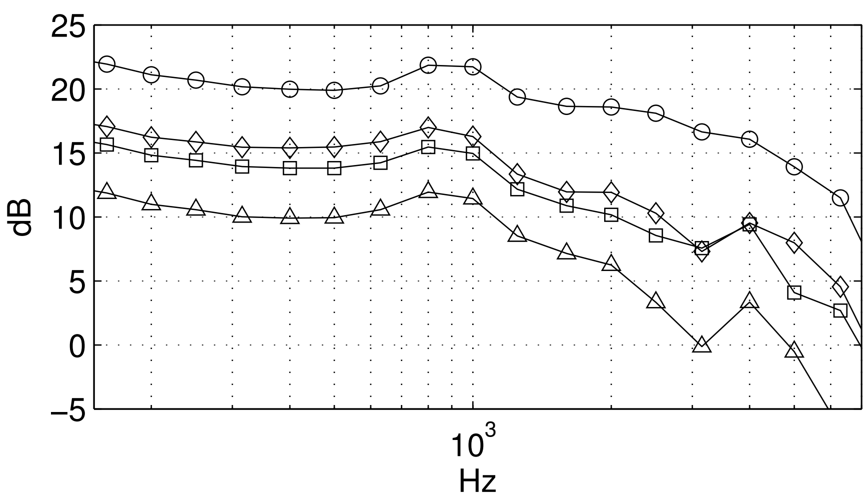


Figure 4.4: Transfer functions for HATS from microphone position 1 to microphone positions 2-5 (∘, ◊, □, △).

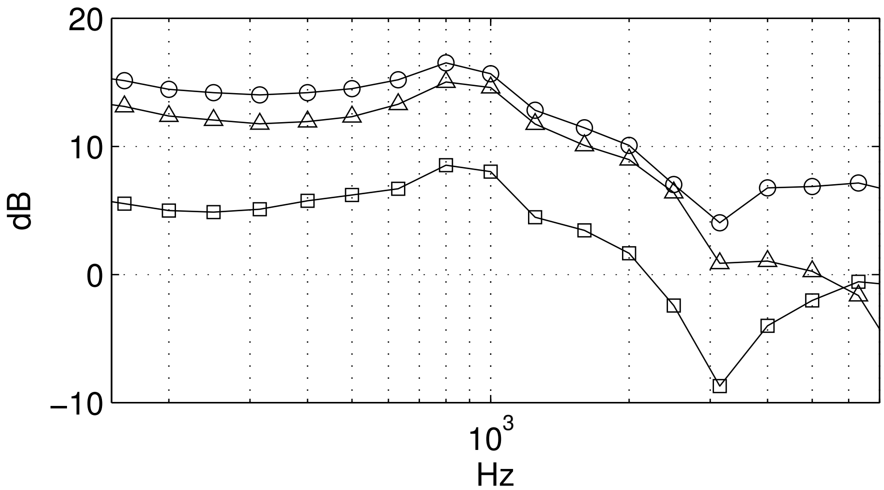


Figure 4.5: Transfer functions for HATS from microphone position 1 to microphone positions 6-8 (○, □, △).

To study the effect of the upper body simulator in HATS some extra measurements were conducted. The measurement process was repeated for the bare head of the HATS and also when the torso was not covered with the vest. In addition MLS transfer function measurements were done using MRP as reference.

In Figures 4.6 and 4.7 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 microphone position 1. 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. Near 3 kHz a dip is also seen for the chest position. The dips in both cases are repeated on higher frequencies in comb filtering shape.

The vest that is used in measurement has a remarkable effect on wideband measurements. The vest attenuates the reflection and increases the delay difference of the reflected and the direct sound. The dip for the chest position does not occur if vest is not on. It seems that the dip shifts to higher frequency because the delay difference in direct and reflected sound goes smaller. For microphone position 1 we can see that the dip is deeper if the vest is taken off. This implies that the vest really attenuates the reflection.

Finally as a benchmark for the consideration we see that if the torso is not present the transfer function for microphone position 1 has no fluctuation. It follows the pure spherical head model throughout the considered frequency band.

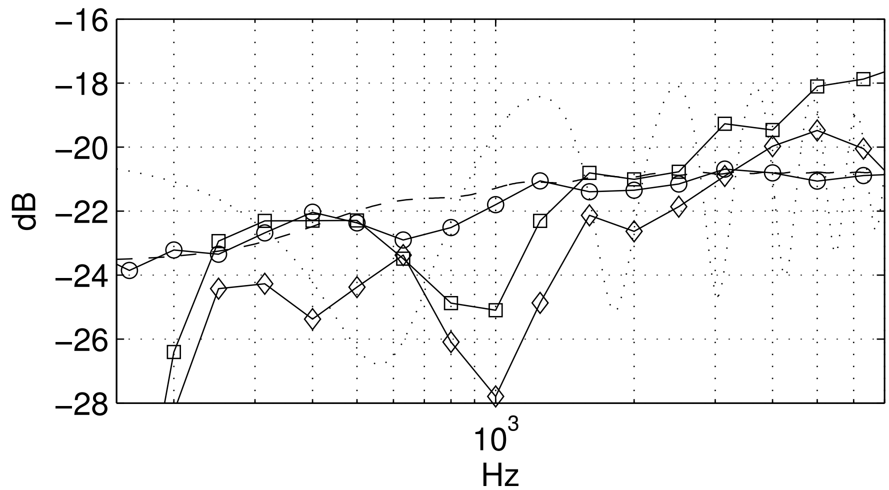


Figure 4.6: By power spectrum means measured transfer functions from MRP to microphone position 1 for cases were HATS was with (□) and without vest (◊). Similar MLS measurement for the bare head (○). Bare head (- -) and infinite baffle (··) models are included.

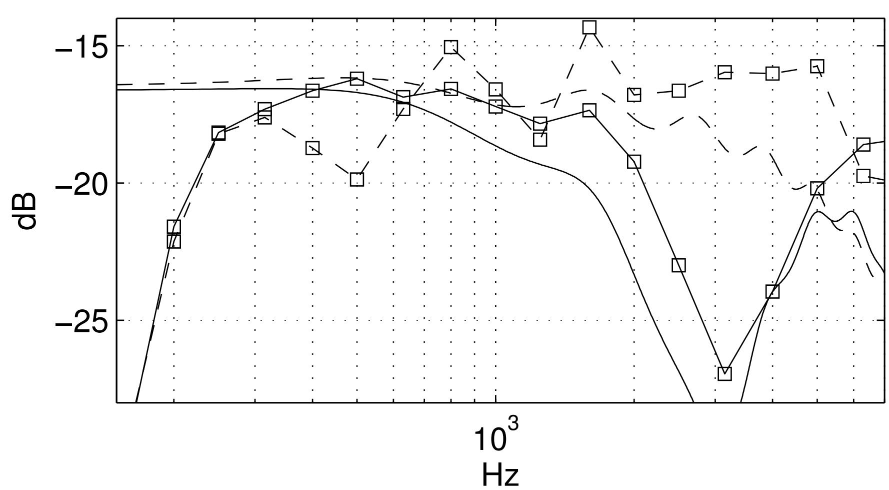


Figure 4.7: By power spectrum means measured transfer functions from MRP to microphone position 7 for cases were HATS was with (□–) and without vest (□- -). Two corresponding models are included. In the models the positions are on in 10 mm (- -) and 90 mm (–) distance from the chest.

4.4 Directivity of test subjects

The directivity to different positions is discussed for HATS in the previous chapter extensively. Test subjects were measured using similar measurement positions. Therefore the directivity itself is not discussed here again. Instead the differences of averaged natural and HATS speech are discussed and how the speech content affects on the directivity.

4.4.1 Speech content and variation

Long-term averages are easy to use as spectral basis for directivity discussion and after all the HATS for example tries to correspond to an averaged human head and upper body. Nevertheless there are aspects that have to be covered to ensure the reliability of long term averaged speech. On the other hand we can just accept the variations in speech and suppose that the average is representative and enhance the average by using some kind of weighting to averaging.

Before we go further in comparing the averaged test subjects to HATS, few aspects of natural speech are covered to show how they affect on the directivity. A direct motivation for this is to see if speech material and the subjects change the measurement results. One self-evident fact is that the physical characteristics like gender, height, weight, etc. affect, but that variation factor can be controlled by selecting the subjects carefully.

The analysis of variance (ANOVA) concept was represented in Section 4.1.1. ANOVA analysis was applied to the speech data for most of the acquired measurement data. It is self-evident that the speech is not a simple set of data which meets some standardized distribution. By ANOVA the data can be classified using factors and test if they are significant. In speech these factors are by far the phoneme, the speech volume etc.

The obvious hypothesis is that the factors in speech affect on the results. So by ANOVA analysis the hypothesis is tested by statistical means. The analysis gives also a course view on the characteristics of the speech data.

The data for the ANOVA was gathered in the following way. Four factors were selected under which the data was classified:

1. Speech volume
2. Phoneme
3. Test subject
4. 1/3-octave band.

The transfer functions from microphone position 1 to positions 2-8 were all analyzed separately. In other words the data from measurement cases 1.3-1.5 were used as a starting point (see Table 3.5). The segmented parts, i.e. phonemes of the sentence, were windowed

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Speech volume	121.3	2	60.64	2.01	0.1334
Person	46733.6	12	3894.47	129.38	0
Phoneme	34831.4	23	1514.41	50.31	0
Frequency	101629.8	16	6351.86	211.02	0
Error	3417608.8	113540	30.1		
Total	3600924.9	113593			

Sequential (Type I) sums of squares.

Figure 4.8: ANOVA analysis results for transfer function measurement data for microphone position 4.

without overlap and for each window a transfer function was calculated. Windowing parameters were the same as listed in Section 4.1. There were replicates for each cell because there was more than one window of data for every phoneme as well as for every phoneme instance.

The ANOVA gave as a result P-value for each factor. As an example the ANOVA table for microphone position 4 is seen in Figure 4.8. P-values were for all seven channels less than 0.05 except for speech volume. So all the factors excluding speech volume for all seven channels were significant. The result goes with the hypothesis that the selected factors are significantly affecting the directivity. The speech volume however seems not to have much effect on the data.

If the sums of squares are considered we see that the error is extensive. So the ANOVA model does not fit to the data well. In other words the data is scattered also within the factors significantly. For example this means that one test subject utters the same phoneme differently on each instance.

Because extensive amount of error was found in the ANOVA analysis the analysis was applied in smaller parts of the whole data. First the speech volume factor was omitted. The ANOVA analysis was applied only for the normal speech level data. Further on the analysis was applied for each frequency on the normal speech level. The results were similar to the values in Table 4.8. P-values stayed less than 0.05 and the residual error was significant although the speech level and frequency factors were omitted. So the data seemed to be really scattered as mentioned above.

4.4.2 Long-term averaged speech and HATS

The overall characteristics of the sound field around the head and upper body were discussed using the HATS as subject. Nevertheless the key issue of this theses was to compare real human directivity to directivity of the HATS. The comparison is done using a long-term averaged speech over all data as the estimate for real speech.

The difference is considered basically by dividing the transfer function for the averaged test subjects $T_{\mathrm{TS}}^{\mathrm{ave}}(f)$ with the transfer function for HATS $T_{\mathrm{HATS}}(f)$

$$T_{\mathrm{diff}}(f) = T_{\mathrm{TS}}^{\mathrm{ave}}(f)/T_{\mathrm{HATS}}(f). \tag{4.8}$$

The transfer functions for HATS can be seen in Figures 4.4 and 4.5. All the transfer functions are referred to microphone position 1. In an ideal case the reference should be omitted in the division. Here the far field differences cause some fluctuation in the scheme. In the averaging the coherence weighting was applied as discussed in section 4.1.

The results are seen in Figures 4.9 and 4.10. A model is included in the figures for some of the channels. Over all we see that the differences are not that much dependent of the position. Especially near cheek the curves are almost similar as the model predicts. Second essential point is the fluctuation in the curves. The results are next discussed in details.

In all results a peek is seen around 800 Hz. Mostly this is caused by the difference of the references. It seems that the reflection from the chest is significantly different for HATS than for averaged test subjects.

In the models for all positions the mouth cross-section area for HATS was larger than for test subjects. Especially near the cheek in Figure 4.9 for position 2-5 we see that the larger mouth is more directional if the general trend is considered. Moreover the difference is similar near the cheek regardless of the position.

In the curves in Figure 4.9 for the chest positions also the near chest reflection difference is seen as a peak around 3 kHz (see Figure 4.7). The mouth cross-section area difference seems to affect less near the chest, which is seen if the general trend is considered. For microphone position 6 there is almost 5 dB overall level difference. This is caused mostly by the difference of the dimensions. In other words the microphone 6 is closer to the mouth for test subjects.

4.4.3 Mouth aperture size and 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 Figure 4.9 were modelled using two different mouth aperture sizes. As it is seen the general trend of the measurement results follow the modelled curves if the fluctuation in frequency is omitted.

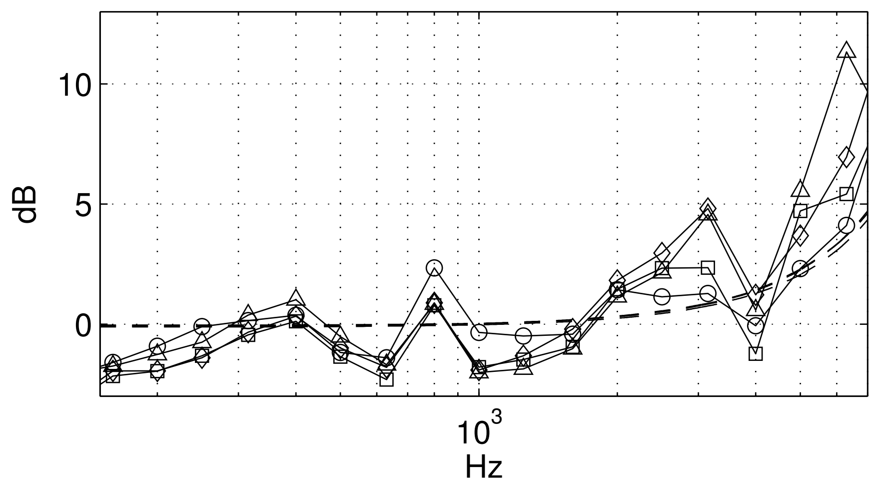


Figure 4.9: Difference of averaged test subjects speaking in normal speech volume and HATS. The comparison is done for transfer functions from microphone position 1 to 2-5 (∘, ◊, □, △). A positive dB value implies that the HATS is more directional on that frequency (see Equation 4.8). Three corresponding models are included for microphone positions 2, 4, and 5. The mouth aperture radii for corresponding HATS and averaged person in the direct models are 1.5 cm and 0.5 cm.

Because this 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 [11, 3, 29, 13].

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

As we already saw in Section 4.4.1 there are several factors that cause variance in speech data. Same of these factors should directly be linked to the mouth aperture size. For example two factors listed below are intuitively self-evident:

1. Continuous phonation $\leftrightarrow$ natural speech.
2. Vowels grouped by openness.

In the measurements continuously phonated phonemes were also measured. It is obvious that within speech the vowels are not so pure, in other words the mouth size is smaller. The vowel classification should also directly be linked to the physical mouth size.

The phoneme grouping for the vowels can be seen in Table 4.1. The recordings were segmented into phoneme groups. The average transfer functions for the phone groups were

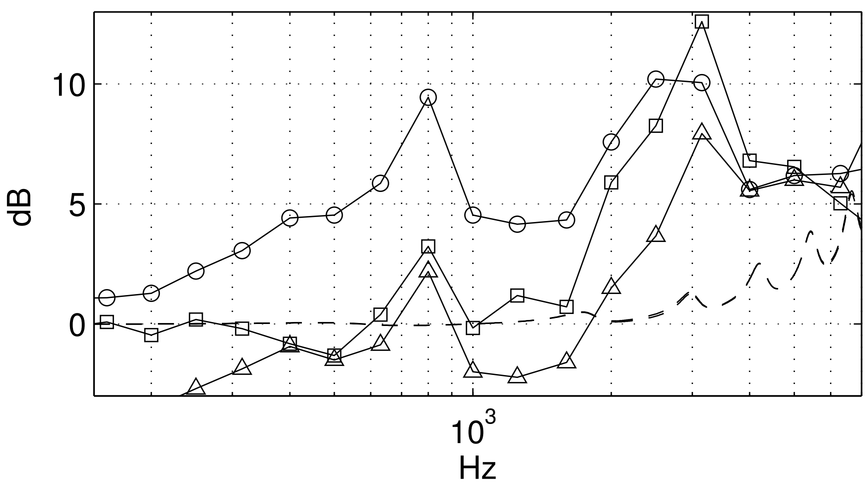


Figure 4.10: Difference of averaged test subjects speaking in normal speech volume and HATS. The comparison is done for transfer functions from microphone position 1 to 6-8 (○, □, △). A positive dB value implies that the HATS is more directional on that frequency (see Equation 4.8). Two corresponding models are included for microphone positions 6 and 7. 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 an 1 cm distance from the baffle.

calculated by weighting with the coherence functions (see Equation 4.5).

One representative microphone position was selected for further consideration. Results are shown as differences between averaged test subjects and HATS as in Figures 4.9 and 4.10. Basically if the mouth is larger the curve should be on high frequencies closer to the 0 dB axis. In both Figures 4.11 and 4.12 we see that the hypothesis made goes with the measurement results if the general trend is considered. Again we have to omit the fluctuation caused by the reference point.

The speech volume by far should correlate the way that we use the mouth and the whole articulation system and so the size of mouth. Also some other prosodial and segmental features in speech should have same kind of link to mouth size. Nevertheless these considerations are not included here because straight mouth aperture dependence was not found.

4.5 Handset effect

The starting point for this study was the transfer function measurements of microphones that are widely used in mobile phone industry. The results and discussion, so far, have

Table 4.1: Finnish vowels grouped by openess [32, 2].

Group nro	Including	Description
1	/i, y, u/	vowels: close
2	/e, ø, o/	vowels: close-mid
3	/ɑ, æ/	vowels: open

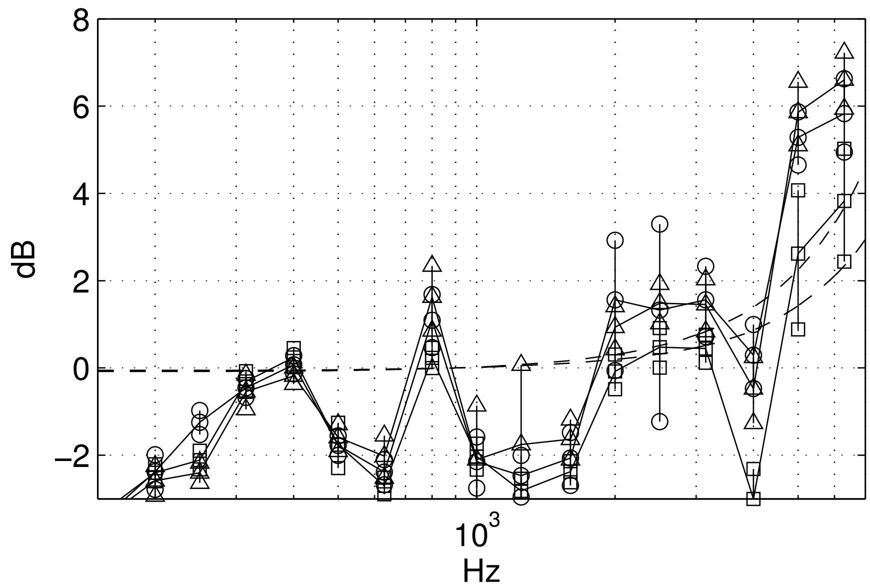


Figure 4.11: Transfer function difference between averaged test subjects and HATS in microphone position 4 for open vowels (□), close-mid (○), and close (△). Two modelled cases are included for microphone position 4. The mouth aperture radii for corresponding HATS and averaged person in the direct models are 0.5 cm and 1.5 cm (same as in Figure 4.9) and for other curve 0.5 cm and 1 cm. 95% confidential intervals can be seen for each frequency.

been concentrated only on measurements that were done using measurement microphones. Positions were selected so that they corresponded the positions of the microphones of mobile phones in use. However when the microphone is attached to a large handset it should change the response if the considered wavelength is small compared to the phone dimensions. Further on the hand holding the phone should also change the responses. Therefore a set of phone and headset measurements were included in measurement activity (see Tables 3.4 and 3.5).

As already discussed the phone should affect the sound field. For example a handset

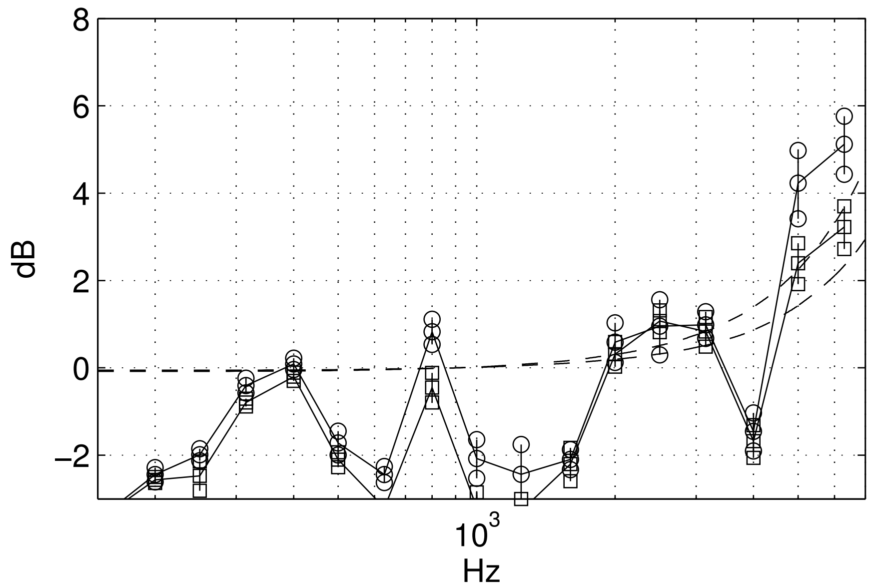


Figure 4.12: Transfer function difference between averaged test subjects and HATS in microphone position 4 for vowels articulated within the sentence (○) and separately articulated vowels (□). The averages are done from cases 1.1 and 1.3 in the table 3.5. The same modelled cases are included as in Figure 4.11. 95% confidential intervals can be seen for each frequency.

and a hand shadow the microphone and therefore the microphone construction can become directional although the microphone itself is omnidirectional. Nevertheless the effects are fairly small compared to the confidential intervals of the results. For example the maximum dimension of the phones (see Table 3.3) is about 10 cm. That corresponds to wavelength around 3 kHz.

In Figure 4.13 we see the results for two phone models. The other results for the phone equipment are not included but they go along with these results. As we can see the curves are very similar with the curves that were measured by measurement microphones. Some overall level difference can be seen between measurement microphones and phones. Also we see that the difference between test subjects and HATS on high frequencies is significantly larger, between 5 to 10 db.

One important remark is that the curves for mobile phones are an average of position. This is because the holding of the phones was not regulated in test subject measurements. When the different positions (see Figure 4.4) are averaged in the linear scale the large values are emphasized in logarithmic scale. This is one obvious reason for the difference between

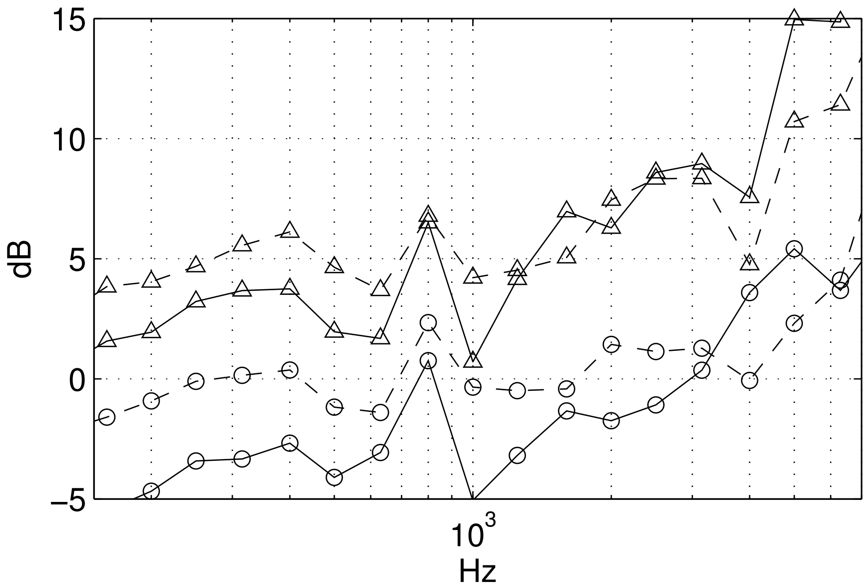


Figure 4.13: Difference between measurements for averaged test subjects and HATS with Nokia 9110 (○–) and Nokia 8310 (△–). Corresponding measurement microphone cases are included for positions 2 (○- -) and 4 (△- -). For separation reasons 6dB offset is added to Nokia 8310 and microphone position 4 curves.

phone and measurement microphone results.

Finally it is worthwhile to mention that the microphones in phones are normally tuned with cavities and holes and moreover it is not rear to use a directional microphones. In these cases the directional issues play a more significant role. These issues are omitted here.

4.6 Reliability discussion

4.6.1 SNR and coherence

The simplest means to consider the reliability of the measurement data is to examine the coherences and signal-to-noise ratios (SNR). These concepts were defined in Section 4.1. SNR characterizes the background noise ratio of the data. The coherence is calculated between two signals and it basically characterizes the similarity of the signals. Therefore it also is an indicator for the noise. In transfer function considerations it can be used directly as the reliability indicator for the transfer function of the considered two signals. Both SNR and coherence are estimated by cross spectrum estimate and auto spectrum estimate.

The coherence function $\gamma_{xy}(f)$ can have values between 1 and 0. If coherence is 1 the signals are coherent. On the other hand, if signals are completely uncorrelated so that $G_{xy}(f) = 0$, then the coherence is zero (see Equation 4.4). When the coherence function is greater than zero but less than unity, one or more of the following four main conditions exist:

1. Extraneous noise is present in the measurements.
2. Resolution bias errors are present in the transfer function estimates.
3. The system is not linear.
4. The output y is due to other inputs besides the input x.[6]

For speech signals the coherence and SNR vary strongly depending on the content of the speech. With the HATS the starting point for the reliability consideration is totally different. The excitation signal was selected so that the output from the mouth was at its maximum performance on considered frequencies. Therefore it is adequate if the reliability considerations are done only for the test subject data.

The SNR and coherences were calculated by averaging the data as was done for the transfer function estimates. The noise estimate was taken from the end of each recording starting from the end of the last segmented phoneme. By this overall average consideration we do not see if parts of the speech material and so the transfer function estimates are badly corrupted by the noise because of silent phonemes or such. Nevertheless these corruptions have to be accepted because long term averaging for the speech was selected as the starting point of the study. The corruption was reduced using the coherence weighting for the transfer functions.

Figures of SNR and coherence are not included because it was found that the estimates were adequate on all frequency bands that were considered. The data from the measurement case 1.3 (see Table 3.5) was used as the input data for calculations. The worst SNR in microphone position 1 to 8 was greater than 25 dB over the wide band. The coherence between the microphone position 1 and positions 2-8 were greater than 0.8 on narrow band and greater than 0.65 on wide band. The values imply that the results are reliable on wide band and very reliable on narrow band [5, 6].

4.6.2 Confidence intervals

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. Confidential intervals are the 95% probability intervals for normal distribution that is obtained from the distribution of the considered data. As we can see the data has to be assumed as normal distributed. Therefore confidential intervals are mainly used to assess some random error in measurement data. Although the data is not even close to normal distribution the confidential intervals characterize the variance and so the reliability of the data [21, 5].

The confidential interval analysis was applied here to transfer functions of the test subject measurements. As already mentioned the data is varied because of the speech content and test subject as well as the measurement errors. So the confidential intervals here stand for both known varying factors and unknown measurement errors.

The confidential intervals are seen in figures 5.1, 4.11, and 4.12. The intervals stay mostly within a ± 1 dB range. This implyes that the results are very representative and for example the differences between phonemes are statistically significant. The intervals are calculated by fitting the mean and sample standard deviation (SSD) of the data to normal distribution.

One other way to estimate the error for the magnitude of the transfer function $|T_{xy}(f)|$ is to use directly the coherence γ_{xy}. 95% confidential intervals are given approximately by

$$\Big[|T_{xy}(f)|\,(1-2\varepsilon) \leq |T_{xy}(f)| \leq |T_{xy}(f)|\,(1+2\varepsilon) \Big], \tag{4.9}$$

where

$$\varepsilon = \frac{\sqrt{1-\gamma_{xy}^2(f)}}{|\gamma_{xy}(f)|\sqrt{2n}} \tag{4.10}$$

and n is the amount of averaged data samples.[6]

In Figure 4.14 the approximate confidential intervals can be seen as a function of coherence. The value 2228 for n is the average amount of data samples in data for difference of test subject and HATS per case and microphone position (see Figure 5.1. There are 24 phonemes in Finnish language, so the other intervals represent the average intervals for one phoneme.

The approximate confidential intervals seem to give values that go with the direct method of calculating the intervals from the mean and SSD. In Section 4.6.1 the coherence for data was discussed. The approximation predicts smaller than 0.5 dB intervals for coherences less than 0.5. For an average phoneme the approximate intervals are significantly wider. So it seems that the huge amount of data for the main results was a key factor in ensuring the reliability.

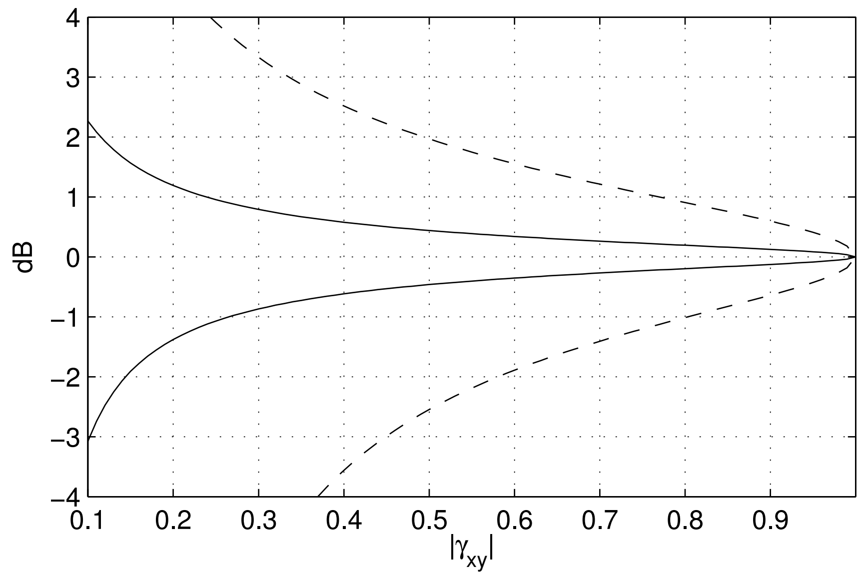


Figure 4.14: Confidential intervals yielded from the Equation 4.9. n is 2228 for solid lines and the same value divided with 24 ($n \approx 93$) for dashed lines.

Chapter 5

Improvement proposals for telephonometry

5.1 Equalization concept

When the directivities of the HATS and the averaged natural speech of a group of test subjects were compared, significant differences were found (see Figures 4.9 and 4.10). 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 which 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 Figures 4.6 and 4.7). For these reasons it is meaningless to yield an equalization scheme for these positions and therefore they are omitted in this approach.

The curves for the positions near cheek in Figure 4.9 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 neighbour frequency bins weighting with Hanning window. Window size was nine so 4 frequency bins were used from both sides for averaging. This window size corresponds to three octaves. The smoothed curves can be seen in Figure 5.1.

The smoothed curves are within about 2 dB on all frequencies. If the confidential intervals are considered the curves are, at least in two groups, statistically different on high frequencies. Monotonic behaviour on some dimension would be interesting to find. If the low frequency level differences are omitted as well as the microphone position 3, we see that the difference on high frequencies gets slightly steeper when the position is in a further distance from the mouth. Still this dependence considering the confidential intervals is not that significant. One interesting fact is that the model (see Figure 4.9) predicts that there should not be any difference. So the dimensions of the head do not comply the simplified sphere model.

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

1. Average curve or curves.
2. Separete curves for each position.
3. A model based position dependent equalization.

A general average would be easiest to use in the actual measurements. The model based approach typically means an interpolation and extrapolation between the positions. For example in this study there is no sufficient amount of positions to implement this scheme. If position dependent equalization curves or several average curves are defined the position of the phone microphone has to be determined accurately. The position depends on the phone design and the handset positioner of the artificial mouth. So the set of equalization curves has to be designed so that they somehow are easy to link to different sizes of phone models.

Finally it is an important question how accurate and extensive in general the equalization scheme should be. The end users hold the phones in various angles. Still the frequency response for the microphone of the phone is normally measured only in one microphone position.

5.1.1 Filter design example

A filter design scheme is next presented. The motivation is to see what kind of filter design fits to the target curves discussed in the previous section. Only the average of the curves in Figure 5.1 is covered as an example. The design for each position separately follows the same steps.

The averaged target curve for the equalization can be seen in Figure 5.2. As we can see the curve has no dips or peaks so it can be modelled 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 the same order than a finite impulse response (FIR) filter.

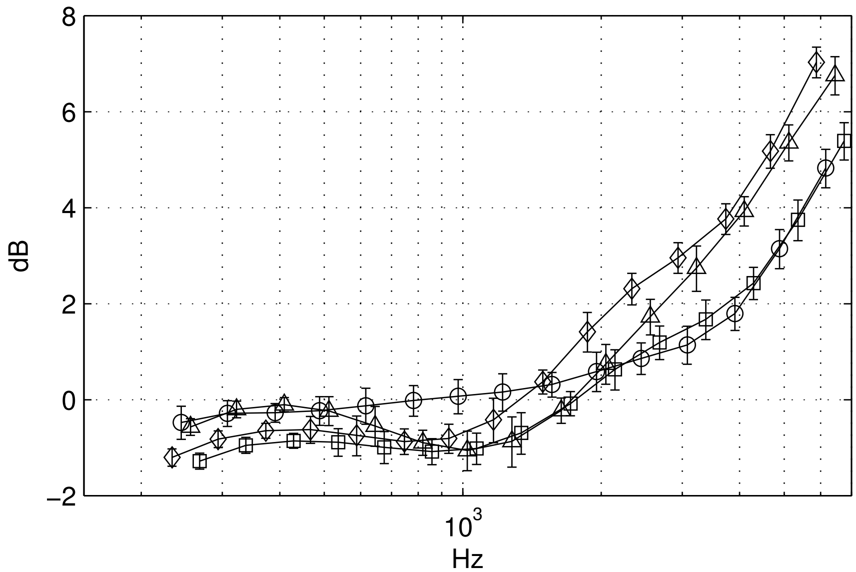


Figure 5.1: Smoothed differences of transfer functions of averaged test subjects and HATS from Figure 4.9. Microphone position 2-5 (○, ◊, □, △). The 95% confidential intervals are included in the figure.

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 5.2. The order three was the lowest design that stayed within ±1 dB from the target response.

5.2 Measurement vest

A dedicated 2 cm thick vest (model DS 0900) should be worn on B&K HATS 4128 in acoustical measurements. The covering should reduce the reflections from the torso so that it resembles better the 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 shoulders are an undesired side effect in those measurements.

The main problem with the torso simulator is the fact that the human upper body is neither hard nor soft completely in acoustical sense. In the shoulder area the vest can give adequately good simulation of the reflection but the situation near chest seems to be substantially different. The ribs and especially the breastbone seem to give a strong reflection directly beneath the clothing and skin. So the bare torso of the HATS could simulate in fact

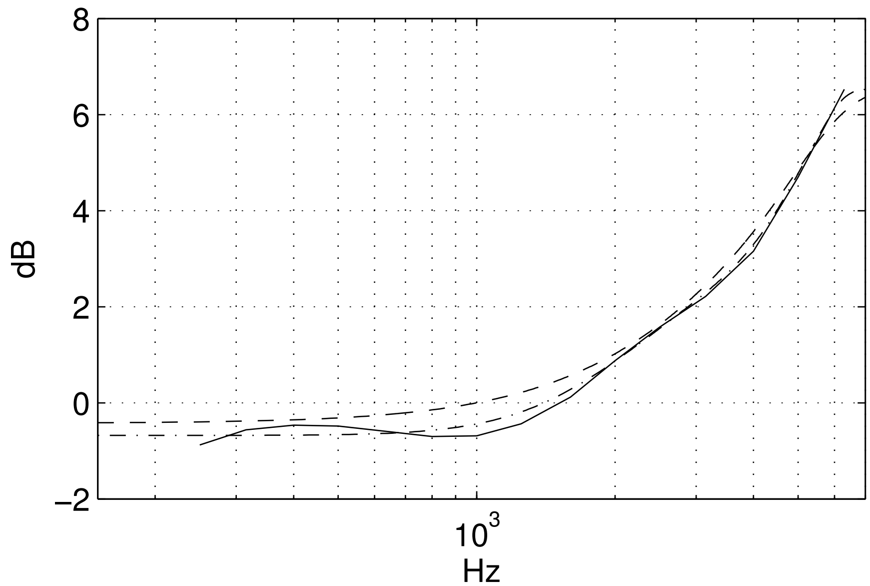


Figure 5.2: Target response (–) for the filter design to compensate the directivity of HATS to correspond to the directivity of averaged person near cheek. Target response is the average of the curves in Figure 5.1. Order 3 (- -) and 7 (- ·) IIR filter responses ($F_s = 32$ kHz).

better on chest than the torso with covering.

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 6 to 8 follow the same positions as the microphones in these phone accessories. By considering the measurement results for the position 1 and 7 in the Figures 4.6 and 4.7 we see significant differences between bare torso and the torso with vest.

First of all, the vest attenuates the reflections from the torso. This is seen for example in the far field position. The dip near 800 Hz is smaller because the cancelling reflection from chest is weaker.

More important issue is the responses to positions on the chest. The vest shifts the microphone the amount of its thickness from the hard surface of the torso. The time delay between the reflection and the direct sound is therefore larger and a dip is seen in response (see Figures 4.3 and 4.7). The vest in between attenuates the reflection but it does not cancel it.

When positions near chest are measured, should the vest be used or not? To study this the HATS was measured with the vest and without it. The results were compared as in Figure 4.10 to test subject measurements and they can be seen in Figure 5.3.

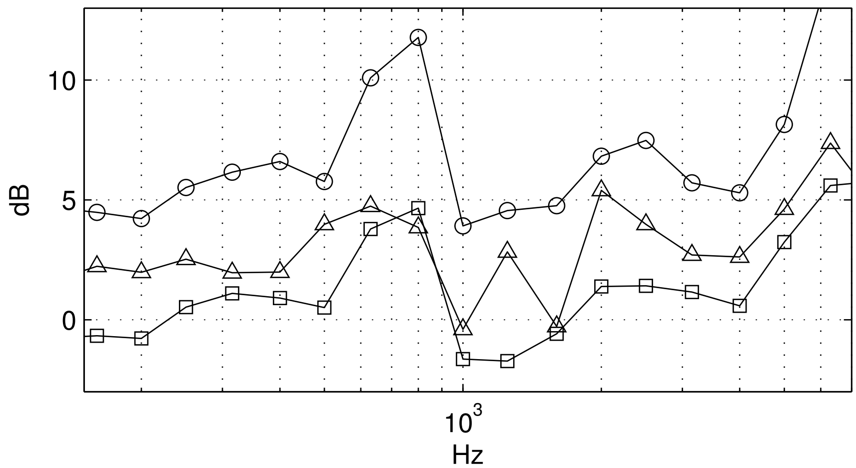


Figure 5.3: Difference of averaged test subjects speaking in normal speech volume and HATS. The comparison is done for transfer functions from microphone position 1 to 6-8 (○, □, △). A positive dB value implies that the HATS is more directional on that frequency (see Equation 4.8). The measurement vest was not woren on HATS.

If the general trend is considered in the curves (Figure 5.3) we see that the difference of the near chest reflection is omitted compared to Figure 4.10 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 an average person in high frequencies because of the difference of the mouth aperture size.

The final conclusion when Figures 4.10 and 5.3 are compared 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 headsets.

Chapter 6

Conclusion

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.

The backgrounds of the subject was discussed in Chapter 2. We saw that the current literature and the standardization are more or less inadequate to fulfil the open questions of the study objectives. One essential point is that there is just a couple of products on artificial mouth markets. Therefore the competition between the manufacturers does not drive the improvement of products and standardization.

A measurement system was gathered and a set of measurements was conducted to empirically see into the subject. The aim was to acquire a set of transfer functions to seven points near the cheek and the chest. Measurements were repeated for over ten test subjects and finally also for HATS.

To help the interpretation of the measurement results a theoretical modelling scheme was applied. Considering in parallel the measurements outcome, the modelling predicted surprisingly accurate the directivity characteristics. Two separate models were implemented: one for the bare head and one for the head and torso together.

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 position. 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 were 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.

Two improvement proposals were introduced in Chapter 5 to enhance the correspondence of the HATS measurements to an averaged person. An equalization scheme for the handset measurements was discussed in the first section of the chapter. 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 in 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.

The reliability of the results was ensured mainly by careful design of measurement setup and process. The measurement procedures were kept simple and the positioning of the microphones was done using a dedicated positioning device. The distribution of the speech data was assumed to be normal and so the 95% confidential intervals were obtained. Confidential intervals, coherences, and SNR implied that the measurements and the results were adequately reliable.

As we saw there are open questions on the field of the directivity of artificial mouths. The product that was studied in this thesis 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 in the thesis.

References

[1] ANSI. Determination of sound power levels of noise sources in anechoic and hemi-anechoic rooms. Technical Report ANSI Standard S12.35-1990 (R1996), American National Standards Institute, New York, U.S.A., 1996.

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

[3] 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*, 90:799–828, 1991.

[4] Bruce A. Bartlett. Tonal effects of close microphone placement. *J. of the Audio Engineering Society*, 29(10), Oct. 1981.

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

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

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

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

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

[10] D. Byrne et al. An international comparison of long-term average speech spectra. *J. of the Audio Engineering Society*, 96(4), Oct. 1994.

[11] 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*, 96:2088–2100, 1994.

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

[13] 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.

[14] James L. Flanagan. Analog measurements of sound radiation from the mouth. *J. of the Acoustical Society of America*, 32(12):1613–1620, Dec. 1960.

[15] James L. Flanagan. *Speech Analysis Synthesis and Perception*. Springer-Verlag, 2nd, expanded edition, 1972.

[16] Inc. HEAD acoustics. Head and Torso Simulators. http://www.headacoustics.com/telecom/products/hms.html.

[17] 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*, pages 183–186, Oct. 1999.

[18] IOtech inc. WaveBook Series, High-Speed Waveform Acquisition & Analysis, 2000. http://www.iotech.com/pdf/catpdf/wavebook.pdf.

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

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

[21] Pertti Laininen. *Tilastollisen analyysin perusteet*. Otatieto. Oy Yliopistokustannus University Press Finland Ltd. HYY-yhtymä, 2000.

[22] Eero Lampio. Pallokeinopään kehittäminen ja sen käyttö puhelimen akustisissa mittauksissa. Master's thesis, Teknillinen Korkeakoulu, 1961.

[23] John Laver. *Principles of Phonetics*. Cambridge University Press, 1994.

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

[25] F. S. McKendree. Directivity indices of human talkers in english speech. In *Proceedings - 1986 International Conference on Noise Control Engineering*, pages 911–916. Noise Control Foundation, Jul. 1986.

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

[27] Brüel & Kjær. Head and Torso Simulator (HATS) - Type 4128D, 2003. http://www.bksv.com/1667.htm.

[28] Esa Ranta, Hannu Rita, and Jari Kouki. *Biometria: Tilastotiedettä ekologeille*. Yliopistopaino, Helsinki, 7. painos edition, 1999.

[29] 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*, 100:537–554, 1996.

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

[31] Seppo Uosukainen. Akustiikan kenttäteoria -kurssin opetusmateriaali. Akustiikan ja Äänenkäsittelytekniikan laboratoria, TKK, Espoo, 2002.

[32] Kalevi Wiik. *Fonetiikan perusteet*. WSOY, 2nd edition, 1981.

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