Analysis of Directivity Factors and Indices of Human Speech

Samuel D. Bellows

A senior thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Bachelor of Science

Timothy W. Leishman, Advisor

Department of Physics and Astronomy

Brigham Young University

ABSTRACT

Analysis of Directivity Factors and Indices of Human Speech

Samuel D. Bellows
Department of Physics and Astronomy, BYU
Bachelor of Science

Directivity measurements of human speech reveal important characteristics of sound radiation and are useful in a variety of applications. Previous work in speech directivity has used different approaches; for this work, the directivity factors and indices of speech were measured using both a single and multiple-capture scanning system with human subjects. Analysis in the spherical harmonic domain helped to show important relationships between the different techniques as well as to simplify relations between measured frequency-response functions and the corresponding directivity factor and index functions. The high-resolution results show that while the directivity of human speech is omnidirectional at low frequencies, it becomes more directional at higher frequencies. Furthermore, diffraction lobes play a significant role in the directivity of human speech.

Keywords: Speech, Directivity, Voice, Microphone Array

ACKNOWLEDGMENTS

This research was supported by the College of Physical and Mathematical Sciences. A special thanks to Josh Bodon, Jenny Whiting, and Claire Pincock for the experimental work they carried out, Dr. William Strong for his valuable insights, and my advisor Dr. Timothy Leishman for his encouragement and support.

Parts of this thesis were presented at the 146th International Convention of the Audio Engineering Society in Dublin Ireland, March, 2019

Contents

Chapter 1

Introduction

A thorough understanding of the directivity of a sound source provides important insights into its radiation. In this work, frequency-dependent directivity factors and indices of human speech were derived from high-resolution sampling over a sphere and represented as functions of varying radiation axis as implied by Beranek [1]. While the common approach to the problem uses only a single principal axis of radiation, the general approach allows any designated axis, including but not limited to a traditional axis. While many sources of sound have single well-defined principal axes coinciding with maximum radiation, others do not because of complex radiation mechanisms, diffraction effects, variation over frequency, etc.

In this work, directivity factors and indices of human speech were measured and defined as functions of spherical-coordinate angles and frequency. The results were analyzed using various methods and shown to be more complex than previously assumed. The functions also accentuate directional details beyond those apparent from traditional polar and balloon plots.

Chapter 2

Speech Directivity Measurements

2.1 Previous Speech Directivity Measurements

Over the decades, numerous techniques for measuring speech directivity have been employed. Challenges inherent in measuring speech directivity include the often opposing demands of accurately measuring a dynamic field while also seeking a high spatial sampling density. Despite the best attempts to produce the same speech sounds, each human speech utterance will vary much more than that of an electronically excited apparatus such as a head and torso simulator (HATS) or a loudspeaker. Thus, in measuring the dynamic field, it is ideal to sample all spatial sampling points during the same capture. However, because of limited measurement equipment or other practical constraints, this leads to lower spatial sampling resolution than if measurements are carried out over multiple captures.

Because of the advantages of high-resolution, multiple capture measurements, various methods of relative calibration have been employed to compensate for differences in the speech utterances. In Dunn and Farnsworth's measurements of speech directivity [2], one microphone was used to measure 80 different positions and a second fixed microphone was used to ensure that the relative

power between speech repetitions was reasonable. Their technique required over 5,000 repetitions of a speech utterance in order to sample all spatial positions in the desired 1/3rd octave bands. Moreno and Pretzschner [3] also used a single calibration microphone along with a single measurement microphone in their multiple-capture measurements.

In order to increase spatial sampling resolution, circular microphone arrays began to be employed by researchers. McKendree [4] used a single arc with 30°azimuthal resolution, taking all measurements in a single capture. Instead of a single horizontal microphone array, Chu and Warnock [5] utilized two vertical arrays offset azimuthally by 90°, each having approximately 20° latitudinal resolution. By rotating the talker azimuthally 90° in 15° steps, they swept out a half-sphere with 141 unique sampling positions. In order to calibrate between captures, they used the microphone directly above the talker as a reference. Halkosaari et al. [6] measured near-field speech directivity with a helmet and fixed microphone positions. In order to compare between different measurement sets, they employed a transfer function relative to a reference position.

2.2 Measurement Technique

2.2.1 Measurement Apparatus

For this work, the directivites were measured using both a lower-resolution, single-capture system and a higher-resolution, multiple-capture system in an anechoic chamber. The first incorporated a constellation of microphones placed at the 32 vertices of a Catalan pentakis dodecahedron surrounding the seated talker at a nominal radius. The second employed a vertical semi-circular arc with 37 array microphones spaced at 5° polar-angle increments, beginning at 0°, directly above the talker, and ending at 180°, directly beneath. In order to sweep out the entire measurement sphere, the talker's chair was rotated azimuthally in 5° increments for a total of 2,522 unique sampling positions. Figure 2.1 illustrates the sampling positions and densities for both schemes.

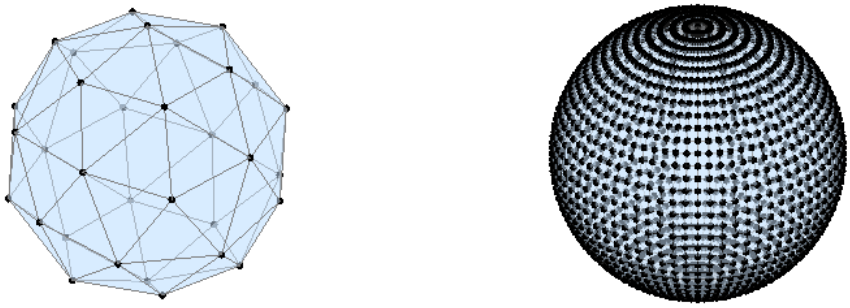

Figure 2.1 Single and multiple-capture sampling positions.

For both systems, each capture consisted of the participant speaking six phonetically balanced sentences. The talker's mouth was positioned consistently at the center of the measurement sphere with the aid of a head restraint on the chair. A reference microphone was also placed near his or her mouth to allow for compensation between takes using a frequency-response technique. For the multiple-capture system, nine talkers were recorded: four males and five females; for the single-capture system, two talkers were recorded: one male and one female.

2.2.2 Frequency-Response Technique

In order to compensate for differences in speech utterances between captures, the frequency response function between the reference microphone and each array microphone was calculated. First, the cross-spectral density between the reference microphone and each array microphone and the auto-spectral density of the reference microphone was calculated by averaging between windowed blocks of each speech utterance. The frequency response is given as the ratio of the cross-spectral density

and the auto-spectral density as

$$H(\theta_s, \phi_s, f) = \frac{G_{ab}(\theta_s, \phi_s, f)}{G_{aa}(f)}, \tag{2.1}$$

where the subscript s indicates the sampling position. Once the narrow-band frequency response was calculated, it was energetically summed into 1/3rd octave bands for further analysis.

Chapter 3

Analysis

3.1 Fourier Analysis on the Sphere

The spherical harmonics $\{Y_n^m\}_{n\in\mathbb{N}_0,|m|\leq n}$ of degree n and order m, defined as

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\theta,\phi)e^{im\phi}, \tag{3.1}$$

[7] form an orthonormal basis for $L^2(\mathbb{S}^2)$, the space of Lebesgue square-integrable functions on the unit sphere $\mathbb{S}^2$; thus, any function $f \in L^2(\mathbb{S}^2)$ may be written as a linear combination of spherical harmonics:

$$f = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}(f)_n^m Y_n^m. \tag{3.2}$$

The coefficients $\{(f)_n^m\}_{n\in\mathbb{N}_0,|m|\leq n}$ are given by the inner product of the function f with the spherical harmonic Y_n^m

$$(f)_n^m = \langle f, Y_n^m\rangle = \int_{\mathbb{S}^2} f Y_n^{m*} d\Omega, \tag{3.3}$$

where $*$ denotes complex conjugation. The spherical-harmonic expansion is a Fourier mapping $\mathcal{F}: L^2 \to \ell^2$, which maps Lebesgue square-integrable functions to square-summable sequences.

The mapping is isometric, in that the norms in both spaces are equal:

$$||f||_{L^2} = ||(f)_n^m||_{\ell^2}, \tag{3.4}$$

where

$$||f||_{L^2} = \langle f, f^* \rangle^{1/2} \tag{3.5}$$

and

$$||(f)_n^m||_{\ell^2} = (\sum_{n=0}^{\infty} \sum_{m=-n}^{n} |(f)_n^m|^2)^{1/2}. \tag{3.6}$$

This gives Parseval's relation:

$$\int_{\mathbb{S}^2} |f|^2 d\Omega = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} |(f)_n^m|^2. \tag{3.7}$$

To facilitate the use of the spherical harmonics with operators on $L^2(\mathbb{S}^2)$, the double indices n and m can be rewritten as a single index ℓ using the bijection given in [8]:

$$\ell = n(n+1) + m, \tag{3.8}$$

thus allowing the spherical harmonics to be written as $\{Y_\ell\}_{\ell=0}^{\infty}$.

3.1.1 Operators

The use of operators on $L^2(\mathbb{S}^2)$ assists with the analysis of the functions which are elements of this space. Let $\mathcal{B}$ denote a bounded linear operator that maps $f \in L^2(\mathbb{S}^2)$ to $g \in L^2(\mathbb{S}^2)$. Because $L^2(\mathbb{S}^2)$ is a separable Hilbert space, $\mathcal{B}$ admits an infinite matrix representation $\mathbf{B}$, with elements $b_{n,p}^{m,q}$ that describe the relation between the input direction Y_p^q and output direction Y_n^m [8]. The output row index is given from the single index conversion of n, m and the input column index is given from the single-index conversion of p, q. If $(f)_n^m$ are the spherical harmonic coefficients of f and $(g)_n^m$ are the spherical harmonic coefficients of g, $\mathcal{B}$ can be written as [8]

$$(g)_n^m = \sum_{p=0}^{\infty} \sum_{q=-p}^{p} b_{n,p}^{m,q} (f)_p^q. \tag{3.9}$$

Spectral Truncation Operator

The spectral truncation operator of degree N projects the spherical harmonic coefficients onto the degree-limited space $\mathcal{H}_N$, truncating the spherical harmonic expansion to degree N. The operator's matrix elements are given by

$$b_{n,p}^{m,q} = \begin{cases} \delta_{n,p}^{m,q} & n \in \{0,1,\ldots,N\},\ m \in \{-n,-n+1,\ldots,n\} \\ 0 & otherwise \end{cases}, \tag{3.10}$$

where δ is the Kronecker delta [8]. The use of this operator spatially smooths functions by eliminating high spatial frequencies and can reduce measurement noise in those frequencies, much as a low-pass filter can eliminate undesired high-frequency noise in the time-domain.

Rotation Operator

The rotation operator $\mathcal{D}(\varphi,\vartheta,\omega)$ rotates a function using Euler angles in the *zyz* convention, where ω is the first rotation about the z axis, ϑ is the second rotation about the y axis, and φ is the final rotation about the z axis. The operator's matrix elements are given by

$$d_{n,p}^{m,q} = D_{m,q}^{n}(\varphi,\vartheta,\omega)\delta_{n,p}, \tag{3.11}$$

where $D_{m,q}^{n}$ is the Wigner D-matrix [8].

Symmetry Operator

A function f is symmetric about the xz plane ($\phi = 0$) if $f(\theta,\phi) = f(\theta,-\phi)$. Using spherical harmonic coefficients $(f)_n^m$ given by 3.3,

$$(f)_n^m = \int_{\mathbb{S}^2} f(\theta,\phi) Y_n^{m*}(\theta,\phi) d\Omega, \tag{3.12}$$

The symmetry relation yields

$$(f)_n^m = \int_{\mathbb{S}^2} f(\theta,-\phi) Y_n^{m*}(\theta,\phi) d\Omega. \tag{3.13}$$

Reversing the direction of integration in the ϕ direction and using the relations $Y_n^m(\theta,-\phi) = Y_n^{m*}(\theta,\phi)$ and $Y_n^m(\theta,\phi) = (-1)^m Y_n^{-m*}(\theta,\phi)$, one finds that

$$\begin{aligned}(f)_n^m &= \int_{\mathbb{S}^2} f(\theta,\phi) Y_n^{m*}(\theta,-\phi) d\Omega \\ &= \int_{\mathbb{S}^2} f(\theta,\phi) Y_n^m(\theta,\phi) d\Omega \\ &= \int_{\mathbb{S}^2} f(\theta,\phi)(-1)^m Y_n^{-m*}(\theta,\phi) d\Omega \\ &= (-1)^m (f)_n^{-m}.\end{aligned} \tag{3.14}$$

For functions that are not symmetric about the *xy* plane, an average about the plane is performed in the spherical harmonic domain using

$$(f_s)_n^m = \frac{1}{2}[(f)_n^m + (-1)^m (f)_n^{-m}], \tag{3.15}$$

where $(f_s)_n^m$ are the symmetrized coefficients.

The operator's matrix elements are given by

$$b_{n,p}^{m,q} = \frac{1}{2}[\delta_{n,p}^{m,q} + (-1)^m \delta_{n,p}^{m,-q}]. \tag{3.16}$$

In order to symmetrize about any arbitrary plane, the rotation operator $\mathcal{D}$ may be applied before the symmetry operator.

3.1.2 Computation of Spherical Harmonic Coefficients

For this work, the least-squares method was employed to compute the spherical harmonic coefficients from the discretely sampled sphere. The least-squares method assumes that the value of a function f at a point $\Omega_s \in \mathbb{S}^2$ is a linear combination of the infinite-dimensional basis vectors $\{Y_\ell\}_{\ell=0}^{\infty}$.

$$f(\Omega_s) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (f)_n^m Y_n^m(\Omega_s) = \sum_{\ell=0}^{\infty} (f)_\ell Y_\ell(\Omega_s) \tag{3.17}$$

If f is band-limited to degree N ($f \in \mathcal{H}_N$), the series may be truncated at degree N. If the function is sampled at S points, a system of S equations in $(N+1)^2 = L$ variables may be formulated as

$$\mathbf{f} = \mathbf{Ya}, \tag{3.18}$$

where

$$\mathbf{f} = \begin{bmatrix} f(\Omega_1) \\ f(\Omega_2) \\ \vdots \\ f(\Omega_S) \end{bmatrix}, \tag{3.19}$$

$$\mathbf{Y} = \begin{bmatrix} Y_0(\Omega_1) & Y_1(\Omega_1) & \dots & Y_L(\Omega_1) \\ Y_0(\Omega_2) & Y_1(\Omega_2) & \dots & Y_L(\Omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_0(\Omega_S) & Y_1(\Omega_S) & \dots & Y_L(\Omega_S) \end{bmatrix}, \tag{3.20}$$

and

$$\mathbf{a} = \begin{bmatrix} (f)_0 \\ (f)_1 \\ \vdots \\ (f)_L \end{bmatrix}. \tag{3.21}$$

The lowest possible bound for the system to be well-conditioned is when $S = L$; however, in most cases S must be much greater than L [9]. Thus, for a configuration with S sampling points, the upper-bound for the degree of the coefficients that can be assured to be exact is $\sqrt{S} - 1$. However, the positioning of the sampling points, especially when not distributed uniformly on the sphere, may result in decreased degree. For the single-capture method, with $S = 32$, the upper-bound is

$N = 4$, for the multiple-capture method, with $S = 2522$, the upper-bound is $N = 49$. For this work, the single-capture results were expanded to $N = 4$ and the multiple-capture results were expanded to $N = 10$.

3.1.3 Error on the Sphere

A continuous mean-square deviation between two functions f and g with spherical harmonic expansion coefficients $(f)_n^m$ and $(g)_n^m$ on the sphere can be defined as

$$E[f,g] = \left[\frac{1}{4\pi}\int_{\mathbb{S}^2} |f-g|^2 d\Omega\right]^{1/2} \tag{3.22}$$

Using the relation

$$|f-g|^2 = (f-g)(f^*-g^*) \tag{3.23}$$

The error can be written as

$$\begin{aligned} E[f,g] &= \left[\frac{1}{4\pi}\int_{\mathbb{S}^2} ff^* + gg^* - fg^* - gf^* d\Omega\right]^{1/2} \\ &= \left[\frac{1}{4\pi}(\langle f,f\rangle + \langle g,g\rangle - \langle f,g\rangle - \langle g,f\rangle)\right]^{1/2} \\ &= \left[\frac{1}{4\pi}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} |(f)_n^m|^2 + |(g)_n^m|^2 - (f)_n^m (g)_n^{m*} - (g)_n^m (f)_n^{m*}\right]^{1/2} \end{aligned} \tag{3.24}$$

and expressed on a decibel scale as

$$E_{dB} = 20\log_{10}(1+E). \tag{3.25}$$

3.2 Wave-Equation and Source-Centering

The use of the wave equation and its general solution in spherical polar coordinates allows for the analysis of the acoustic pressure field beyond the measurement sphere. The solution to the time independent wave equation for an outward propagating wave is

$$p(r,\theta,\phi,k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} C_n^m(k) h_n^{(2)}(kr) Y_n^m(\theta,\phi), \tag{3.26}$$

where $h_n^{(2)}$ is the spherical Hankel function of order n of the second kind and $C_n^m(k)$ are the pressure expansion coefficients at a given wave number k. The pressure expansion coefficients $C_n^m(k)$ are closely related to the spherical harmonic coefficients $(f)_n^m$ of a given k by

$$C_n^m(k) = \frac{(f)_n^m}{h_n^{(2)}(kR)}, \tag{3.27}$$

where R is the measurement sphere radius.

For further analysis using the wave equation, it is desirable that the acoustic center of the source be located at the origin. However, due to measurement imperfections and the frequency dependence of the acoustic center [10], the source center may not be located at the origin. The source centering algorithm presented by Hagai et al. uses a translation matrix which operates on the pressure expansion coefficients in the measured frame of reference and provides them in the new frame of reference with origin (r'', θ'', ϕ'') [10]. However, because of errors introduced from the truncated spherical harmonic expansion series [10], a direct expansion from the new reference frame may have less error.

The pressure expansion coefficients $C_n^m(k)$ are then given by solving the following system of equations:

$$\mathbf{p} = \mathbf{H}\mathbf{c}, \tag{3.28}$$

where

$$\mathbf{H} = \begin{bmatrix} h_0^{(2)}(kr_1')Y_0^0(\theta_1',\phi_1') & h_1^{(2)}(kr_1')Y_1^{-1}(\theta_1',\phi_1') & \dots & h_N^{(2)}(kr_1')Y_N^N(\theta_1',\phi_1') \\ h_0^{(2)}(kr_2')Y_0^0(\theta_2',\phi_2') & h_1^{(2)}(kr_2')Y_1^{-1}(\theta_2',\phi_2') & \dots & h_N^{(2)}(kr_2')Y_N^N(\theta_2',\phi_2') \\ \vdots & \vdots & \ddots & \vdots \\ h_0^{(2)}(kr_S')Y_0^0(\theta_S',\phi_S') & h_1^{(2)}(kr_S')Y_1^{-1}(\theta_S',\phi_S') & \dots & h_N^{(2)}(kr_S')Y_N^N(\theta_S',\phi_S') \end{bmatrix}, \tag{3.29}$$

$$\mathbf{c} = \begin{bmatrix} C_0^0(k) \\ C_{-1}^1(k) \\ \vdots \\ C_N^N(k) \end{bmatrix}, \tag{3.30}$$

and

$$\mathbf{p} = \begin{bmatrix} p(r_1', \theta_1', \phi_1') \\ p(r_2', \theta_2', \phi_2') \\ \vdots \\ p(r_S', \theta_S', \phi_S') \end{bmatrix}, \tag{3.31}$$

where the primed coordinates (r', θ', ϕ') are the coordinates of the measurement system relative to the translated origin (r'', θ'', ϕ'').

After computing the expansion coefficients $C_n^m(k)$ in a new reference frame, an objective function designed to minimize the energy in high-degree terms is computed. By minimizing the objective function with respect to position, one arrives at an estimate of the acoustic center. For this work, the objective function J_2 defined in [11] was used, where J_2 is given as

$$J_2 = \frac{\sum_{n=0}^{N} \sum_{m=-n}^{m} n |C_n^m(k)|^2}{\sum_{n=0}^{N} \sum_{m=-n}^{m} |C_n^m(k)|^2}. \tag{3.32}$$

This objective function was chosen because it was found to give convex patterns [11], necessary to find a minimum.

3.3 Directivity Factor and Index

The directivity factor Q was defined by Beranek as "the ratio of the intensity on a designated axis of a sound radiator at a stated distance r to the intensity that would be produced at the same position by

a point source if it were radiating the same total acoustic power as the radiator" [1, p. 109]. If one assumes far-field radiation, such that the angular dependence of the directivity factor is no longer dependent on the distance r, the directivity factor may be found from the far-field relation between acoustic pressure and time-averaged intensity

$$\langle I(\theta,\phi,f)\rangle_t = \frac{|p(\theta,\phi,f)|^2}{2\rho_0 c}. \tag{3.33}$$

The time-averaged acoustic power of the source is then given as

$$\langle W\rangle_t = \int_{\mathbb{S}^2} \langle \vec{I}\rangle_t \cdot \vec{n} dS. \tag{3.34}$$

Thus, in terms of the frequency response function H, a directivity factor function may be written as

$$Q(\theta,\phi,f) = \frac{4\pi|H(\theta,\phi,f)|^2}{\int_0^{2\pi}\int_0^{\pi}|H(\theta,\phi,f)|^2\sin\theta d\theta d\phi}. \tag{3.35}$$

Writing writing the frequency response function H with spherical harmonic expansion coefficients a_n^m,

$$H(\theta,\phi,f) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_n^m(f)Y_n^m(\theta,\phi, \tag{3.36}$$

the directivity factor function may be further simplified to the form

$$Q(\theta,\phi,f) = |\sum_{n=0}^{\infty}\sum_{m=-n}^{n} \tilde{a}_n^m(f)Y_n^m(\theta,\phi)|^2, \tag{3.37}$$

where the power-normalized coefficients $\tilde{a}_n^m$ are given as

$$\tilde{a}_n^m(f) = a_n^m(f)\sqrt{\frac{4\pi}{\sum_{n=0}^{\infty}\sum_{m=-n}^{n}|a_n^m(f)|^2}}. \tag{3.38}$$

Using these same coefficients to write the power-normalized frequency response

$$\tilde{H} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \tilde{a}_n^m(f) Y_n^m(\theta, \phi), \tag{3.39}$$

the directivity factor function is given as the magnitude-squared response.

$$Q = |\tilde{H}|^2. \tag{3.40}$$

The directivity index function expresses the result on a decibel scale as

$$DI(\theta, \phi, f) = 10 \log_{10} Q(\theta, \phi, f) = 20 \log_{10} |\tilde{H}(\theta, \phi, f)|. \tag{3.41}$$

3.4 Selection of Axis

Because directivity factors and indices are typically expressed as functions of frequency for only a single designated axis, two approaches were used to condense the information contained in the angular dependence: on-axis directivity and maximum-axis directivity.

On-axis directivity provides the directivity factor and index along a single axis directly in front of the talker's mouth. However, because of human speech radiation, diffraction, and absorption characteristics, this is not always the axis of maximum radiation. The maximum-axis directivity thus traces the direction and value of maximum radiation over frequency.

Chapter 4

Results

4.1 Source Centering

The objective function J_2 (Eq. 3.32) indicated that the acoustic center of the measured directivity at low frequencies was near the array origin. Across the frequency range of 160 Hz to 250 Hz, the average displacement of the acoustic source center and the origin of the array was found to be less than 1 cm. Figure 4.1 shows the plotted objective function J_2 for a female talker in various planes. For this specific frequency the source center was located at (0.5 cm, 0 cm, 0 cm) in 0.5 cm increments. Thus, for these measurements, the effects of source centering were minimal.

4.2 Symmetrizing

Because of the anatomy of the human body, symmetric results about the xz plane are expected for speech directivity. Using this expectation, some researchers, only measured half of the possible azimuthal positions [5], [12]. When measurements have been made over the entire sphere, the results of both hemispheres may be averaged and symmetrized using the symmetry operator.

Figure 4.2 shows the $1/3^{rd}$ octave measured directivity factor function of a male talker at the

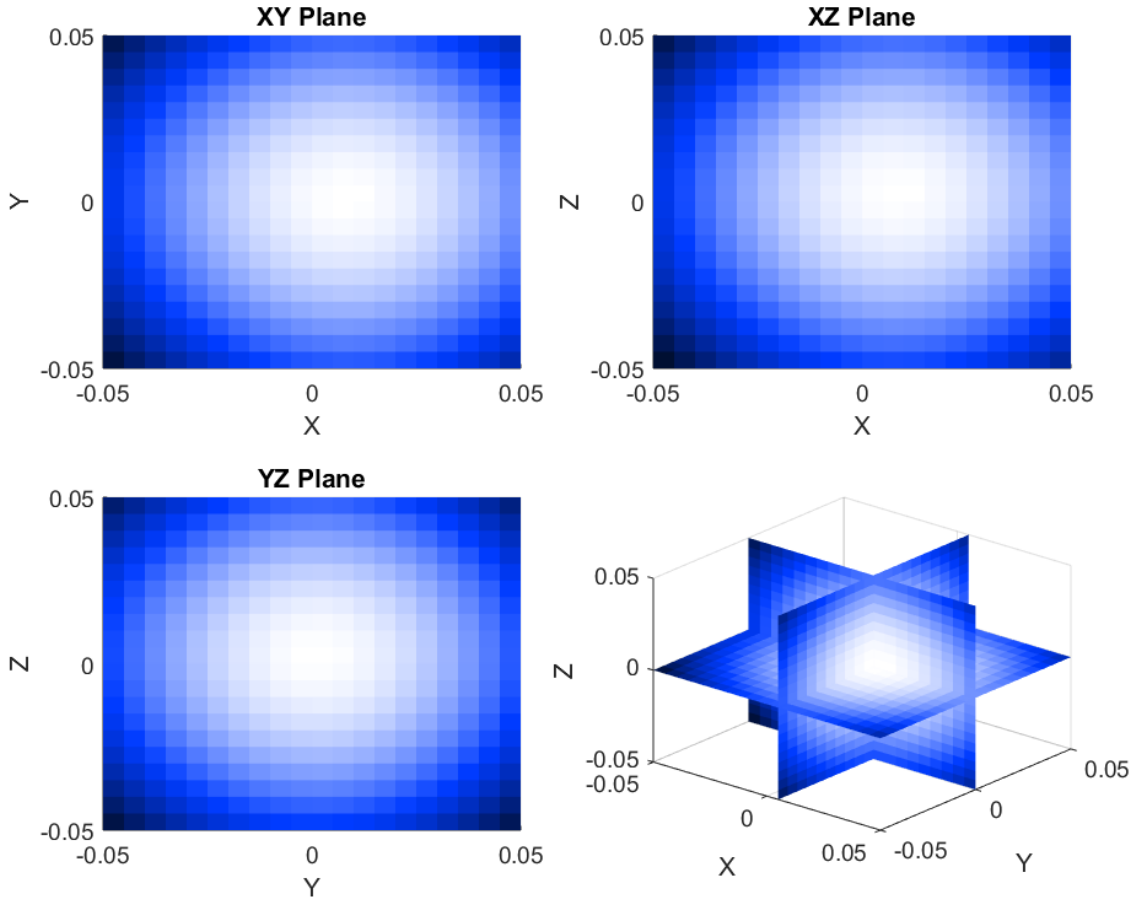

Figure 4.1 The J_2 objective function for female talker at the 250 Hz $1/3^{rd}$ octave band.

200 Hz center frequency. The results are presented as both balloon and corresponding polar plots in the transverse, frontal, and median planes. The polar plots are also overlaid on the balloon for clarity. Significantly, these plots should not be confused with typical directivity balloon and polar plots as they represent the function defined in Eq. 3.35. The results show quasi-symmetry in the transverse and frontal plane. Figure 4.3 shows the directivity factor function after the symmetric operator has been applied; the averaging between sides results in true symmetry in the transverse and frontal planes.

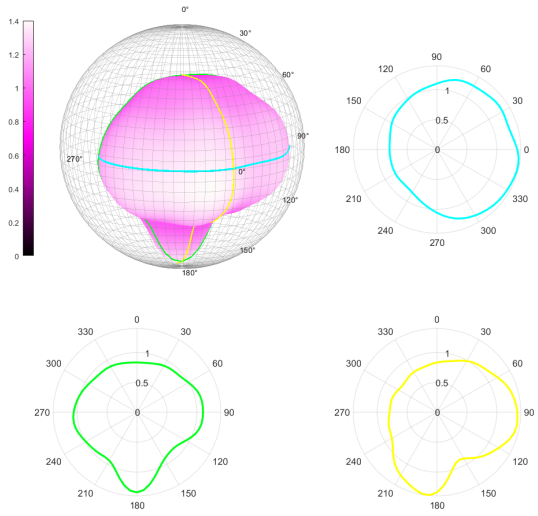

Figure 4.2 Male talker directivity factor function for 200 Hz 1/3rd octave band before symmetrizing.

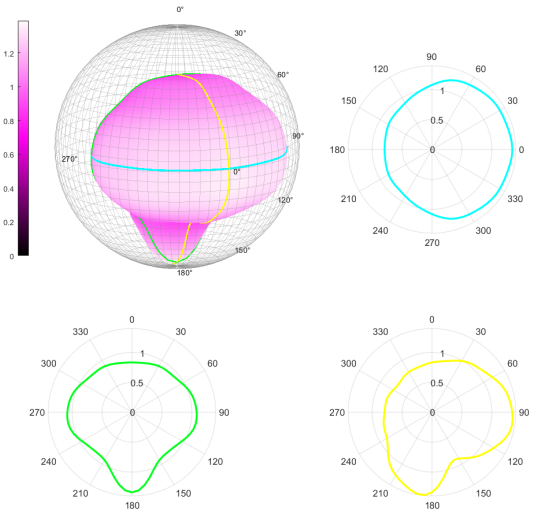

Figure 4.3 Male talker directivity factor function for 200 Hz 1/3rd octave band after symmetrizing.

4.3 High-Resolution Results

The high-resolution directivity factor and index functions of human speech were shown to vary significantly in detail from previous assumptions. As anticipated, they tended to be omnidirectional at lower frequencies, but had complex patterns at higher frequencies that contrasted with idealizations, such as nearly hemispherical radiation. In some cases the latter may be a useful approximation, but the actual patterns are much more intricate.

Figure 4.4 shows the symmetrized directivity factor function of a male talker in the 125 Hz 1/3rd octave band. As expected for lower frequencies, the source is nearly omnidirectional, with a value near 1 over the entire sphere.

Figure 4.5 shows a female talker's directivity factor function in the 400 Hz 1/3rd octave band, illustrating radiation patterns occuring around her head and body in the transverse and frontal planes. The pronounced effects of lobes in the 1.25 kHz 1/3rd octave band are visible in Figure 4.6, showing concentrated intensity approximately 30° above the talker's mouth. At the 3.15 kHz 1/3rd octave

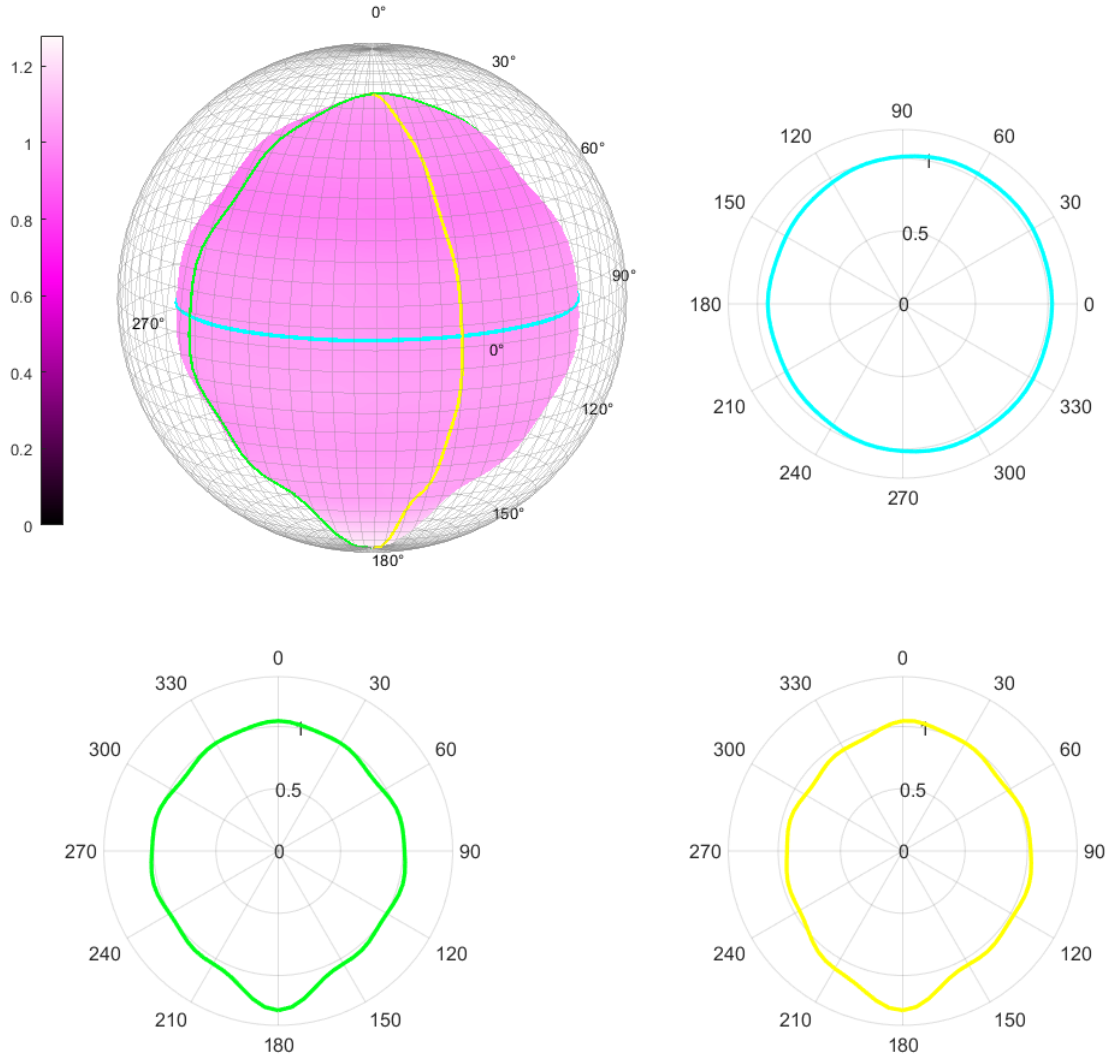

Figure 4.4 Male talker's symmetrized directivity factor function at the 125 Hz 1/3rd octave band.

band for a male talker, three lobes can be seen in the median plane, two with directivity factors near 4, illustrated in Figure 4.7.

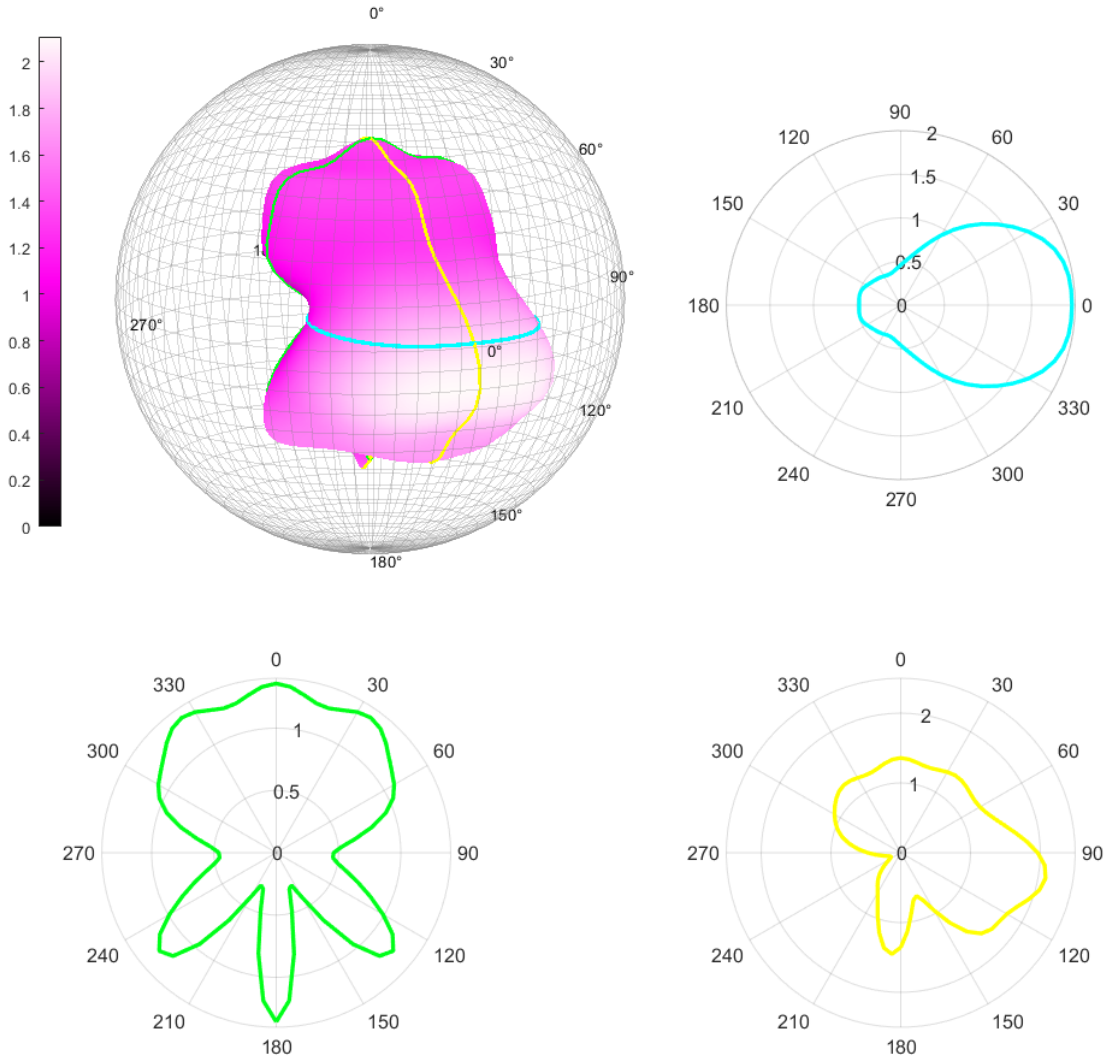

Figure 4.5 Female talker's symmetrized directivity factor function at the 400 Hz 1/3rd octave band.

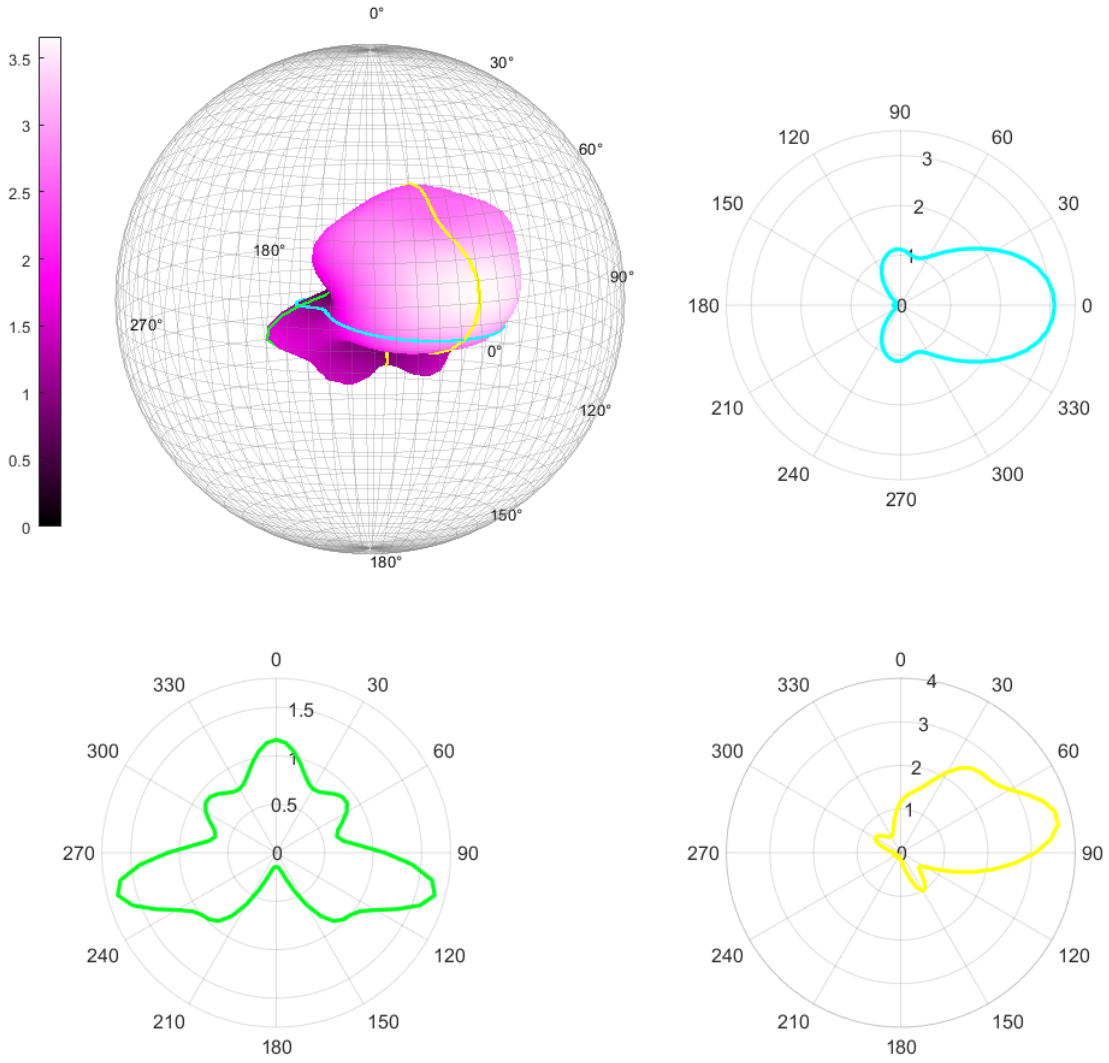

Figure 4.6 Male talker's symmetrized directivity factor function at the 1.25 kHz 1/3rd octave band.

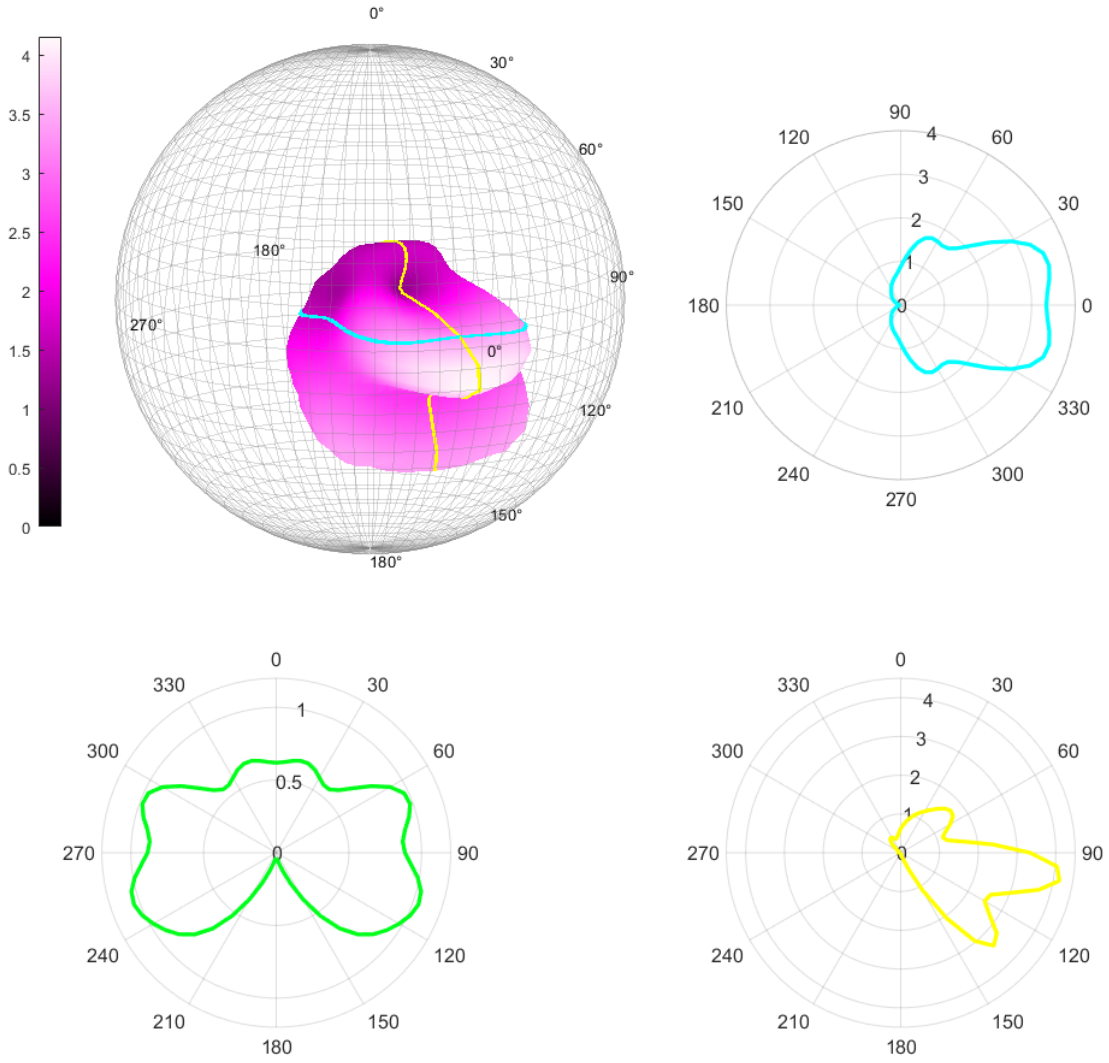

Figure 4.7 Male talker's symmetrized directivity factor function at the 3.15 kHz 1/3rd octave band.

Figure 4.8 shows the on-axis directivity factor and index values derived from a symmetrized average of six talkers, three male and three female. Because of shifting directivity lobes and nulls, the factor oscillates over frequency as they pass through the axis directly in front of the mouth. By way of comparison, Figure 4.9 illustrates the maximum-axis directivity factor. Because the point of maximum directivity factor is no longer located directly in front of the talker's mouth, the colatitude and azimuthal position have been traced outward in the polar plots with increasing frequency. As frequency increases, the radiation is also seen to become more directional as expected.

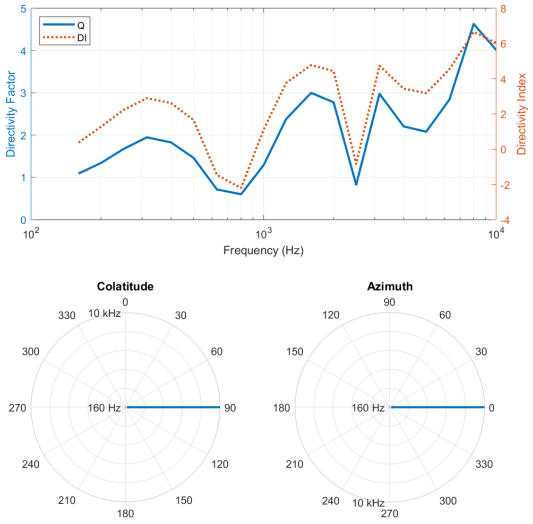

Figure 4.8 Fixed-axis average talker directivity factor and index.

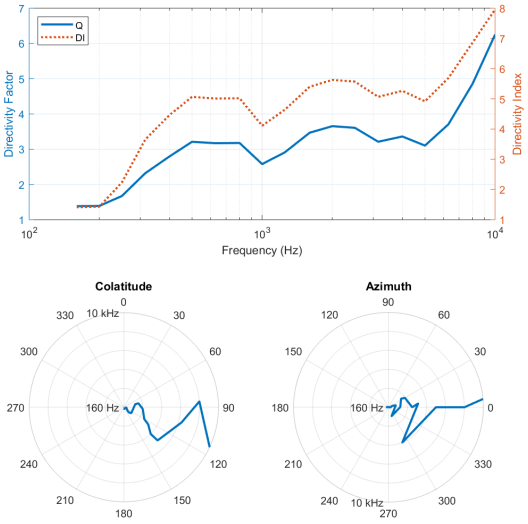

Figure 4.9 Maximum-axis average talker directivity factor and index.

4.4 Comparison to Lower-Resolution Results

Comparing the high-resolution multiple-capture results to lower-resolution single-capture results emphasizes not only consistency and repeatability across measurement systems, but also the potential benefits of the high-resolution system. Figure 4.10 shows the a male talker's directivity factor function at the 200 Hz 1/3rd octave band measured using both the single-capture and multiple-capture systems. To assist in the visualization of the different resolutions, the sampling grids of both methods have been overlaid in gray. Furthermore, each balloon's mesh has been generated by creating planar faces that connect sampling points, and both balloons are shown with interpolated color over the planar mesh faces. As is evident from the overlaid sampling grids, the single-capture measurement lacks fine details but preserves gross directivity-factor trends.

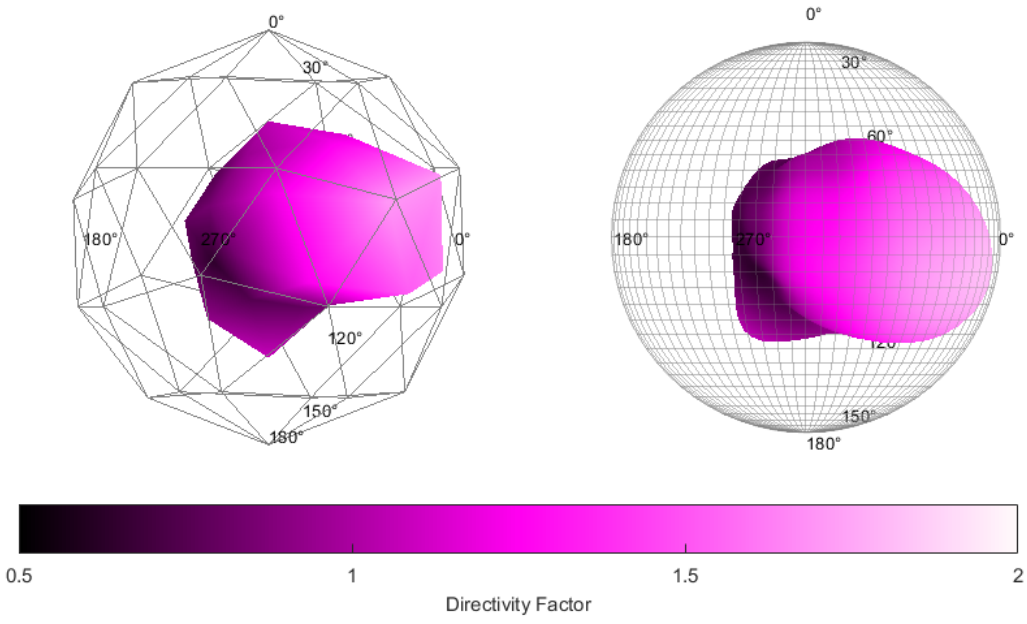

Figure 4.10 Symmetrized male talker's directivity factor function in the 200 Hz 1/3rd octave band measured using a single and multiple capture system.

Figure 4.11 plots the maximum-axis directivity factor for the same male talker, the solid line for the single-capture measurement and the dashed line for the multiple-capture measurement. This result reaffirms that while there are differences between high-resolution multiple-capture measurements and single-capture measurements, both systems reveal similar trends, both in magnitude and direction of the frequency-dependent directivity of human speech.

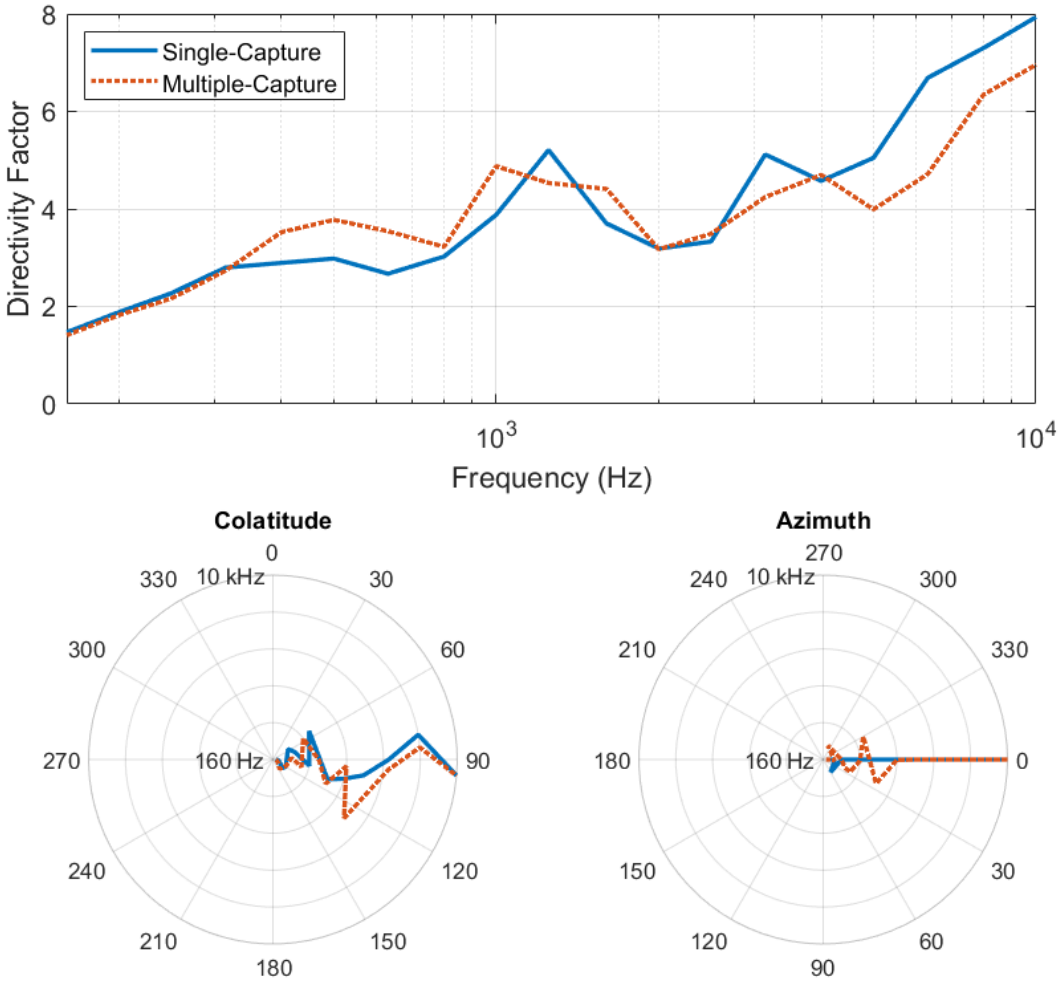

Figure 4.11 Single and multiple-capture system directivity factor maximum-axis values.

In particular, both systems demonstrate that speech is much more directional than the hemispherical assumption often used, which would result in a constant directivity factor $Q = 2$ across all frequencies. As illustrated in the measurements, speech directivity tends to be omnidirectional at low frequencies, corresponding to $Q = 1$. However, by 500 Hz, $Q > 2$ and by 1 kHz the directivity factor takes on increasingly higher values.

Another revealed feature is the effect of diffraction lobes. Instead of a constant increase across frequency, undulations can be seen in the directivity factor of Fig. 4.11 as areas of constructive and destuctive interference form due to diffraction around the body, evidenced by peaks in the directivity factor near 400 Hz, 1.25 kHz, and 3 kHz. The frequencies at which these lobes occur are not unique to this specific talker: these lobes can be identified for the other talkers shown in Figs. 4.5, 4.6, and 4.7. The polar plot representing the median plane in Fig. 4.5 shows a lobe forming just below the talker's mouth in the 400 Hz 1/3rd octave band. The lobes forming in Figs. 4.6 and 4.7 were previously mentioned and occurred in the 1.25 kHz and 3.15 kHz 1/3rd octave bands respectively. While men and women have different fundamental frequencies and unique vocal timbres, the diffraction is primarily depedent on the anatomical features of the seated talkers and the chair, resulting in greater similarities at the frequencies for which the lobes occur. Above these frequencies, speech becomes increasingly directional and concentrated in front of the talker's mouth, and some diffraction effects become less pronounced.

Chapter 5

Errors and Limitations

5.1 Effect of Symmetrizing

The use of the symmetry operator can be an important means of validation of directivity measurements. For symmetric radiation patterns, such as those generated by human speech, there should be no deviation from one hemisphere to the other; therefore, the repeatability and reliability of measurements can be gauged by the amount of symmetry present. Figure 5.1 shows the deviation from symmetrizing the frequency response functions from 160 Hz to 10 kHz for both a KEMAR HATS and the average of six human talkers, using Eq. referr. Because the HATS is electronically excited, the directivity patterns it produces should be more repeatable and therefore more symmetric than those generated by the human subjects. For both the average talker and the HATS, the error from symmetrizing remained less than 1 dB across the given frequency range, indicating good agreement between measured hemispheres.

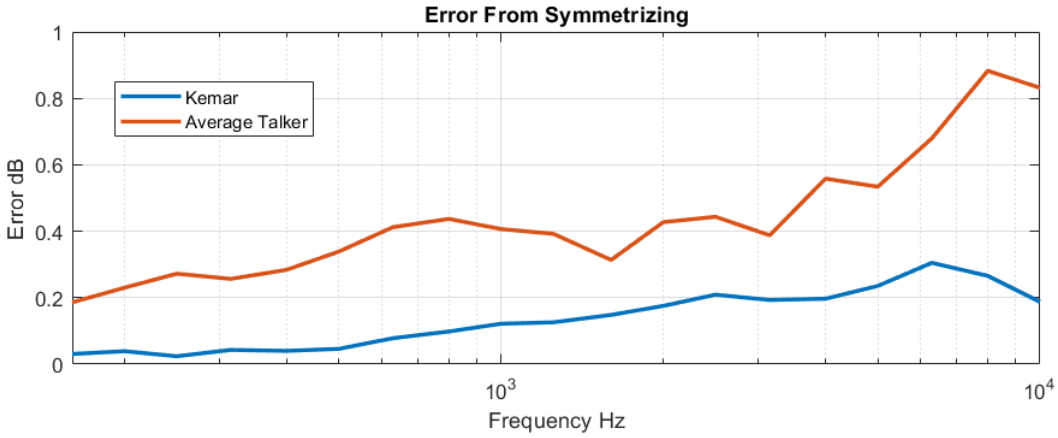

Figure 5.1 Deviation from the symmetric operator applied to the frequency-response functions of a KEMAR HATS and of the average of six human talkers.

5.2 Individual Variations

Deviations between individuals and the average of six talkers were also observed. Figure 5.2 plots the deviation of the frequency-response functions over frequency between the six individual talkers to the average of all six. In general, the deviation is near 0.5 dB at low frequencies and centers around 1.0 dB as frequency increases. As noted in [13], part of the variation at higher frequencies is due to the effects of diffraction caused by the individual characteristics of the talkers, such as anatomical features.

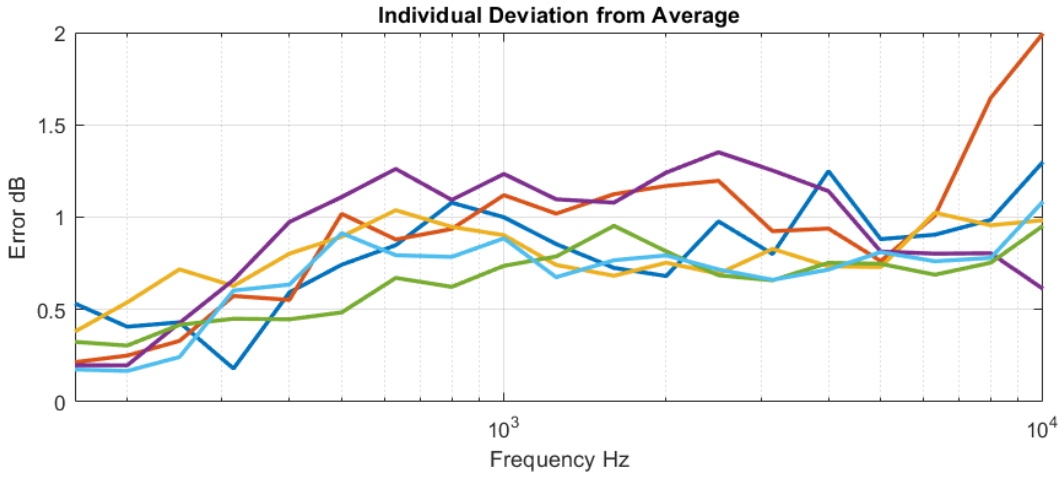

Figure 5.2 Variation between individual talkers and their average.

Chapter 6

Conclusions

The directivity factor and index functions reveal important characteristics of the radiation of human speech. In order to measure these function, both a single and multiple-capture system were employed; analysis and comparison in the spherical harmonic domain demonstrated significant similarities between both measurement sets. The results show that the radiation of the human voice is more complex than previously assumed, revealing that speech is roughly omnidirectional at lower frequencies and progressively more directional at higher frequencies. Because of directivity effects, such as shifting lobes, different approaches to selecting an axis of radiation for calculating the directivity factor and index of a source yield different values. Because of individual variations between talkers, a larger data set would be useful in continuing to study vocal directivity. Future measurements made with higher resolution and over a larger sampling population would assist in continuing to understand the spatial and frequency relationships of the directivity of human speech.

Bibliography

[1] L. L. Baranek, *Acoustics* (Acoustical Society of America, New York, 1986).

[2] H. Dunn and D. W. Farnsworth, "Exploration of Pressure Field Around the Human Head During Speech," Journal of the Acoustical Society of America **10,** 184–199 (1939).

[3] A. Moreno and J. Pretzschner, "Human Head Directivity in Speech Emission: A New Approach," Acoustics Letters **1,** 78–84 (1978).

[4] F. S. McKendree, "Directivity Indices of Human Talkers in English Speech," Paper presented at Inter-Noise 86 Conference: Progress in Noise Control, Cambridge, USA pp. 911–916 (1986).

[5] W. T. Chu and A. C. Warnock, "Detailed Directivity of Sound Fields Around the Human Head During Speech," National Research Council Canada **IRC-RR-104,** 1–47 (2002).

[6] T. Halkosaari, M. Vaalgamaa, and M. Karjalainen, "Directivity of Artificial and Human Speech," Journal of the Audio Engineering Society **53,** 620–631 (2005).

[7] T. M. Dunster, in *NIST Handbook of Mathematical Functions*, F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, eds., (Cambridge University Press, New York, 2010).

[8] R. A. Kennedy and P. Sadeghi, *Hilbert Space Methods in Signal Processing* (Cambridge University Press, 2013).

[9] B. Rafaely, *Fundamentals of Spherical Array Processing* (Springer-Verlag, Berlin Heidelberg, 2015), Vol. 8.

[10] I. B. Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic Centering of Sources Measured by Surrounding Spherical Microphone Arrays," Journal of the Acoustical Society of America **130,** 2003–2015 (2011).

[11] N. R. Shabtai and M. Vorländer, "Acoustic Centering of Sources with Higher-Order Radiation Patterns," Journal of the Acoustical Society of America **137,** 1947–1961 (2015).

[12] B. B. Monson, E. J. Hunter, and B. H. Story, "Horizontal Directivity of Low- and High-Frequency Energy in Speech and Singing," Journal of the Acoustical Society of America **132,** 433–441 (2012).

[13] S. D. Bellows and T. W. Leishman, "High-Resolution Analysis of the Directivity Factor and Directivity Index Functions of Human Speech," Paper presented at the 146th Convention of the Audio Engineering Society, Dublin, Ireland (2019).