

High-resolution spherical directivity of live speech from a multiple-capture transfer function method

Timothy W. Leishman, Samuel D. Bellows, Claire M. Pincock, and Jennifer K. Whiting

Citation: *The Journal of the Acoustical Society of America* **149**, 1507 (2021); doi: 10.1121/10.0003363

View online: <https://doi.org/10.1121/10.0003363>

View Table of Contents: <https://asa.scitation.org/toc/jas/149/3>

Published by the [Acoustical Society of America](#)



Across Acoustics

The official podcast highlighting authors' research from our publications

High-resolution spherical directivity of live speech from a multiple-capture transfer function method

Timothy W. Leishman,^{a)} Samuel D. Bellows, Claire M. Pincock,^{b)} and Jennifer K. Whiting

Acoustics Research Group, Department of Physics and Astronomy, Brigham Young University, N284 Eyring Science Center, Provo, Utah 84602, USA

ABSTRACT:

Although human speech radiation has been a subject of considerable interest for decades, researchers have not previously measured its directivity over a complete sphere with high spatial and spectral resolution using live phonetically balanced passages. The research reported in this paper addresses this deficiency by employing a multiple-capture transfer function technique and spherical harmonic expansions. The work involved eight subjects and 2522 unique sampling positions over a 1.22 or 1.83 m sphere with 5° polar and azimuthal-angle increments. The paper explains the methods and directs readers to archived results for further exploration, modeling, and speech simulation in acoustical environments. Comparisons of the results to those of a KEMAR head-and-torso simulator, lower-resolution single-capture measurements, other authors' work, and basic symmetry expectations all substantiate their validity. The completeness and high resolution of the measurements offer insights into spherical speech directivity patterns that will aid researchers in the speech sciences, architectural acoustics, audio, and communications.

© 2021 Acoustical Society of America. <https://doi.org/10.1121/10.0003363>

(Received 12 August 2020; revised 15 December 2020; accepted 29 December 2020; published online 5 March 2021)

[Editor: Zhaoyan Zhang]

Pages: 1507–1523

I. INTRODUCTION

Speech is a ubiquitous faculty of human expression that affects personal and group communication, human-machine interaction, sound in acoustical environments, and signals in audio and telecommunication systems. Despite its importance in these and other areas and its prominence in many scientific studies and technologies, the fine details of speech radiation and diffraction are not well documented or understood. To increase knowledge in this area and advance future research, the authors have conducted high-resolution, spherically sampled measurements that thoroughly characterize speech directivity patterns and provide functional results for general usage.

Past challenges in measuring live speech with sufficient angular resolution have resulted in a lack of accessible, detailed data. Professionals have consequently relied on low-resolution, plane-polar, and even roughly estimated directivities for their work. Meanwhile, high-resolution spherical directivities of loudspeakers (e.g., with 5° uninterpolated polar and azimuthal-angle resolutions) have become virtually standardized tools for simulating and improving sound systems.^{1–3} It stands to reason that similar directivities should become available for human speech to better characterize its properties and enhance modeling, simulation, and optimization for many applications.

In the past, various researchers have explored the directivity of live speech,^{4–12} a combination of speech and

singing,^{13,14} or singing alone.^{15,16} Others have investigated radiation patterns from head-and-torso simulators (HATSs) and compared them to those of human speech^{17–24} or singing.^{25–29} Some have studied speech^{30–32} or HATS^{33–36} directivities in relationship to theoretical head models. Others have investigated speech, singing, or HATS directivities connected with theoretical or physical models of baffled vocal tracts for distinct vowels, including the effects of higher-order cross-modes.^{37–39} Still others have explored measured or perceived spectral variations in audio signals from microphone placements at different radii and angles.^{5,40–42} These efforts have afforded valuable insights into human voice directivity. However, the results are insufficient for many modern applications, and each has suffered from distinct limitations.

Few directivity measurements have been taken, plotted, or tabulated over a complete sphere (or an entire hemisphere with assumed symmetry), or else sampling was insufficient to characterize the frequency-dependent spherical directivity thoroughly over useful bandwidths. As limited spherical data subsets, conventional plane-polar and line plots for one, two, or three distinct planes only minimally characterize spherical directivity functions. The completeness and resolution recommended by current loudspeaker standards are much more advantageous for understanding and modeling directivities than these traditional representations.

In some cases, measurements have suffered from spatial aliasing or processing errors. Most did not utilize the corrective, noise-reducing capabilities of narrowband signal processing in the complex frequency domain. Many produced full, half, or 1/3-octave-band results, but the graphical or

^{a)}Electronic mail: tim_leishman@byu.edu

^{b)}Current address: MD Acoustics, LLC, 170 South William Dillard Drive, Suite A103, Gilbert, AZ 85233.

tabular presentations were not particularly useful for detailed analysis, modeling, and design needs. The results typically failed to leverage physical symmetry about the median talker plane; researchers did not acquire complete spherical data to validate or symmetrize measurements or enhance averaging. Several studies failed to utilize live, phonetically balanced speech, relying instead upon simulated speech, limited phonemes, or phonemic glissandi.

This work's primary aims were to overcome such limitations, acquire detailed, high-resolution spherical directivities of averaged live speech, and disseminate them in electronic formats that facilitate straightforward applications. Multichannel, multiple-capture (rotate-and-repeat) digital recordings of several seated male and female subjects in an anechoic chamber formed the study's basis. In many ways, past incognizance of related methodologies, their potential for high-resolution measurements, and associated mitigation of repetition variations in narrow or broad spectral bands has hampered spherical directivity work. Some researchers have sensibly divided $1/n$ -octave-band levels from sampling microphone signals by those of reference microphone signals. However, this paper shows that additional processing enhancements provide desirable benefits for narrowband and broader-band directivities.

The subjects spoke efficient, phonetically balanced passages while the measurement system captured their radiated sounds, which were inherently affected by diffraction and absorption. With 5° equiangular sampling in both the polar and azimuthal angles, the procedure employed 2522 unique measurement positions over a complete sphere. Subsequent signal processing yielded frequency response functions (FRFs or transfer functions) between a near-field reference microphone output in the rotating reference frame and those of more distant semicircular array microphones in a fixed reference frame. Ratios of averaged cross-spectra to averaged input auto-spectra produced the FRFs, led to effective coherent output spectra, improved noise immunity, and yielded normalized directivity functions for both narrow and energetically summed broader bands.

Directivity balloons and polar plots in the median, frontal, and transverse planes provided convenient visualizations of the results. Complementary coherence balloons also demonstrated relevant qualities. Subsequent spherical harmonic expansions of the angularly sampled narrowband FRFs and associated broader-band summations of coherent output spectra provided continuous functions for directivity reconstructions, angular resampling, smoothing, and other purposes.

Section II explains the measurement methods and demonstrates how they enabled directivity acquisition for subsequent sharing. Section III presents illustrative results. Section IV provides further analysis and discussion, including comparisons of live speech directivities to those of a KEMAR HATS, lower-resolution results acquired using a quasi-spherical 32-point microphone array, and the results of other authors.

II. METHODS

A. Measurement system and procedure

Figure 1 depicts the speech directivity measurement system used in the investigation, located in an anechoic chamber with an 80 Hz cutoff frequency (below the typical talkers' fundamental frequencies). The chair could adjust vertically or horizontally within the rotating reference frame to locate the talker mouth opening at the microphone array and corresponding measurement sphere's circular center (see Fig. 2). The rotation apparatus comprised an axle-and-gear mechanism driven by a stepper motor, allowing azimuthal rotation of the seated talker via computer control. Once the mouth aligned initially toward the polar angle $\theta = 90^\circ$ and the azimuthal angle $\phi = 0^\circ$, a head restraint minimized later movement within the rotating reference frame. It consisted of a thin plastic web attached to a narrow metal framework, which connected firmly to the chair.

Each subject wore a small 5.4 mm diameter head-worn pre-polarized condenser microphone, positioned consistently at the corner of his or her mouth to provide a reference signal. Two additional precision 1/2 in. (12.7 mm) pre-polarized condenser microphones provided alternative reference signals from more distant fixed positions within the rotating reference frame.

The semicircular array included 37 precision 1/2 in. (12.7 mm) pre-polarized condenser microphones, each with approximately 50 mV/Pa sensitivity and relative calibration over frequency, to produce spatially sampled signals of the radiated field. They fell at $\Delta\theta = 5^\circ$ polar-angle increments with a fixed radial distance of either $a = 1.83$ m or $a = 1.22$ m from the spherical center using adjustable support rods. Custom phantom power converters supplied the array microphone preamplifiers and two reference microphone preamplifiers utilizing integrated electronics piezo-electric (IEPE).

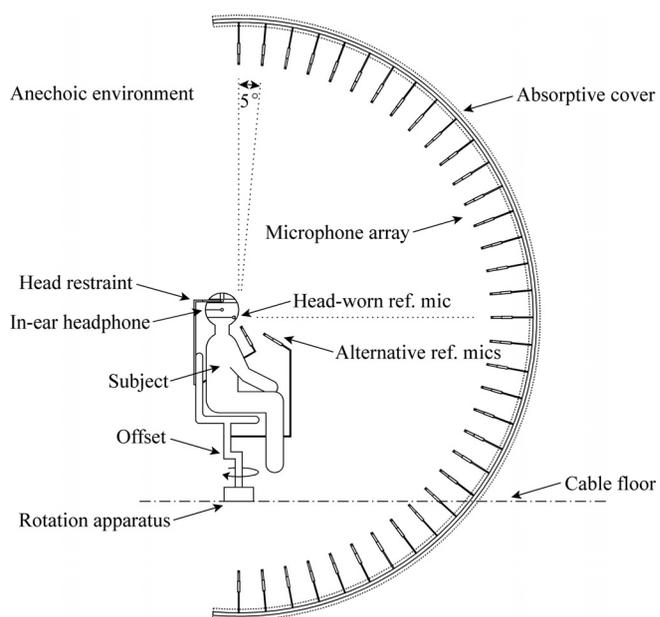


FIG. 1. Diagram of the speech directivity measurement system.

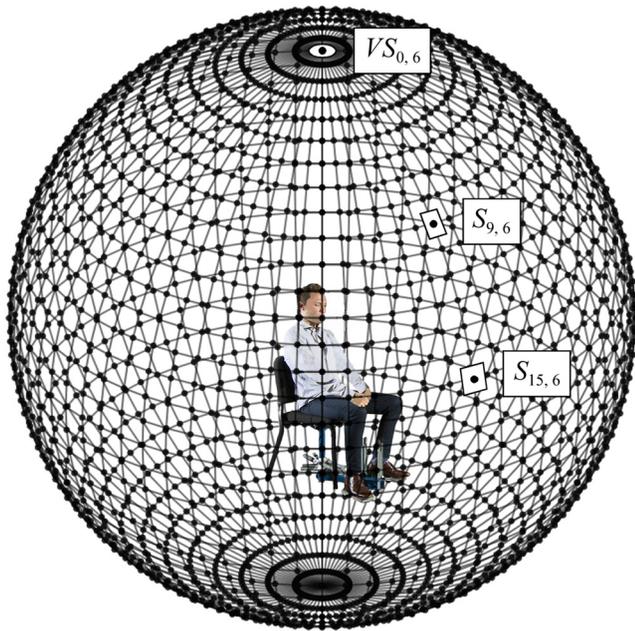


FIG. 2. (Color online) Sampling positions of the speech directivity measurement system. Effective sampling areas differed, as suggested by those illustrated about the points labeled $VS_{0,6}$, $S_{9,6}$, and $S_{15,6}$, where $V = 72$. For convenience in summations, $S_{0,v}$ and $S_{36,v}$ segmented the sampling areas of the two poles into V equal sectors, where $v = 0, 1, 2, \dots, V - 1$. The total polar sampling areas were then $VS_{0,v}$ and $VS_{36,v}$.

The microphone signals fed five eight-channel Focusrite RedNet 4 digital audio interfaces for recording at 24 bits and 48 kHz sampling frequency. At higher frequencies, e.g., above 4 kHz, signal-to-noise ratios were generally lower behind each talker than in front. This effect was due partly to diffraction about the subject's head, body, and chair and partly to the 40-channel gain settings made at $\phi = 0^\circ$ that remained fixed for all other azimuthal angles. The recordings thus proceeded in 40-channel files for each azimuthal angle through Reaper Digital Audio Workstation software.

The main subjects included three men and three women, all American-English speakers between ages 18 and 35. Each memorized the short sentences proposed by Lai *et al.*,⁴³ which contained the most commonly used phonemes. Before a measurement sequence began, the subject spoke the sentences for an initial recording via the head-worn reference microphone. Subsequent playback through a small in-ear headphone prompted and paced the talker for additional practice. While pacing was the primary aim of the technique, pitch matching and repetition consistency may also have improved from its usage. Once the subject repeated the passages confidently and consistently, the 40-channel directivity measurement sequence began.

The talker repeated the same six sentences in step with continued in-ear prompts following each $\Delta\phi = 5^\circ$ azimuthal rotation relative to the fixed microphone array. In case of a word or phoneme repetition error, the recording technician stopped the multichannel recording and repeated it at the same angle. The procedure did not expressly monitor or

compensate for vocal fatigue, e.g., as perceived by the talker or determined by physiological indicators, including reduced ranges and control of fundamental frequency and level.^{44,45} However, slight variations in these particular indicators did not appear to affect the directivities appreciably—especially with the study's signal-processing methods, subject averaging, and broadband representations. Consequently, they did not constitute repetition errors.

As depicted in Fig. 2, the multiple-capture method produced 2522 unique sampling positions that thoroughly assessed the speech radiation. The complete measurement procedure, including setup, practice, 72 passage repetitions, and takedown, took approximately 2 h per subject. The full 360° rotation in the azimuthal angle offered several benefits, including examinations of typical symmetry assumptions about the median plane and associated measurement validations. It also allowed symmetrizing of data or a quasi-doubling of subject averaging.

A second, single-capture system facilitated methodological comparisons for two additional subjects (one male and one female). It comprised a constellation of 32 microphones with positions at the apexes of a Catalan pentakis dodecahedron of nominal radius $a = 1.83$ m and angles confirmed by an altazimuth-mounted laser pointer at the spherical center.⁴⁶ Recordings from this system were more efficient than those of the multiple-capture system because they required no passage repetitions and only about 30 min per subject. However, inherently limited angular resolution resulted from the 32 unique measurement positions. An $a = 1.83$ m radius version of the higher-resolution multiple-capture system also allowed comparative measurements of the two subjects.

The $a = 1.22$ m radius multiple-capture system assessed the GRAS KEMAR 45BC HATS directivity for comparison with those of the six primary talkers. The HATS sat on a narrow stand in place of the chair (see Fig. 3) as it rotated azimuthally and its mouth opening remained fixed at the center of the measurement sphere. A 20 Hz to 21.5 kHz linear sine sweep drove the HATS as the reference signal for FRF measurements based on spectral averages over five iterations.

B. Narrowband signal processing

The signal processing for the systems included computations of the FRFs between a reference signal (e.g., from the head-worn microphone) and each of the array signals. The multiple-capture system involved $U = 37$ polar angles and $V = 72$ azimuthal angles on the measurement sphere. (The $a = 1.22$ m array involved only $U = 36$ polar angles because the rotation apparatus obstructed the nadir microphone.) The FRFs over the entire sphere were then $H_{u,v}(f) = H(\theta_u, \phi_v, f)$, where $u = 0, 1, 2, \dots, U - 1$, $v = 0, 1, 2, \dots, V - 1$, $\theta_u = u\Delta\theta$, $\phi_v = v\Delta\phi$, and $\Delta\theta = \Delta\phi = 5^\circ$. Each azimuthal angle included repetitions of the zenith and incorporated nadir measurements.

Transfer function calculations for each measurement position employed Welch's method,^{47,48} discrete Fourier transforms performed on the six-sentence passage involved 48 000-sample block sizes (1 s record length, 1 Hz



FIG. 3. (Color online) The GRAS KEMAR 45BC HATS positioned on a stand in the directivity measurement system. A fixed microphone attached to the stand rotated with the HATS to provide an alternative reference signal.

narrowband resolution), a Hann window, and 75% overlap. This approach typically produced 64 or more block averages over the six sentences. The autospectrum $G_{aa_v}(f)$ of the reference microphone signal $a_v(t)$ followed similarly for each azimuthal-angle increment (continuous frequency f and time t are used here only for convenience). The method likewise led to an average cross-spectrum $G_{a\tilde{b}_{u,v}}(f)$ between the reference microphone signal and each array microphone signal $\tilde{b}_{u,v}(t)$, where the tilde suggests the ideal signal $b_{u,v}(t)$ contained uncorrelated noise $n_{u,v}(t)$ [i.e., $\tilde{b}_{u,v}(t) = b_{u,v}(t) + n_{u,v}(t)$]. The FRF then followed as $H_{u,v}(f) = G_{a\tilde{b}_{u,v}}(f)/G_{aa_v}(f)$, a least squares estimator reducing bias associated with the noise.⁴⁸ These calculations assumed the system between the reference and array microphones was linear and approximately time-invariant following each azimuthal-angle increment (see Secs. II C and III A for further discussion).

The previously mentioned spectra and the autospectrum $G_{\tilde{b}_{u,v}\tilde{b}_{u,v}}(f)$ of each array signal allowed calculation of the microphone position's coherence function as $\gamma_{a\tilde{b}_{u,v}}^2(f) = |G_{a\tilde{b}_{u,v}}(f)|^2 / G_{aa_v}(f) G_{\tilde{b}_{u,v}\tilde{b}_{u,v}}(f)$. This calculation provided means of testing signal associations in the assumed linear input-output model⁴⁸ and quantifying and visualizing confidence in the FRFs over the full measurement sphere. The frequency-dependent values related to causality, signal-to-noise ratio, and other signal and system characteristics. While speech is neither an entirely deterministic nor random broadband process, the coherence function distinguished valid spectral regions for the FRFs and their ensuing directivities. Because acceptable FRFs required sufficient radiated sound levels at both the reference and array microphone positions, calculated signal-to-noise ratios

$SNR_{a\tilde{b}_{u,v}}(f) = 10 \log \left\{ \gamma_{a\tilde{b}_{u,v}}^2(f) / [1 - \gamma_{a\tilde{b}_{u,v}}^2(f)] \right\}$ enabled further monitoring of uncorrelated noise in the associated signals.⁴⁹

C. Broader bands

For some applications, directivities in $1/n$ -octave or other broader bands are more practical than those in narrow bands. Accordingly, one might consider summing $|H_{u,v}(f)|^2$ directly into the broader bands with equal weighting per frequency. However, radiated speech amplitude is not spectrally uniform; its frequency dependence differs from any transfer function's dependence, suggesting that the latter should receive a proportional weighting. One approach to the problem might involve multiplying each $|H_{u,v}(f)|^2$ by the reference input autospectrum $G_{aa_v}(f)$, which coincidentally yields the coherent output spectrum $G_{bb_{u,v}}(f) = |H_{u,v}(f)|^2 G_{aa_v}(f) = \gamma_{a\tilde{b}_{u,v}}^2(f) G_{\tilde{b}_{u,v}\tilde{b}_{u,v}}(f)$. This product represents the array microphone signal energy that is fully coherent with the reference microphone signal.

In theory, the coherent output spectrum eliminates uncorrelated noise measured at the array and optimizes signal identification due to the measurement system's input. Of course, the reference microphone must adequately detect the speech signal generated from within the time-varying vocal tract and produce a linearly related signal with negligible noise contamination. The reference signal then yields the same theoretical coherent output spectrum as the true source signal in the vocal tract, despite modifications by linear transfer functions to the external reference microphone and through its transduction mechanism.^{48,50} In practice, time-delay bias errors, estimation errors, and deviations from linear time-invariant (LTI) assumptions reduce the full benefits of coherent output spectra, which highlights the importance of reducing those effects. Corrective strategies include estimating the magnitude-squared FRF as $|H_{u,v}(f)|^2 = |G_{a\tilde{b}_{u,v}}(f)/G_{aa_v}(f)|^2$ and averaging many data blocks to reduce spectral-estimate variances and narrowband measurement noise. The strategies may also involve large block sizes and consequently narrower frequency bins that improve the broader bands' overall level estimates.⁵⁰ Talker restraints and prompts reduce the measurement system's time-varying features.

Because the reference microphone position influences the spectral content of $G_{aa_v}(f)$, and the latter varies with passage repetition at each azimuthal angle, a *global* frequency-dependent weighting of $|H_{u,v}(f)|^2$ would improve broader-band directivities. The sound power spectrum could form such a weighting. However, for multiple-capture live-speech measurements, it must incorporate azimuthal variations in $\hat{a}_v(f)$, the Fourier transform of $a_v(t)$, and $G_{aa_v}(f)$ rather than assuming they are consistent, as for a single-capture measurement or a perfectly repeating source. Following algebraic manipulation, an equating of the sound power formulations for multiple and single-capture measurements yields an effective input autospectrum

$$G_{aa,eff}(f) = \frac{\sum_{u=0}^{U-1} \sum_{v=0}^{V-1} S_{u,v} |\hat{a}_v(f)|^2 |H_{u,v}(f)|^2}{\sum_{u=0}^{U-1} \sum_{v=0}^{V-1} S_{u,v} |H_{u,v}(f)|^2}, \quad (1)$$

where $S_{u,v}$ is the effective sampling area of each microphone position (see Secs. II A and II F). The effective coherent output spectrum for each sampling position then follows as

$$G_{bb,eff_{u,v}}(f) = |H_{u,v}(f)|^2 G_{aa,eff}(f). \quad (2)$$

In this work, calculations of $1/n$ -octave-band directivities followed by adding each bin of $G_{bb,eff_{u,v}}(f)$ to a broader band when it fell entirely within the band limits. If it spanned a limit, its energy divided proportionately between the adjacent bands.

D. Spherical harmonic expansions

An expansion of the measured complex FRFs using spherical harmonics led to the continuous angular function

$$H(\theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n a_{nm}(f) \tilde{Y}_n^m(\theta, \phi), \quad (3)$$

where $\tilde{Y}_n^m(\theta, \phi)$ are normalized spherical harmonics of degree n and order m (see the terminological discussions in Refs. 51–53). The expansion coefficients $a_{nm}(f)$, to truncated expansion degree N , followed from either a quadrature method or a least squares approximation based on the discrete sampling positions.^{54,55} Similar expansions applied to the real-valued $G_{bb,eff_{u,v}}(f)$ or broader-band results. Because higher-degree spherical harmonics typically correspond to more considerable spatial variations, an appropriate choice of N can smooth data and reduce unwanted spatial noise.

E. Normalized directivity functions

Normalizing the narrowband FRFs by the maximum found at any sampling position on the sphere for a given frequency resulted in the sampled complex directivity function

$$D_{u,v}(f) = \frac{H_{u,v}(f)}{H_{(u,v)_{\max|H|}}(f)}, \quad (4)$$

where $(u, v)_{\max|H|}$ represents the index pair with the maximum FRF magnitude given by $\max_{u,v} |H_{u,v}(f)|$ [i.e., the maximum of all $|H_{u,v}(f)|$ for the given f]. The associated decibel beam pattern for directivity balloons and polar plots then followed as⁵⁶

$$B_{u,v}(f) = 20 \log |D_{u,v}(f)|. \quad (5)$$

From a degree N truncation of Eq. (3), a similar normalization of $H(\theta, \phi, f)$ on the sphere yielded an angularly continuous complex directivity function

$$D(\theta, \phi, f) = \frac{H(\theta, \phi, f)}{H[(\theta, \phi)_{\max|H|}, f]}, \quad (6)$$

where $(\theta, \phi)_{\max|H|}$ is the angle with maximum FRF magnitude given by $\max_{\theta, \phi} |H(\theta, \phi, f)|$ for a given f . The associated decibel beam pattern was then

$$B(\theta, \phi, f) = 20 \log |D(\theta, \phi, f)|. \quad (7)$$

Similar formulas applied to the real-valued $G_{bb,eff_{u,v}}(f)$ or broader-band results.

Some authors have attached phases to $1/n$ -octave-band directivities for array radiation predictions.^{57,58} While Eqs. (4) and (6) contain phase information, the utility of phase for $1/n$ -octave-band speech directivities is indefinite.

F. Directivity deviation

An energetic, area-weighted, root-mean-square deviation (AWRMSD) provided means of globally quantifying differences between any directivity $D_{u,v}(f)$ and a reference directivity $D_{ref,u,v}(f)$. Suitable comparisons required alternative normalization such that the area-weighted mean magnitude of each function became unity for each frequency bin. This step was necessary because the normalization in Eq. (4) depends upon specific angles of maximum radiation that may vary over frequency and between talkers.⁵⁹

The area weights were the effective sampling areas for each microphone position (see Fig. 2), defined in Eqs. (4)–(6) of Ref. 56 as

$$S_{u,v} = \begin{cases} 2a^2 \Delta\phi \sin^2\left(\frac{\Delta\theta}{4}\right) = \frac{4\pi a^2}{V} \sin^2\left(\frac{\Delta\theta}{4}\right); & u = 0, 36 \\ 2a^2 \Delta\phi \sin(\theta_u) \sin\left(\frac{\Delta\theta}{2}\right); & 1 \leq u \leq 35, \end{cases} \quad (8)$$

which sum to the total area S of the measurement sphere. A quadrature rule based on these weights allows approximate numerical integration of a function $f(\theta, \phi)$ over the sphere as

$$\int_0^{2\pi} \int_0^\pi f(\theta, \phi) a^2 \sin \theta d\theta d\phi \approx \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} S_{u,v} f_{u,v}. \quad (9)$$

Consequently, at each frequency, the mean-normalized directivity function takes the form

$$\tilde{D}_{u,v}(f) = \frac{H_{u,v}(f) S}{\sum_{u=0}^{U-1} \sum_{v=0}^{V-1} S_{u,v} |H_{u,v}(f)|}. \quad (10)$$

With this result, the AWRMSD becomes

$$\sigma_{\text{AWD}}(f) = \sqrt{\frac{1}{S} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} S_{u,v} \left[|\tilde{D}_{u,v}(f)| - |\tilde{D}_{\text{ref},u,v}(f)| \right]^2}, \quad (11)$$

with a decibel form

$$L_{\text{AWRMSD}}(f) = 20 \log_{10}[1 + \sigma_{\text{AWD}}(f)]. \quad (12)$$

The latter maps $\sigma_{\text{AWD}} = 0$ to 0 dB and $\sigma_{\text{AWD}} = 1$ to 6 dB. A single-number value also follows by energetically averaging this level over frequency. Sections III and IV utilize these metrics to compare directivities.

III. RESULTS

A. Narrowband directivities

Figure 4 shows the normalized 630 Hz narrowband (1 Hz bandwidth) directivity for a female talker using two visualization techniques. In each case, the mouth axis falls at $(\theta, \phi) = (90^\circ, 0^\circ)$. Since the sampling positions fell at a fixed radius $a = 1.83$ m over the measurement sphere, color alone might feasibly represent the normalized levels at the various angles, as shown in Figs. 4(a) and 4(b). The color scales from -40 to 0 dB over the sphere; values less than

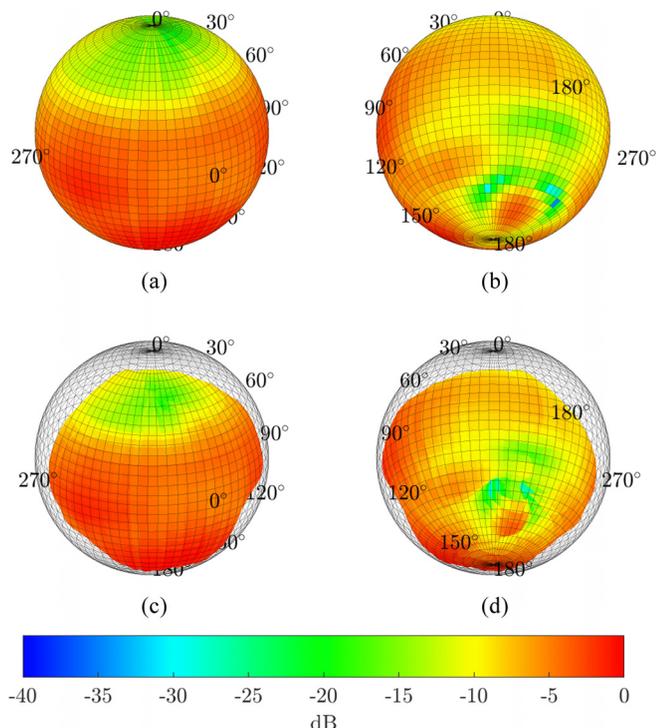


FIG. 4. (Color online) Two visualization schemes for the narrowband (1 Hz resolution) directivity of a female talker at 630 Hz, where the mouth axis falls at $(\theta, \phi) = (90^\circ, 0^\circ)$. (a) Viewed from above the equator and to the talker's right, color over a sphere on a dB scale represents a level at each sampling position relative to the maximum. (b) The same as (a) but viewed from below and to the talker's left. (c) Viewed from above and to the talker's right, a varying parametric surface (balloon) radius also represents the relative level from the circumscribed spherical cage (0 dB) to the origin (-40 dB). (d) The same as (c) but viewed from below and to the talker's left.

-40 dB maintain the same color as -40 dB. A visualization enhancement follows by representing the levels at the various angles with parametric surface radii ranging from -40 to 0 dB to form a distinct shape or "balloon," as shown in Figs. 4(c) and 4(d). Both plots include faceting to produce gray mesh lines and consistent color across each planar face. The circumscribed spherical cages signify the maximum normalized directivity value of 0 dB. The outline of a planar slice through a balloon and the origin constitutes a directivity polar plot.

Figure 5(a) shows a similar narrowband balloon smoothed by interpolating color across the mesh line segments and planar faces, a technique employed in subsequent plots. A view from behind the balloon follows later in Fig. 8(a). Figures 5(b)–5(d) are polar plots for the transverse, frontal, and median planes, respectively, which also overlay the balloon of Fig. 5(a) to help orient the viewer and augment visualization. Polar plots in the transverse and frontal planes exhibit quasi-symmetry because of the seated talker's anatomical symmetries, which affect radiation, diffraction, and absorption. While the level is generally highest in front of the talker, the maximum occurs at about 40° downward from the horizontal in the median plane. This result could differ for a standing talker.

Figure 6 shows balloon and polar plots of the calculated coherence at the same frequency. Here, the radii and color scale range only from 0.9 to 1.0 to emphasize the generally high coherence values over the sphere while highlighting reduced-value regions. These plots are useful for identifying problems such as incoherent noise, distortion, insufficient gain, poorly functioning microphones, or bad connections.

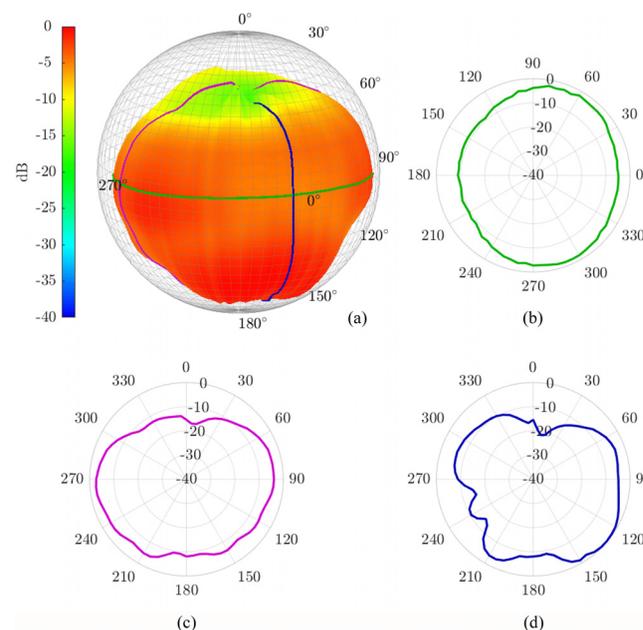


FIG. 5. (Color online) Balloon and polar plots for the narrowband (1 Hz resolution) 630 Hz directivity of a female talker. (a) Directivity balloon using an interpolating color scheme. (b) Transverse polar plot. (c) Frontal polar plot. (d) Median polar plot. The polar plots also overlay the balloon plot.

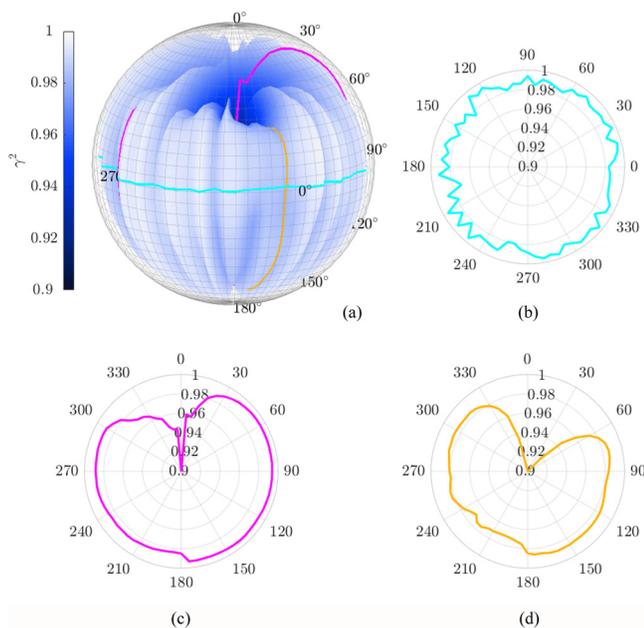


FIG. 6. (Color online) Coherence balloon and polar plots associated with the narrowband 630 Hz directivity measurement of the female talker in Figs. 4 and 5. The plots are on a scale from 0.9 to 1. (a) Balloon. (b) Transverse plane. (c) Frontal plane. (d) Median plane. The polar plots also overlay the balloon plot.

Coherence tends to drop whenever speech levels detected by the array or reference microphones drop, resulting in a decreased signal-to-noise ratio.

It is noteworthy that sound generation by the KEMAR HATS mouth simulator was spectrally and energetically replete due to its swept-sinusoidal excitation and transduction and virtually time-invariant because of its fixed properties. Sound produced by the human talkers was less spectrally replete and more time-variant due to the characteristic fluctuations of vocal-tract configurations.^{37–39,60} However, for both the HATS and talkers, the gross diffraction of external sound about the heads and bodies and propagation into the free field were roughly time-invariant following each azimuthal rotation. Time-dependent talker vocal-tract geometries may have varied somewhat with rotation angle and repetition, but regularly prompted repetitions of set phonetically balanced passages aimed to improve consistency and minimize the impact on directivity. Nevertheless, as shown later in Sec. IV A, frequency-dependent spatially averaged coherence values for the HATS were consistently higher over its usable bandwidth than those of the human talkers.

Section II C suggested that a talker would not produce a consistent level at each azimuthal angle. The 630 Hz, auto-spectral polar plot from the reference-microphone and array balloon plot in Fig. 7 affirm this expectation. The variations are evident to a lesser extent from the longitudinal banding of the coherence balloon of Fig. 6 and its transverse polar plot; reductions in radiated field levels in Fig. 7 correspond roughly to regions of reduced coherence. Furthermore, for the narrow 1 Hz bandwidth, one sees increased asymmetry near the directivity balloon’s zenith [Fig. 5(c)] and

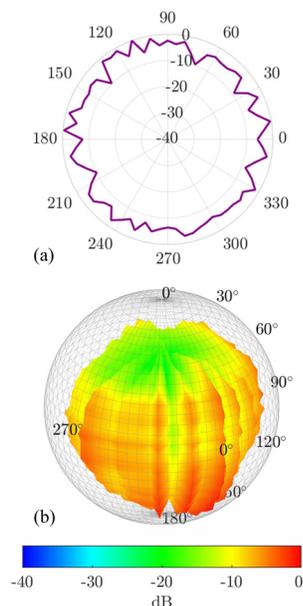


FIG. 7. (Color online) Autospectral plots associated with the narrowband 630 Hz directivity measurement of the female talker in Figs. 4–6. (a) Reference microphone polar plot. (b) Microphone array balloon plot. Both plots show expected amplitude variations for each repetition at 5° azimuthal-angle increments.

connected distortions in the median plane [Fig. 5(d)]. Both effects relate to an area of reduced coherence and are further evident from a comparison of the FRF-based directivity balloon and the associated coherence balloon viewed from behind the talker in Fig. 8. Despite these behaviors, the images reveal nearly symmetric diffraction patterns about the talker’s body and chair.

Figure 9 depicts two views of a lower-resolution directivity balloon produced by the same female talker at the same frequency and nominal radius $a = 1.83$ m, but acquired using the single-capture array of 32 microphones. As in Fig. 5(a), the flat faces depict planar color interpolations between contiguous sets of sampling points. Figure 9(a), showing the balloon from in front of the talker, and Fig. 9(b), from behind, have features roughly similar to those in Fig. 5, including reduced levels above and behind the talker. However, Fig. 9(b) also illustrates a significant problem that may arise from conflicting talker and array symmetries and coarse sampling resolution.

The Catalan pentakis dodecahedron array geometry, used elsewhere by other researchers,^{61,62} afforded useful comparative sampling with the highest quasi-uniform sampling channel count from the available 40 channels. However, for Fig. 9, the talker median plane did not align with one of the array’s 15 mirror-image planes of its icosahedral symmetry, while the frontal plane did. Consequently, the measurement arrangement produced visible directivity asymmetries, such as the -20 dB patch seen in Fig. 9(b) only on the balloon’s left side. Asymmetries also appear to a lesser extent in Fig. 9(a) but are not noticeably present in Fig. 5, measured with high-resolution sampling. Because sparse sampling schemes can significantly affect directivity measurement

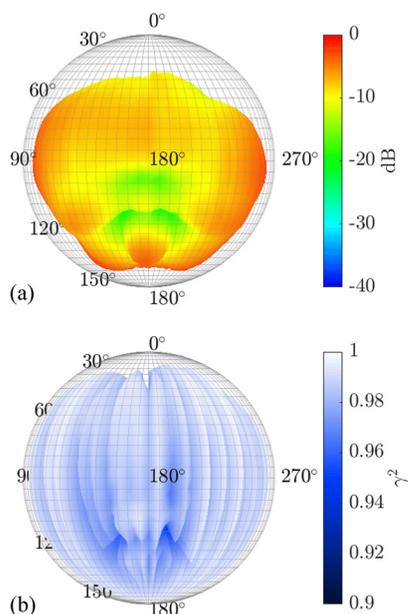


FIG. 8. (Color online) Balloon plots for the narrowband 630 Hz measurement of the female talker described by Figs. 4–7, viewed from a vantage point behind the talker. (a) Transfer-function-based directivity balloon. (b) Associated coherence balloon.

quality, their alignments should account, whenever possible, for critical source radiation, diffraction, and symmetry features. In this regard, higher-resolution measurements are typically more accommodating than lower-resolution measurements.

B. One-third-octave-band directivities

Figure 10 presents energetically weighted and summed directivity results over the 630 Hz 1/3-octave band (see Sec. II C) and energetically averaged across three female talkers measured at radius $a = 1.22$ m. Many of the 630 Hz narrowband directivity features in Fig. 5 are still apparent but are smoother due to the broader bandwidth and subject averaging. The smoothing has notably impacted artifacts caused by reduced coherence near the zenith and increased balloon symmetry. The frequency-averaged $L_{AWRMSD}(f)$ of the individual directivity (see Sec. II F) compared to that of the female talkers' reference average directivity was 1.6 dB. Figure 11 shows similar results for the 1.6 kHz 1/3-octave band averaged across the three male talkers measured at the same radial distance.

C. Directivities based on spherical harmonic expansions

The directivity results in Fig. 12 are similar to those in Fig. 11 but based on a degree $N = 10$ spherical harmonic expansion of the measured data. The frequency-averaged $L_{AWRMSD}(f)$ between the expansion and raw data for this band was 0.6 dB. Figure 13 presents similar spherical harmonic directivity results for the KEMAR HATS over the same band. With the expanded male average as the reference, it had a frequency-averaged $L_{AWRMSD}(f)$ of 2.1 dB. In

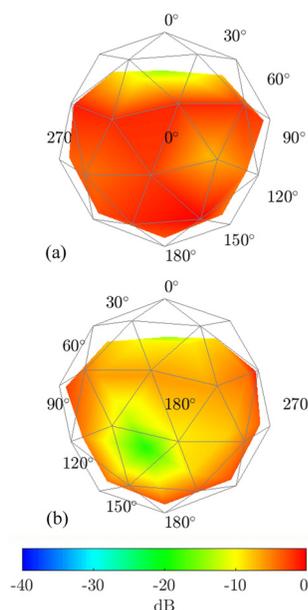


FIG. 9. (Color online) The narrowband 630 Hz directivity of the female talker as in Figs. 4–8, but measured with a 32-point single-capture array. (a) Front view. (b) Back view.

both cases, the highest levels occurred toward the front and in an upward inclination.

Figure 14 shows a sequence of balloons from a vantage point to a female talker's right, each expanded to degree $N = 15$ with increasing 1/3-octave-band center frequencies. Reference 63 provides several additional 1/3-octave-band directivity plots and animations for the human talkers and the KEMAR HATS, including balloon rotations and evolutions over frequency. This reference archives the average talker directivity results in various 1/3-octave bands, expanded with spherical harmonics to degree $N = 10$. The included electronic files contain tabulations for architectural acoustics simulations and other applications.

IV. ANALYSIS AND DISCUSSION

A. Usable bandwidth

As expected from physical arguments and as borne out by the measurements, the speech radiation was nearly omnidirectional at low frequencies, close to the fundamental frequencies of speech. This effect was less so for the females because their average fundamental frequencies were nearly double those of the males, while their pertinent anatomical dimensions for radiation and diffraction were closer to those of the males. Fundamental frequencies varied for the talkers over time, but as indicated for the male and female talkers in Fig. 15(a), the frequency-dependent time and area-weighted average coherences over the sphere provided insights into the functional spectral contents of their speech. Discrete coherence values were usually greater than or less than the average values at any given angle and frequency. The spatial average dropped consistently below 0.8 above about 5 kHz. At higher frequencies, coherence values directly behind a talker usually fell about 0.1 below the

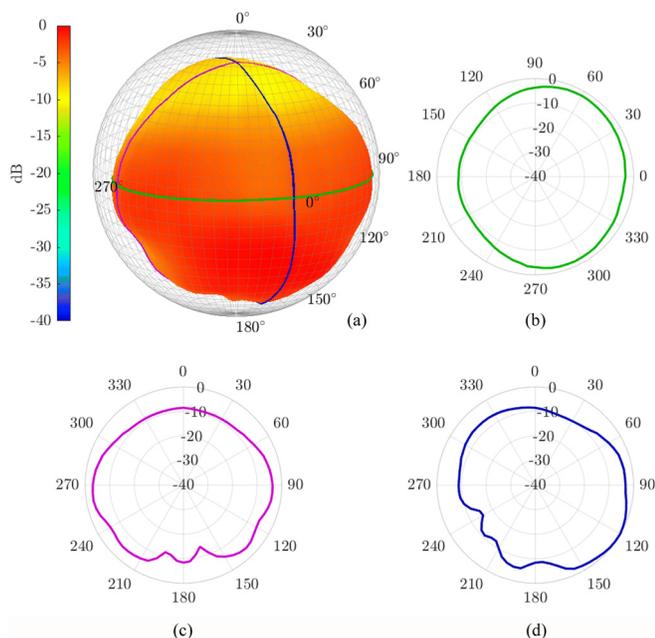


FIG. 10. (Color online) Female speech directivity averaged over three talkers for the 630 Hz 1/3-octave band. As expected from anatomical and seating symmetries, the patterns in the transverse and frontal planes are quasi-symmetric, with diffraction effects apparent behind and below the talkers. (a) Balloon. (b) Transverse plane. (c) Frontal plane. (d) Median plane.

average while those in front rose roughly 0.1 above it. As mentioned earlier, recording gains were set only once for the entire measurement sequence as a talker faced the semi-circular array. The results may have improved by optimizing signal-to-noise ratios and dynamic ranges for each

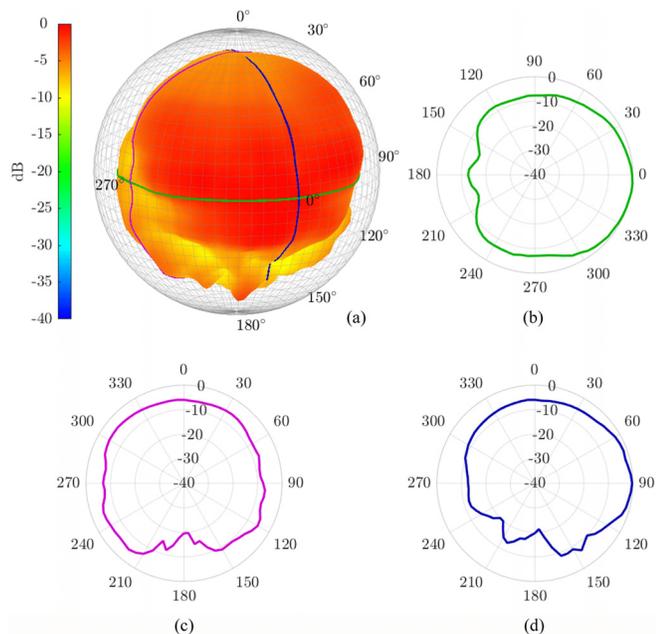


FIG. 11. (Color online) Male speech directivity averaged over three talkers for the 1.6 kHz 1/3-octave band. The highest levels occur in front of the seated talkers with an upward inclination, above the transverse plane. The patterns in the transverse and frontal planes are quasi-symmetric, with diffraction effects apparent behind and below the talkers. (a) Balloon. (b) Transverse plane. (c) Frontal plane. (d) Median plane.

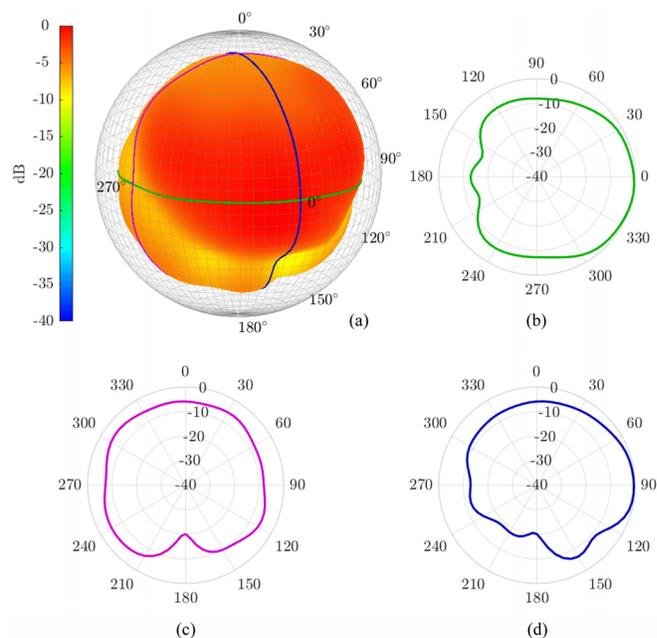


FIG. 12. (Color online) Average male speech directivity in the 1.6 kHz 1/3-octave-band, as in Fig. 11, but based on a degree $N = 10$ spherical harmonic expansion. The expansion smooths certain details while maintaining the general directional characteristics. (a) Balloon. (b) Transverse plane. (c) Frontal plane. (d) Median plane.

azimuthal angle. In general, a talker's spatially averaged coherence was less than that of the KEMAR HATS over its usable bandwidth.

As suggested in Secs. II C and III B, the use of coherent output spectra, broader bands, and multiple-subject

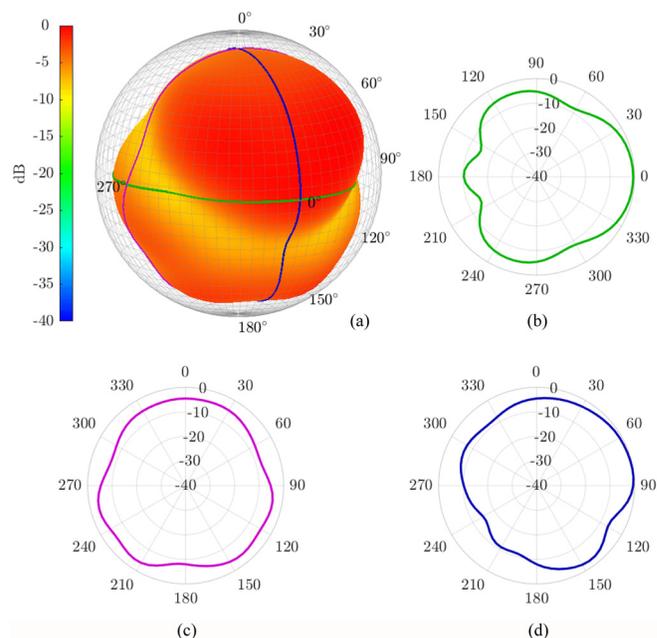


FIG. 13. (Color online) The KEMAR HATS directivity for the 1.6 kHz 1/3-octave band with a degree $N = 10$ spherical harmonic expansion. (a) Balloon. (b) Transverse plane. (c) Frontal plane. (d) Median plane.

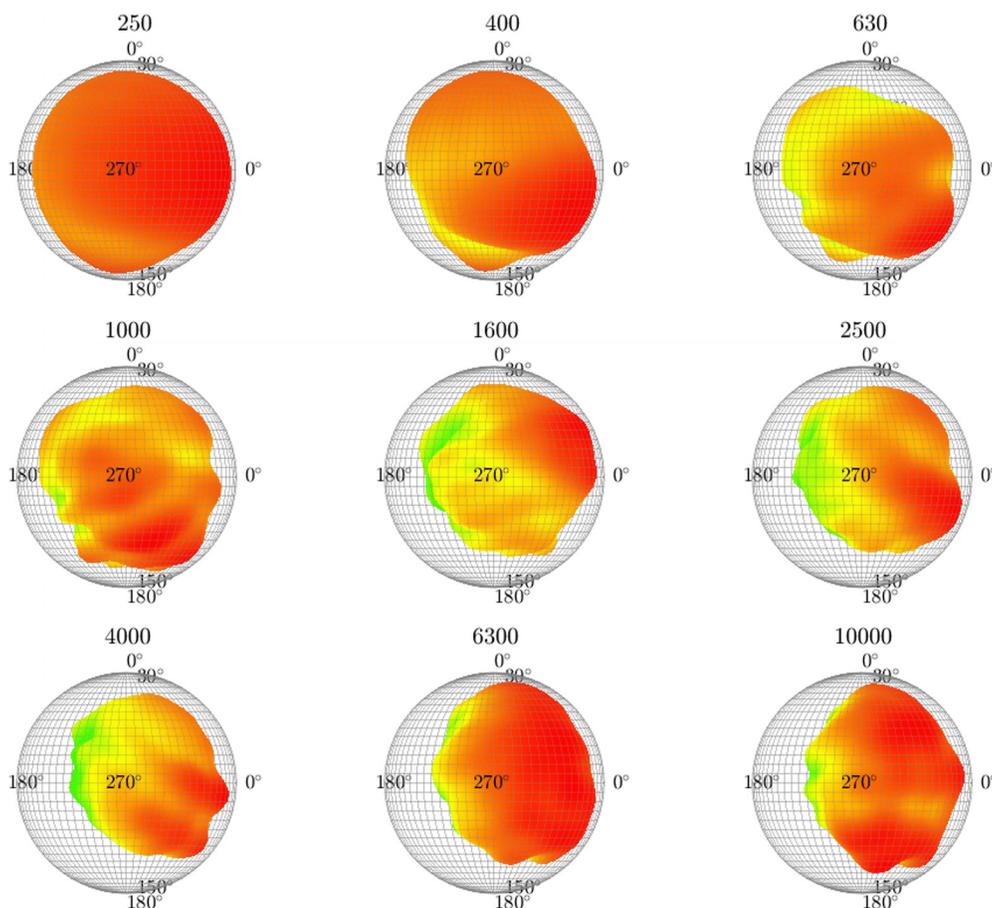


FIG. 14. (Color online) Directivity balloon side views for a single female talker in 1/3-octave bands, based on degree $N = 15$ spherical harmonic expansions of the measured data.

averaging served to isolate speech signals from zero-mean uncorrelated noise at any angle. The cross-spectrum played a crucial role in eliminating the noise in the coherent output spectra and transfer functions, but the latter two were only valid to the degree that the cross-spectral and autospectral estimates were valid. Coefficients of variation provided one means of predicting estimation errors. The estimates of the narrowband coherent output spectra and transfer functions (see Secs. II B and II C) depended upon both the coherence at any given frequency and the number of time-record averages.⁵⁰ As stated earlier, the latter was approximately 64 per talker and effectively 384 over six talkers.

Because of the gradually falling high-frequency coherence, it became necessary to estimate the usable directivity measurement bandwidth. At 10 kHz, the spatially averaged narrowband coherence ranged among the talkers from approximately 0.3 to 0.5, with an average of about 0.4. With 64 averages, the 0.4 value produced a narrowband coefficient of variation of approximately 11% for the average transfer function magnitude [see Eqs. (4)–(7)], suggesting that a 10 kHz narrowband limit was reasonable. The coefficient of variation for the average coherent output spectrum at this frequency was approximately 25%, but with 384 effective averages, it dropped to around 10%. Furthermore,

once the many narrowband spectral components combined into relatively broad high-frequency 1/3-octave bands, errors for overall band-level estimates dropped much further.⁵⁰ This suggested that the 10 kHz and even higher 1/3-octave bands were useful.

The zenith microphone captured signals at the same position for each azimuthal-angle increment and thus should have yielded similar transfer functions and effective coherent output spectra in each case. The 1/3-octave-band zenith directivity values for the two talkers and KEMAR had frequency-dependent standard deviations σ represented in Fig. 15(b) for the 72 increments. The talker standard deviations remained under 0.5 dB between their fundamental frequencies and about 6 kHz. They steadily rose from 0.3 dB at 5 kHz to approximately 1 dB at 10 kHz. Below the fundamental frequencies, they rose more dramatically. The standard deviation for the KEMAR HATS was relatively low over the plotted bandwidth except below its low-frequency response roll-off.

B. Directivity pattern characteristics

Section III A suggested that speech directivity is approximately symmetric in the transverse and frontal planes due to a subject’s anatomical and seating symmetries

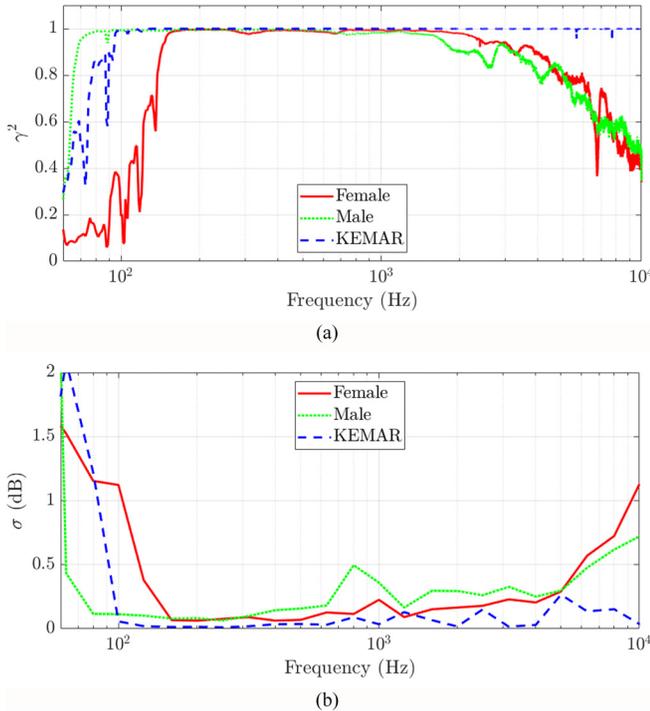


FIG. 15. (Color online) (a) Area-weighted spatially averaged coherence over the sampling sphere from 60 Hz to 10 kHz for a single male talker, a single female talker, and the KEMAR HATS. Frequencies with higher coherence roughly correspond to those with higher signal levels over the sphere. Because of lower fundamental frequencies, the male talker has higher coherence below 160 Hz than the female talker. (b) Standard deviations of the 1/3-octave-band zenith directivity values for the two talkers and KEMAR over the 72 azimuthal measurement angles.

about the median plane while the median plane’s results involve asymmetries. These characteristics help substantiate speech directivity measurements and are apparent in the measured results.

Both radiation from the mouth and nostrils and diffraction about a talker and chair cause directivity patterns to evolve remarkably over frequency. The variations in Fig. 14 and the images of Ref. 63 demonstrate several significant frequency-dependent behaviors about a seated talker that are characteristic of speech directivity. Diffraction effects behind and below the head, body, and chair appear in these and other balloon plots, with noteworthy lobe structures to the front and sides. Above the quasi-omnidirectional frequencies, a broad downward-oriented frontal lobe appears around 315 Hz. Different lobes emerge with increasing frequency until a broad upward-oriented frontal lobe dominates the radiation pattern at about 1.25 kHz. New lobes continue to form with increasing frequency, each with interesting angular orientations. Above 5 kHz, radiation is typically dominant in front of the talker, as one might expect from geometric-acoustic arguments.⁶⁴ However, the frequency-dependent pattern shows that a principal radiation axis directly in front of a seated talker’s mouth is not necessarily the axis of maximum radiation at all frequencies.⁵⁹ Table I provides the average on-axis directivity index and normalized sound power levels for all six talkers

TABLE I. On-axis directivity index and normalized sound power level averaged over three male and three female subjects, by 1/3-octave band.

Frequency (Hz)	Directivity index (dB)	Normalized power level (dB)
80	0.0	-28.4
100	0.0	-27.9
125	0.4	-31.0
160	1.3	-28.1
200	1.8	-2.8
250	2.8	0.0
315	3.2	-6.0
400	2.5	-5.7
500	1.3	-3.6
630	0.9	-4.7
800	0.5	-5.5
1000	0.7	-10.7
1250	3.7	-14.5
1600	5.7	-16.1
2000	4.4	-19.3
2500	2.7	-24.2
3150	4.0	-26.8
4000	5.2	-30.1
5000	4.0	-37.8
6300	3.7	-35.4
8000	6.0	-35.8
10000	6.3	-38.9

over several 1/3-octave bands. References 59 and 65 provide further discussion of these topics, based on the same data.

The KEMAR HATS directivity results, measured at $a = 1.22$ m, were similar in many regards to those of the human talkers but differed in particulars for several ostensible reasons. First, the HATS had no legs or chair during the measurements. Figure 13 shows that levels increased behind and below the manikin compared to those of the average male. Second, despite the manikin’s average or median anatomical design,⁶⁶ its features differed from those of this investigation’s specific talkers. Third, KEMAR produced distinctive radiation associated with its fixed mouth aperture and lack of nasal cavity or openings, which again varied from the talker anatomies and the characteristics of running human speech.

C. Comparisons

1. Theoretical predictions

Researchers have compared human speech radiation to the radiation produced by a point source or small radially oscillating cap set in a rigid spherical baffle, with dimensions similar to those of the human mouth and head, respectively.^{4,6,24,30,33-35} Figure 16(a) compares the average female 630 Hz 1/3-octave-band directivity of Fig. 10, in the transverse plane, to that of the KEMAR HATS and the axisymmetric directivity modeled from a 0.85 cm radius cap set in a 9.0 cm radius sphere. Figure 16(b) compares the average male 1.6 kHz 1/3-octave-band directivity from Fig. 11 to that of KEMAR and the directivity

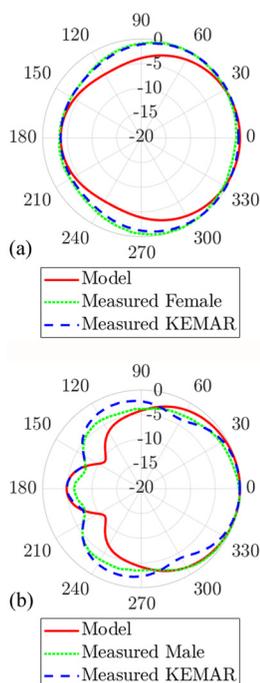


FIG. 16. (Color online) Transverse polar plots comparing narrowband directivities produced by theoretical models of spherically baffled vibrating caps with those produced in 1/3-octave bands by live speech and the KEMAR HATS. (a) Comparisons to average female directivity in the 630 Hz 1/3-octave band. (b) Comparisons to average male directivity in the 1.6 kHz 1/3-octave band.

modeled using a 0.90 cm radius cap set in a 9.5 cm radius sphere. Both Figs. 16(a) and 16(b) show reasonable agreement in front of and behind the head but significant deviations to the sides. This effect follows partly from the diffraction and scattering occurring about the shoulders, body, and chair—all features entirely neglected by the spherical model. The figure includes only transverse polar plots because the frontal and median-plane plots depart considerably from the modeled directivity; the seated subjects and KEMAR lacked significant axial symmetry in those planes.

Through theoretical models, computations, and physical models of distinct, baffled vocal-tract vowel configurations, Blandin *et al.*^{37,38} and Brandner *et al.*³⁹ found that internally propagating cross-modes impacted associated high-frequency directivities in relatively narrow bands. The predicted effects were not apparent in the present study because the data block sizes, long-term averages, and human subject averages coalesced many running-speech phonemes into wider 1/3-octave bands.

2. Average vs KEMAR directivities

Figure 17 shows the 1/3-octave-band $L_{AWRMSD}(f)$ [see Eqs. (8)–(12)] between the KEMAR HATS directivity and a six-talker male-female average directivity as the reference. The energetically averaged level from 160 Hz to 10 kHz was 1.6 dB. The figure also shows the $L_{AWRMSD}(f)$ between the

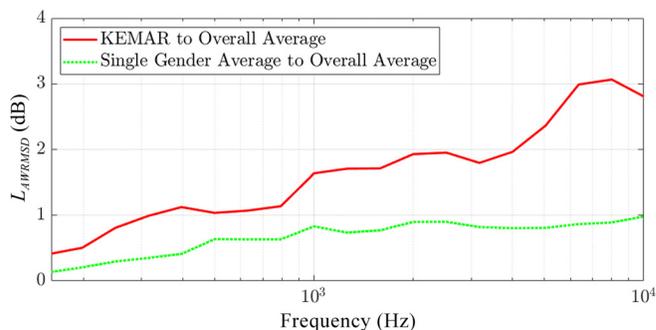


FIG. 17. (Color online) The $L_{AWRMSD}(f)$ between the directivity of the KEMAR HATS and live speech averaged over three males and three females, plus the $L_{AWRMSD}(f)$ between the average of one gender and the overall average.

average single-gender directivity and the overall average directivity as the reference. Because the latter derives from an equal number of male and female talkers, the $L_{AWRMSD}(f)$ between the average male or female directivity and the overall average is the same. The energetically averaged $L_{AWRMSD}(f)$ from 160 Hz to 10 kHz was 0.7 dB. The graphs show that as frequency increases, deviations between directivity patterns tend to increase. This effect is due to the distinct radiation and diffraction characteristics of the HATS and individual talkers.⁵⁹

3. Previous measurements

Figure 18(a) presents 1/3-octave-band directivities of a soprano vocalist, based on post-processing and degree $N = 4$ spherical harmonic expansions of 32-point recordings published by Weinzierl *et al.*^{61,62} Figure 18(b) shows degree $N = 10$ expansions of the 2522-point measurements averaged across three female talkers of the present study. The two sampling methods produce comparable results at lower frequencies, with similar orientations of major lobes and other features. However, the 2522-point measurements allow higher-degree expansions and reveal important details beyond those available from the lower-resolution results, even with the subject averaging (also compare Fig. 14 for a single female talker expanded to degree $N = 15$).

Figure 19 shows the $L_{AWRMSD}(f)$ between the results of a single female talker and the average female talker, both expanded to degree $N = 10$, and those of the soprano vocalist expanded to degree $N = 4$. Both curves show trends of increasing deviation with increasing frequency, just as for the KEMAR HATS in Fig. 17. The energetic averages of the $L_{AWRMSD}(f)$ curves over the 1/3-octave bands from 160 Hz to 10 kHz were 1.3 dB for the single female talker and 1.9 dB for the soprano vocalist.

Figure 20 compares directivity plots for an average talker derived from Chu and Warnock’s tables¹⁰ to the average talker directivity taken from the 2522-point measurements, expanded to degree $N = 15$. Chu and Warnock¹⁰ sampled at 92 unique positions over a partial hemisphere and assumed symmetry about the median plane. Their results compared reasonably with those of Dunn and

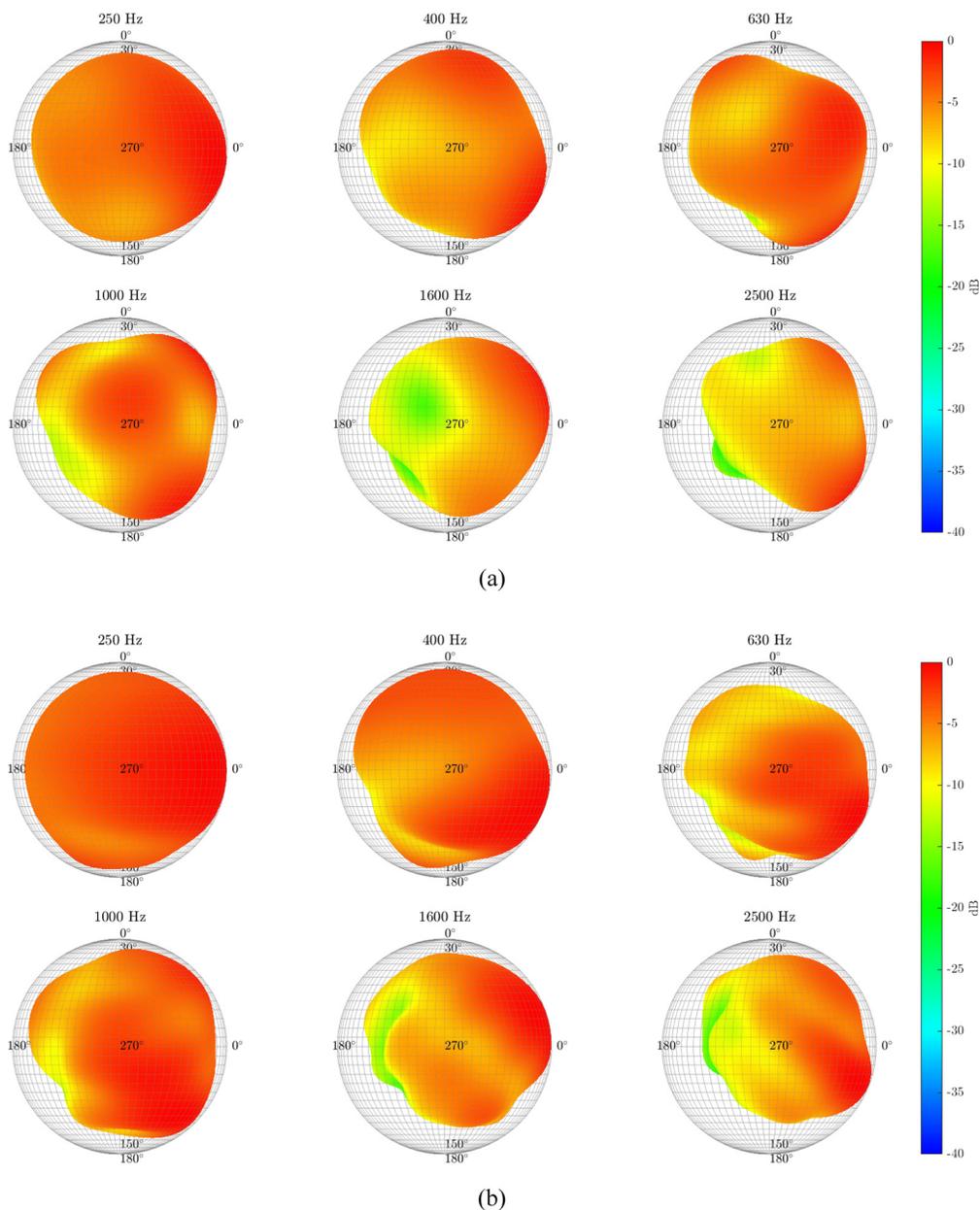


FIG. 18. (Color online) Comparative side views of 1/3-octave-band directivity balloons based on spherical harmonic expansions. (a) A soprano singer reported in Weinzierl *et al.* (Refs. 61 and 62) sampled at 32 points and expanded to maximum degree $N=4$. (b) An average of three female talkers from the present study sampled at 2522 points and expanded to $N=10$.

Farnsworth⁵ (20 maximum positions over a partial hemisphere of fixed radius), Moreno and Pfretzschner⁶ (720 effective circular positions in the transverse and median planes), and McKendree⁸ (seven positions in a right transverse semicircle plus an eighth in a front median semicircle). The frequency-dependent results of the present study likewise agree with those of Chu and Warnock, including diffraction effects around the back of the average talker. However, the 2522-point measurements again reveal more richness of detail over a complete sphere.

Through previously available data, some authors have noted nearly omnidirectional radiation at low frequencies and the formation of a rear shadow region and a persistent

rear lobe at higher frequencies. They have also observed the apparent formation, shifting, and multiplication of other lobes in the front hemisphere with increasing frequency plus other effects. Some have noted that the median plane's dominant radiation axis often falls below the transverse plane and shifts with frequency. This work confirms and expands upon their observations through more complete representations.

D. Angular sampling resolution

The comparisons in Fig. 18 demonstrate increasing discrepancies between degree $N=4$ and degree $N=10$

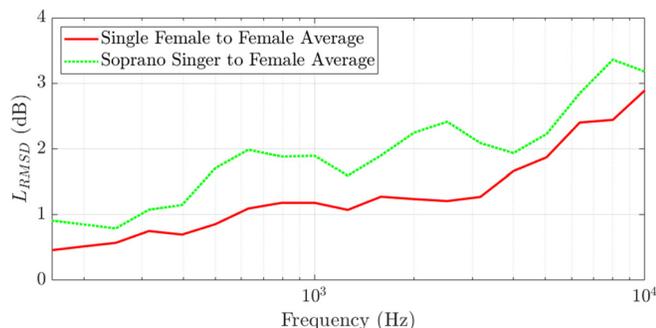


FIG. 19. (Color online) The $L_{AWRMSD}(f)$ between the expanded degree $N=10$ directivity of a single female talker and the degree $N=10$ female average in Fig. 18(b), plus the $L_{AWRMSD}(f)$ between the expanded degree $N=4$ directivity produced by the soprano vocalist in Fig. 18(a) and the same female average.

expansions with increasing frequency, some of which may result from spatial aliasing produced by the 32-point measurements. Although the degree $N=4$ plots in Fig. 18(a) appear functional, comparisons to Figs. 14 and 18(b) show that many significant details are lost at higher frequencies, escaping representation except through denser sampling and higher-degree expansions (also compare Fig. 20). While Figs. 12, 13, and 18(b) present results from degree $N=10$ expansions, the 2522-point measurements allow even higher-degree expansions, such as those depicted in Figs. 14 and 20(b). These illustrate initial requirements to capture fine details and the value of higher-resolution measurements.

E. Limitations

The results of this work derived from averaged autospectra, cross-spectra, FRFs, and effective coherent output spectra over six short phonetically balanced sentences; they did not focus on individual phonemes or the dynamic directivities associated with changes in vocal-tract configurations.^{37–39} Variations in fundamental frequency and spectrum occurred over the sentences and with each azimuthal-angle increment, but the same 1/3-octave bands likely subsumed them. All subjects in the study were seated for practicality and stability within the rotating reference frame. Future measurements could include standing talkers to explore differences in the diffraction and absorption of upright bodies.

While additional subjects would statistically improve the results, this paper’s primary aim was to address measurement and processing methods and provide initial outcomes and comparisons. Because high-resolution directivity measurements are tedious and time-consuming, evaluation of substantial subject sample sizes would require considerable effort for planning, subject selection, training, measurement, and processing. As mentioned earlier, a symmetrizing of spherical measurements about the median plane⁶³ can effectively double the number of subject quasi-averages and reduce the required number of talkers.

The 2522-point measurements produced relatively high angular resolutions, but sampling remained limited. Spherical harmonic expansions converged well to the

measured patterns within the possible number of coefficients, but more research is necessary to substantiate the required frequency-dependent sampling density for full speech bandwidth. As mentioned in Sec. IV A, the estimated upper spectral bound for this study, based on coherence values, transfer functions, coherent output spectra, broadband summation, and multiple-subject averaging, was roughly 10 kHz, but this limit may require additional investigation. Radiation above this frequency, though down considerably in amplitude, could involve more complex directivity patterns, necessitating increased sampling resolution. Future studies could explore these topics.

V. CONCLUSIONS

This work has explored the measurement of time-averaged directivity of live phonetically balanced speech using a multiple-capture transfer function method. It has allowed analysis of directivities in 5° polar and azimuthal-angle resolutions, as advocated currently for loudspeakers. The measurements leveraged transfer functions, coherence functions, and block averaging to mitigate repetition variations in amplitude and spectrum, plus noise in narrow measurement bands. Broader-band directivities followed from effective coherent output spectra, narrowband summations, and subject averaging. Comparisons to the results of theoretical models, measured HATS directivities, and other researchers’ work served to validate the results. The details have shown that while low-frequency speech content is much less directional than high-frequency content, radiation from the mouth and nostrils and diffraction about the talker and chair cause directivity patterns to evolve remarkably over frequency.

Some results of the investigation are available in archived electronic files for readers to use in their work. The authors hope that these resources will increase understanding of speech radiation and diffraction and enhance scientific studies and applications involving the human voice. The files will help improve computational models and applications in architectural acoustics, audio, and communications.

Future work in this area could expand upon these methods and results, adapting them to higher resolutions as warranted. It could explore individual phoneme directivities, dynamic directivities, longer-time averages of continuous speech, radiation and diffraction of standing talkers, and larger talker sample sizes. It could further investigate directivity indices and sound power spectra produced by speech. Comprehensive near-field measurements would lead to a greater understanding of radiated fields at varying distances and allow in-depth studies of microphone placement techniques. The authors encourage research in these and related areas and the application of the results.

ACKNOWLEDGMENTS

The authors express sincere thanks to J. Bodon, J. Eyring, W. Strong, W. Lifferth, J. Petersen, T. Neilsen, and

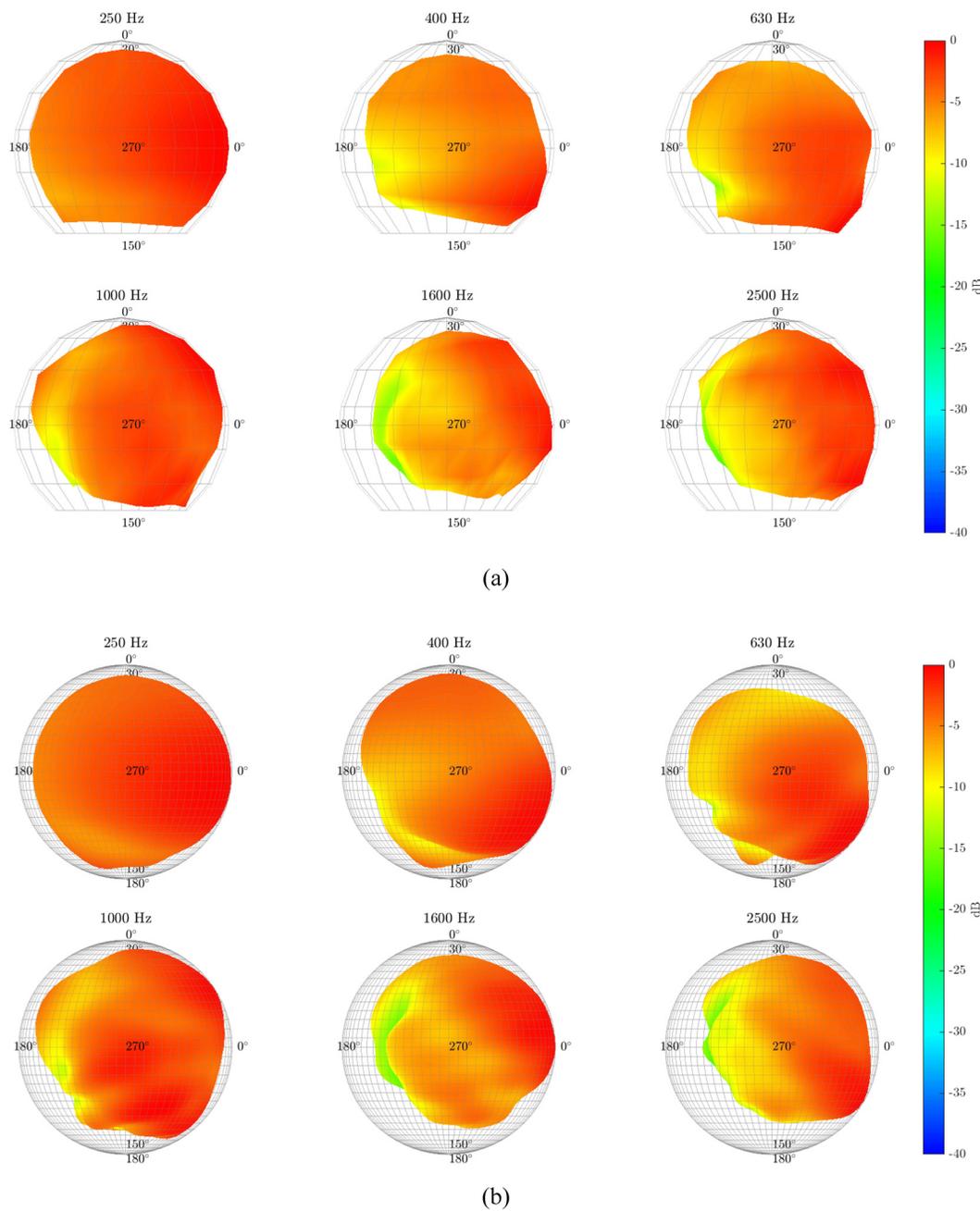


FIG. 20. (Color online) Comparative side views of 1/3-octave directivity balloons. (a) Average speech directivity reported in a partial hemisphere by Chu and Warnock (Ref. 10). (b) Average male and female speech directivity from the present study expanded to degree $N = 15$. The directivities measured by Chu and Warnock did not allow spherical harmonic expansions.

E. Hunter, who all made meaningful contributions and suggestions related to this work. The research was funded in part by the William James and Charlene Fuhriman Strong Family Musical Acoustics Endowed Fellowship Fund and the National Institutes of Health Grant No. R01DC012315. The content is solely the authors' responsibility and does not necessarily represent the views of the National Institutes of Health. The authors have no conflicts of interest to declare.

¹AES56-2008 (r2019): *AES Standard on Acoustics: Sound Source Modeling: Loudspeaker Polar Radiation Measurements* (Audio Engineering Society, New York, 2019).

²CLF Group, "CLF: A common loudspeaker format," *Syn-Aud-Con Newsl.* **32**(4), 14–17 (2004).

³EASE SpeakerLab Software Manual, Rev. 1.0 (Ahnert Feistel Media Group, Berlin, 2016), <http://www.afmg-support.eu/SoftwareDownloadBase/AFMG/EASE%20SpeakerLab/EASE%20SpeakerLab%20User's%20Guide.pdf> (Last viewed July 24, 2020).

⁴F. Trendelenburg, "Beitrag zur Frage der Stimmrichtwirkung" ("Contribution to the question of voice directivity"), *Z. Tech. Phys.* **10**/11, 558–563 (1929).

⁵H. K. Dunn and D. W. Farnsworth, "Exploration of pressure field around the human head during speech," *J. Acoust. Soc. Am.* **10**(1), 184–199 (1939).

⁶A. Moreno and J. Pfretzschner, "Human head directivity in speech emission: A new approach," *Acoust. Lett.* **1**, 78–84 (1978).

- ⁷G. Studebaker, "Directivity of the human vocal source in the horizontal plane," *Ear Hear.* **6**(6), 315–319 (1985).
- ⁸F. S. McKendree, "Directivity indices of human talkers in English speech," *Proceedings of Inter-Noise 86*, Cambridge, MA (July 21–23, 1986), pp. 911–916.
- ⁹A. C. C. Warnock, W. T. Chu, and J.-C. Guy, "Directivity of human talkers," *Can. Acoust.* **30**(3), 36–37 (2002).
- ¹⁰T. Chu and A. C. C. Warnock, "Detailed directivity of sound fields around human talkers," Research Report IRC-RR-104 (Institute for Research in Construction, National Research Council of Canada, Ottawa ON, Canada (2002), pp. 1–47).
- ¹¹B. F. G. Katz, F. Prezati, and C. d'Alessandro, "Human voice phoneme directivity pattern measurements," *J. Acoust. Soc. Am.* **120**, 3359 (2006); for the presentation slides see https://www.researchgate.net/profile/Brian_Katz5/publication/236213231_Human_voice_phoneme_directivity_pattern_measurements/links/565d69b708aefc2aac781c6/Human-voice-phoneme-directivity-pattern-measurements.pdf (Last viewed July 24, 2020).
- ¹²P. Kocon and B. B. Monson, "Horizontal directivity patterns differ between vowels extracted from running speech," *J. Acoust. Soc. Am.* **144**(1), EL7–EL13 (2018).
- ¹³B. F. G. Katz and C. d'Alessandro, "Measurement of 3D phoneme-specific radiation patterns in speech and singing," Scientific Report [Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMS), Orsay, France, 2007], https://rs2007.limsi.fr/PS_Page_14.html (Last viewed July 24, 2020).
- ¹⁴B. B. Monson, E. J. Hunter, and B. H. Story, "Directivity of low- and high-frequency energy in speech and singing," *J. Acoust. Soc. Am.* **132**(1), 433–441 (2012).
- ¹⁵A. H. Marshall and J. Meyer, "The directivity and auditory impressions of singers," *Acustica* **58**, 130–140 (1985).
- ¹⁶B. F. G. Katz and C. d'Alessandro, "Directivity measurements of the singing voice," *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain (September 2–7, 2007), Vol. 4, pp. 1976–1981.
- ¹⁷H. Niese, "Untersuchung über die Knallform bei raumakustischen Impulsmessungen" ("Investigation of the bang shape in room acoustic impulse measurements"), *Hochfrequenztechn. Elektroak.* **65**(3), 98–108 (1956).
- ¹⁸H. F. Olson, "Field-type artificial voice," *J. Audio Eng. Soc.* **20**(6), 446–452 (1972).
- ¹⁹F. Bozzoli and A. Farina, "Directivity balloons of real and artificial mouth simulators for measurement of the speech transmission index," *Proceedings of the 115th Convention of the Audio Engineering Society*, New York (October 10–13, 2003), paper 5953.
- ²⁰F. Bozzoli, M. Viktorovitch, and A. Farina, "Balloons of directivity of real and artificial mouth used in determining speech transmission index," *Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Spain (May 28–31, 2005), paper 6492.
- ²¹F. Bozzoli, P. Bilzi, and A. Farina, "Influence of artificial mouth's directivity in determining speech transmission index," *Proceedings of the 119th Convention of the Audio Engineering Society*, New York (October 7–10, 2005), paper 6571.
- ²²T. Halkosaari, "Radiation directivity of human and artificial speech," M.S. thesis, Helsinki University of Technology, Espoo, Finland, 2004.
- ²³T. Halkosaari and M. Vaalgamaa, "Directivity of human and artificial speech," *Proceedings of the Joint Baltic-Nordic Acoustics Meeting 2004*, Mariehamn, Åland, Finland (June 8–10, 2004).
- ²⁴T. Halkosaari, M. Vaalgamaa, and M. Karjalainen, "Directivity of artificial and human speech," *J. Audio Eng. Soc.* **53**(7/8), 620–631 (2005).
- ²⁵H. Jers and M. Kob, "Nachbildung eines Chores für raumakustische und musikalische Untersuchungen" ("Simulation of a choir for room acoustic and musical investigations"), *Proceedings of 20. Tonmeistertagung*, Karlsruhe, Germany (November 20–23, 1998), pp. 208–217.
- ²⁶M. Kob and H. Jers, "Directivity measurement of a singer," *Collected Papers from the Joint Meeting Berlin 1999: 137th Regular Meeting of the Acoustical Society of America, 2nd Convention of the European Acoustics Association, Forum Acusticum 1999, Integrating the 25th German Acoustics DAGA Conference*, Berlin, Germany (March 14–19, 1999), paper 2aMU19.
- ²⁷H. Jers, "Directivity measurements of adjacent singers in a choir," *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain (September 2–7, 2007).
- ²⁸D. Cabrera, P. J. Davis, and A. Connolly, "Long-term horizontal vocal directivity of opera singers: Effects of singing projection and acoustic environment," *J. Voice* **25**(6), e291–e303 (2011).
- ²⁹M. Brandner, M. Frank, and D. Rudrich, "DirPat: Database and viewer of 2D/3D directivity patterns of sound sources and receivers," *Proceedings of the 144th Convention of the Audio Engineering Society*, Milan, Italy (May 23–26, 2018), e-Brief 425.
- ³⁰G. Stewart, "The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and audition," *Phys. Rev.* **33**(6), 467–479 (1911).
- ³¹K. Sugiyama and H. Irii, "Comparison of the sound pressure radiation from a prolate spheroid and the human mouth," *Acustica* **73**(5), 271–276 (1991).
- ³²C. Pörschmann and J. M. Arend, "Analyzing the directivity patterns of human speakers," *Proceedings of the 46th DAGA 2020*, Hannover, Germany (2020), pp. 1141–1144.
- ³³J. L. Flanagan, "Analog measurements of sound radiation from the mouth," *J. Acoust. Soc. Am.* **32**(12), 1613–1620 (1960).
- ³⁴J. Huopaniemi, K. Kettunen, and J. Rahkonen, "Measurements and modeling techniques for directional sound radiation from the mouth," *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY (October 17–20, 1999), pp. 183–186.
- ³⁵D. Todorovic, "Effect of head cover on directivity pattern of human head," *Proceedings of Forum Acusticum 2005*, Budapest, Hungary (August 29–September 2, 2005), pp. L193–L198.
- ³⁶C. Pörschmann and J. M. Arend, "A method for spatial upsampling of directivity patterns of human speakers by directional equalization," *Proceedings of the 45th DAGA 2019*, Rostock, Germany (2019), pp. 1458–1461.
- ³⁷R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "Influence of higher order acoustical propagation modes on variable section waveguide directivity: Application to vowel [a]," *Acta Acust. Unit. Acust.* **102**, 918–929 (2016).
- ³⁸R. Blandin, A. Van Hirtum, X. Pelorson, and R. Laboissière, "The effect on vowel directivity patterns of higher order propagation modes," *J. Sound Vib.* **432**, 621–632 (2018).
- ³⁹M. Brandner, R. Blandin, M. Frank, and A. Sontacchi, "A pilot study on the influence of mouth configuration and torso on singing voice directivity," *J. Acoust. Soc. Am.* **148**(3), 1169–1180 (2020).
- ⁴⁰B. A. Bartlett, "Tonal effects of close microphone placement," *J. Audio Eng. Soc.* **29**(10), 726–738 (1981).
- ⁴¹E. B. Brixen, "Spectral degradation of speech captured by miniature microphones mounted on persons' head and chest," *Proceedings of the 100th Convention of the Audio Engineering Society*, Copenhagen, Denmark (May 11–14, 1996), paper 4284.
- ⁴²E. B. Brixen, "Near-field registration of the human voice: Spectral changes due to positions," *Proceedings of the 104th Convention of the Audio Engineering Society*, Amsterdam, Netherlands (May 16–19, 1998), paper 4728.
- ⁴³E. M. Lai, G. A. Carrijo, R. Bennett, R. Togneri, M. Alder, and Y. Attikiozel, "An English language speech database at the University of Western Australia," *Proceedings of the ICASSP IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM (April 3–6, 1990), paper S2.17, pp. 101–104.
- ⁴⁴E. J. Hunter, L. C. Cantor-Cutiva, E. van Leer, M. van Mersbergen, C. D. Nanjundeswaran, P. Bottalico, M. J. Sandage, and S. Whitting, "Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue," *J. Speech Lang. Hear. Res.* **63**, 509–532 (2020).
- ⁴⁵P. Bottalico, "Speech adjustments for room acoustics and their effects on vocal effort," *J. Voice* **31**(3), 392.e1–392.e12 (2017).
- ⁴⁶B. T. Thornock, T. W. Leishman, B. E. Anderson, and J. J. Esplin, "Effects of simultaneous sound arrivals on direction-of-arrival estimates of the polar energy time curve," *Appl. Acoust.* **117**, 167–172 (2017).
- ⁴⁷P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.* **15**(2), 70–73 (1967).
- ⁴⁸J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 4th ed. (Wiley, Hoboken, NJ, 2010), Sec. 1.4, Ch. 6, Sec. 11.5.
- ⁴⁹K. G. McConnell and P. S. Varoto, *Vibration Testing: Theory and Practice*, 2nd ed. (Wiley, Hoboken, NJ, 2008), Sec. 5.6.

- ⁵⁰J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, 2nd ed. (Wiley, New York, 1993), Chs. 4, 9, 11.
- ⁵¹D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing* (Springer International, Switzerland, 2017), p. 15.
- ⁵²F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions* (Cambridge University Press, Cambridge, United Kingdom, 2010), Ch. 14.
- ⁵³R. A. Kennedy and P. Sadeghi, *Hilbert Space Methods in Signal Processing* (Cambridge University Press, Cambridge, United Kingdom, 2013), p. 189.
- ⁵⁴B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd ed. (Springer Nature, Cham, Switzerland, 2019), Ch. 3.
- ⁵⁵S. D. Bellows and T. W. Leishman, "Spherical harmonic expansions of high-resolution musical instrument directivities," *Proc. Mtgs. Acoust.* **35**, 035005 (2018).
- ⁵⁶T. W. Leishman, S. Rollins, and H. M. Smith, "An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound," *J. Acoust. Soc. Am.* **120**(3), 1411–1422 (2006).
- ⁵⁷W. Ahnert, J. Raid, S. Feistel, and P. Meyer, "Accurate electroacoustic prediction utilizing the complex frequency response of far-field polar measurements," *Proceedings of the 108th Convention of the Audio Engineering Society*, Paris, France (February 19–22, 2000), paper 5129.
- ⁵⁸S. Feistel and W. Ahnert, "Modeling of loudspeaker systems using high-resolution data," *J. Audio Eng. Soc.* **55**(7/8), 571–597 (2007).
- ⁵⁹S. Bellows and T. Leishman, "High-resolution analysis of the directivity factor and directivity index functions of human speech," *Proceedings of the 146th Convention of the Audio Engineering Society*, Dublin, Ireland (March 20–23, 2019), paper 10173.
- ⁶⁰B. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *J. Acoust. Soc. Am.* **123**(1), 327–335 (2008).
- ⁶¹N. R. Shabtai, G. Behler, M. Vorländer, and S. Weinzierl, "Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments," *J. Acoust. Soc. Am.* **141**(2), 1246–1256 (2017).
- ⁶²S. Weinzierl, M. Vorländer, G. Behler, F. Brinkmann, H. von Coler, E. Detzner, J. Krämer, A. Lindau, M. Pollow, F. Schulz, and N. R. Shabtai, "A database of anechoic microphone array measurements of musical instruments," Technische Universität Berlin (2017), <https://doi.org/10.14279/depositonce-5861.2> (Last viewed July 24, 2020).
- ⁶³S. D. Bellows, C. M. Pincock, J. K. Whiting, and T. W. Leishman, "Average Speech Directivity," Brigham Young University ScholarsArchive (2019). Directivity. 1. <https://scholarsarchive.byu.edu/directivity/1> (Last viewed January 21, 2021).
- ⁶⁴P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 340–341.
- ⁶⁵S. Bellows and T. Leishman, "Single-channel sound power estimation for reverberation effects," *Proceedings of the 149th Convention of the Audio Engineering Society*, Online (October 27–30, 2020), paper 10413.
- ⁶⁶M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**(1), 214–222 (1975).