MULTIPLE SOURCE LOCALIZATION IN THE SPHERICAL HARMONIC DOMAIN USING AUGMENTED INTENSITY VECTORS BASED ON GRID SEARCH

Sina Hafezi, Alastair H. Moore and Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, UK
{s.hafezi14, alastair.h.moore, p.naylor}@imperial.ac.uk

ABSTRACT

Multiple source localization is an important task in acoustic signal processing with applications including dereverberation, source separation, source tracking and environment mapping. When using spherical microphone arrays, it has been previously shown that Pseudo-intensity Vectors (PIV), and Augmented Intensity Vectors (AIV), are an effective approach for direction of arrival estimation of a sound source. In this paper, we evaluate AIV-based localization in acoustic scenarios involving multiple sound sources. Simulations are conducted where the number of sources, their angular separation and the reverberation time of the room are varied. The results indicate that AIV outperforms PIV and Steered Response Power (SRP) with an average accuracy between 5 and 10 degrees for sources with angular separation of 30 degrees or more. AIV also shows better robustness to reverberation time than PIV and SRP.

Index Terms— spherical microphone arrays, localization, direction-of-arrival estimation, spherical harmonic, intensity vector

1. INTRODUCTION

Direction-of-arrival (DOA) estimation is an important task in acoustic signal processing and is used in spatial filtering, source separation, source tracking, environment mapping, dereverberation, speech enhancement and robot audition [1, 2]. Spherical Microphone Arrays (SMAs) have recently become a popular tool in speech acquisition due to their ability to analyse sound in three-dimension [3, 4]. Unlike linear, planar and circular microphone arrays, SMAs have no orientation due to symmetry and therefore provide direction-independent resolution and accuracy. In this paper, we address multiple source localization using SMAs.

A range of methods for source localization using SMAs has been studied over the past decades and can be categorised into four groups of steered response power methods [2, 5–7], subspace methods [8, 9], maximum likelihood methods [10, 11], and intensity vector-based methods [12, 13].

The pseudo-intensity vector method is an attractive DOA estimator for SMAs as it is fast to compute and provides good localization accuracy for a single source [12]. However, as with most localization algorithms, as the level of reverberation and the number of sound sources increase, localization accuracy is reduced [14]. On the other hand, our recently proposed augmented intensity vector, AIV, method [15], has potential for better localization accuracy compared to PIV for a single source. This paper evaluates the performance of AIV method for multiple sources in different conditions of various numbers of sources, various reverberation times (RTs) and various source separations.

This paper is structured as follows. Section 2 briefly reviews the background theory of spherical harmonics, PIV and SRP. Section 3 introduces our previously proposed AIV method. Section 4 explains the smoothing of spatial spectrum obtained from intensity vectors, and finally in Section 5 we evaluate the accuracy and robustness of AIV compared to PIV and SRP in the presence of multiple sources.

2. TECHNICAL BACKGROUND

In this section, we briefly review the Spherical Harmonic Domain (SHD) representation for SMA signals and PIV-based DOA estimation. We also present SRP in the SHD.

2.1. Spherical Harmonics

Consider a point $(r, \Omega) = (r, \theta, \varphi)$ in spherical coordinates with range r, inclination θ and azimuth φ. Let $p(k, r, \Omega)$ denote the sound pressure field at this point where k is the wavenumber. The Spherical Harmonic Transform (SHT) of the soundfield is [16]

$$p_{lm}(k, r) = \int_{\Omega \in S^2} p(k, r, \Omega) Y_{lm}^{*}(\Omega)\, d\Omega, \qquad (1)$$

where $\int_{\Omega \in S^2} d\Omega = \int_0^{2\pi} \int_0^{\pi} \sin(\theta)\, d\theta d\varphi$, and $(.)^*$ denotes the complex conjugate. The spherical harmonic basis function $Y_{lm}(\Omega)$ of order l and degree m (satisfying $|m| \leq l$) is given

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465

by [16]

$$Y_{lm}(\Omega) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}} P_{lm}(\cos(\theta))\, e^{im\varphi}, \quad (2)$$

where P_{lm} is the associated Legendre function and $i^2 = -1$.

The eigenbeams or planewave decomposition of the soundfield are obtained by compensating for the mode strength of the SMA, which depends on both its radius and its configuration (open or rigid sphere), according to

$$a_{lm}(k) = \frac{p_{lm}(k,r)}{b_l(kr)}. \quad (3)$$

The mode strength for a rigid SMA, as used in our experimental study, is given by [16]

$$b_l(kr) = 4\pi i^l \left[j_l(kr) - \frac{j_l^{'}(kr)}{h_l^{(2)'}(kr)} h_l^{(2)}(kr) \right], \quad (4)$$

where j_l is the spherical Bessel function of order l, $h_l^{(2)}$ is the spherical Hankel function of the second kind and of order l, and $(.)^{'}$ denotes the first derivative with respect to argument.

2.2. Pseudo-intensity vectors

The pseudo-intensity vector was proposed in [12] and is an approximation of the active intensity vector. It is calculated from the first order eigenbeams according to

$$\mathbf{I}(k) = \frac{1}{2}\Re\left\{ a_{00}(k)^* . \begin{bmatrix} D_x(k) \\ D_y(k) \\ D_z(k) \end{bmatrix} \right\}, \quad (5)$$

where

$$D_\nu(k) = \sum_{m=-1}^{1} Y_{1m}(\phi_\nu) a_{1m}(k), \quad \nu \in \{x,y,z\} \quad (6)$$

are dipoles steered in the opposite direction of Cartesian axes, given by $\phi_x = (\pi/2, \pi)$, $\phi_y = (\pi/2, -\pi/2)$ and $\phi_z = (\pi, 0)$.

The DOA unit vector $\mathbf{u}(k)$ is given by

$$\mathbf{u}(k) = -\frac{\mathbf{I}(k)}{\|\mathbf{I}(k)\|}, \quad (7)$$

where $\|.\|$ indicates a vector's ℓ_2-norm.

2.3. Steered Response Power in the Spherical Harmonic Domain

Steered Response Power [17] is used here as a baseline method for DOA estimation. In SRP, a beam with an arbitrary directivity pattern is used to scan different look directions to find the source directions as directions corresponding to the highest powers. Beamforming in SHD has been studied in [2] for example, in which the plane wave decomposition (PWD) beamformer was employed. The output of the beamformer steered into an arbitrary look direction Ω is given as [18]

$$y(\Omega) = \sum_k | \sum_{l=0}^{L_d} \sum_{m=-l}^{l} a_{lm}(k) Y_{lm}(\Omega) |^2, \quad (8)$$

where L_d is the maximum order used in the beamforming.

3. AUGMENTED INTENSITY VECTORS

In this section we introduce our recently proposed method, AIV [15], which employs higher order ($l > 1$) eigenbeams to improve the accuracy of DOAs obtained from PIVs. The spatial frequency of the spherical harmonic basis functions increases with the order and so incorporating the information from higher order eigenbeams allows an increased spatial resolution to be obtained.

Consider a plane wave $S(k) = \alpha(k)e^{i\beta(k)}$ with amplitude $\alpha(k)$, phase $\beta(k)$, and DOA $\Omega_u = (\theta_u, \varphi_u)$ arriving from a single source in the far-field. The SHT of this plane wave is given by

$$a_{lm}(k) = S(k) Y_{lm}^*(\Omega_u) + n_{lm}(k), \quad (9)$$

where $n_{lm}(k)$ is a residual due to noise and reverberation.

Approximating $S(k) = \sqrt{4\pi} a_{00}(k)$, by substituting (2) into (9) for $l = 0$ in a noise-free case, AIV aims to minimise the cost function

$$J(k,\Omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} | a_{lm}(k) - \sqrt{4\pi} a_{00}(k) Y_{lm}^*(\Omega) |^2, \quad (10)$$

where L is the maximum spherical harmonic order considered in the optimization.

The optimization is done in the form of a grid search across a grid of discrete look directions $\{\Omega_M\}$ within a search window of size $\triangle\Omega_M = (\triangle\theta_M, \triangle\varphi_M)$ centred on the initial DOA from the PIV. The optimized DOA $\Omega_s(k)$ is the direction in which the cost function $J(k, \Omega_s)$ is minimised

$$\Omega_s(k) = \arg\min_{\Omega} J(k,\Omega), \quad \Omega \in \{\Omega_M\}, \quad (11)$$

which is then converted into Cartesian coordinates to form the optimized DOA unit vector $\mathbf{u}_s(k)$. The Augmented Intensity Vectors $\mathbf{I}_s(k)$ are then formed using (7) and the initial intensity norm $\|\mathbf{I}(k)\|$ as

$$\mathbf{I_s}(k) = -\mathbf{u_s}(k)\|\mathbf{I}(k)\|. \quad (12)$$

4. SMOOTHING OF SPATIAL SPECTRUM

The spatial spectrum is achieved by making a 2D histogram (inclination vs azimuth) using the quantized directions of the

intensity vectors. Due to noisy observations and the presence of multiple irregular peaks in the 2D histogram, we employ smoothing on the spatial spectrum. In our smoothing process, a spatial window is centred at each spatial sample, which is then replaced by the weighted average of the values within the smoothing window. We employ a Gaussian smoothing kernel centred on the look direction Ω expressed as

$$K_{\theta_i,\varphi_i}(\Omega) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\angle(\Omega,\Omega_{\theta_i,\varphi_i})^2}{2\sigma^2}\right), \qquad (13)$$

where $\Omega_{\theta_i,\varphi_i}$ is the direction of inclination θ_i and azimuth φ_i, and σ denotes the standard deviation, which is chosen empirically as defined in Section 5. The kernel is truncated by removing the entries with $K < 0.001$.

5. EVALUATION

5.1. Accuracy

We calculate the DOA estimation error $\varepsilon_{\mathbf{u_o},\mathbf{u_s}}$ (in degrees) between a true DOA unit vector $\mathbf{u}_o$ and an estimated DOA unit vector $\mathbf{u}_s$ as

$$\varepsilon_{\mathbf{u_o},\mathbf{u_s}} = \cos^{-1}\left(\mathbf{u}_o^T\mathbf{u}_s\right). \qquad (14)$$

For multiple sources and equal number of estimated DOAs, the average DOA estimation error depends on how we associate the true DOAs and the estimated DOAs in (14). The average errors for all possible sets of pairs are calculated using (14), and the minimum average error is chosen as the final DOA estimation error.

Three evaluations are conducted using a simulated room environment and a SMA. The Acoustic Impulse Responses (AIRs) of a 32-element rigid spherical microphone array were simulated using Spherical Microphone arrays Impulse Response Generator (SMIRgen) [19] based on Allen & Berkley's image method [20]. The array with radius 4.2 cm is placed at $(2.54, 2.55, 4.48)$ m in a $5 \times 4 \times 6$ m shoebox room. N_s number of sources are distributed on a circle of radius 1 m around the SMA with the same height as the centre of the array. In each trial, the azimuth of the first source is chosen randomly from a uniform distribution around the sphere and the subsequent sources are placed at regularly spaced intervals $\triangle\phi_s$ (the values of N_s and $\triangle\phi_s$ are provided later in each experiment as they differ in each experiment). The source signals consist of different anechoic speech signals randomly selected for each trial from the APLAWD database [21]. The active level of each speech source according to ITU-T P.56 [22], as measured at p_{00}, is set to be equal across all trials. Spatio-temporally white Gaussian noise is added to the microphone signals to produce a signal to incoherent noise ratio (iSNR) of 25 dB at p_{00} for each source.

A sampling frequency of 8 kHz was used with frame length of 4 ms and 50% overlapping of time frames. PIV and

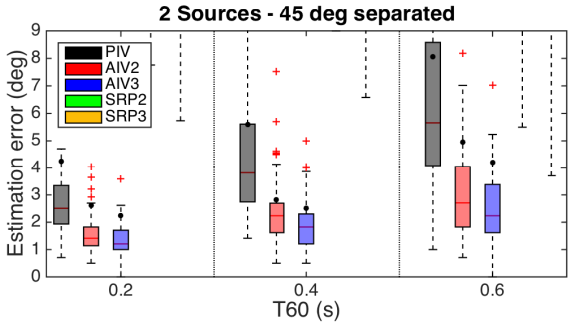

Fig. 1. The effect of T_{60}. The boxes for SRPs are out of the y-axis limit as SRP2 and SRP3 have medians of $\{45.6, 44.5, 64.0\}$ and $\{46.5, 44.7, 53.0\}$ degrees respectively for $T_{60} = \{0.2, 0.4, 0.6\}$ s.

two versions of AIV and PWD-SRP are used with different orders $L = \{2, 3\}$ (respectively referred as AIV2 and AIV3, and SRP2 and SRP3) in (10) and (8). SRP is steered over a search grid of 181×360 degrees (inclination $\times$ azimuth) while AIV has a search window of size 10×10 degrees in (11), centred on an initial direction indicated, in this example, by PIV. The smoothing kernel in (13) has $\sigma = 4$ degrees.

In experiment 1 the effect of Reverberation Time (RT) is evaluated for 2 sources with 45 degrees separation. Figure 1 shows the distribution of errors (100 trials per test condition) for $T_{60} = \{0.2, 0.4, 0.6\}$ s. The boxes show the mean as the black dot, median as the red horizontal line, upper and the lower quartiles, and the whiskers extend to 1.5 times the interquartile range based on Monte Carlo simulations. The median errors are $\{2.5, 3.8, 5.6\}$ degrees for PIV, $\{1.2, 1.8, 2.2\}$ degrees for AIV3, and $\{46.5, 44.6, 53\}$ degrees for SRP3 respectively for all RTs. We can see the clear improvement in AIV compared to PIV as the means, medians and the interquartile ranges are reduced in all RTs.

In experiment 2 the effect of angular separation of two

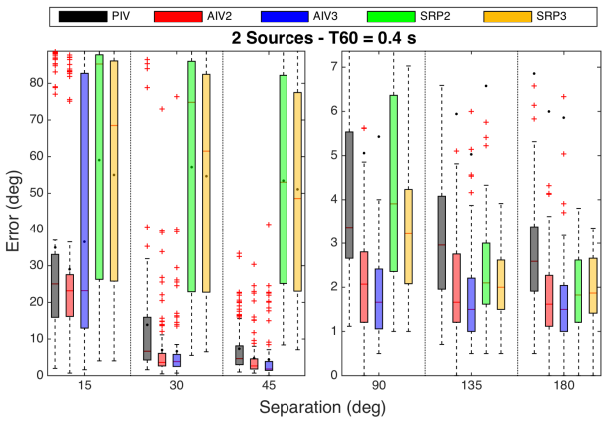

Fig. 2. The effect of separation angle

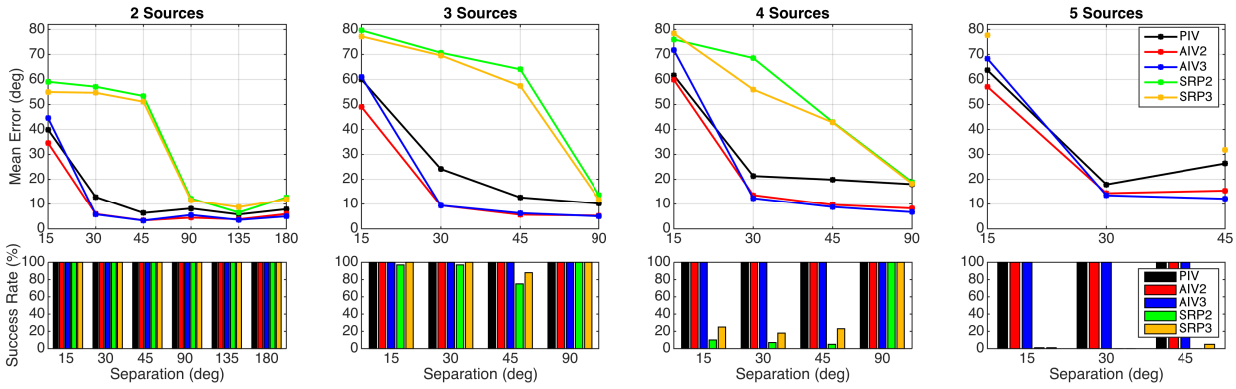

Fig. 3. The effect of number of sources and separation angle

sources is evaluated for a moderate $T_{60} = 0.4$ s. Figure 2 shows the distribution of errors over 100 trials for $\triangle\phi_s = \{15,\ 30\ ,45\ ,90\ ,135\ ,180\}$ degrees. For clarity of display, the results are split in two plots. For separations above 30 degrees, AIVs significantly outperform all methods with the medians around 2 degrees (± 0.5 degree) while PIV and SRPs have very high variations on median as they change from 26 to 3 degrees and 72 to 2 degrees respectively. For the case of $\triangle\phi_s = 15$ degrees we observe poor accuracy for all methods although AIVs still performs most accurately.

In experiment 3 the effect of angular separation and number of sources is evaluated for $T_{60} = 0.4$ s. We define the 'Success Rate' of a method as the number of times (in $\%$) that method successfully estimate the correct number of DOAs corresponding to the number of sources. A failure is when the method's spatial spectrum has fewer peaks than the number of sources. Fig. 3 shows the success rate and the average error of 100 trials as a function of $\triangle\phi_s = \{15,\ 30\ ,45\ ,90\ ,135\ ,180\}$ degrees for $N_s = \{2,\ 3\ ,4\ ,5\}$. The intensity based methods (PIV and AIVs) have full success in estimation whereas SRPs start to fail as the number of sources increases from 3 to 5 and the separation reduces from 45 to 15 degrees especially for five close sources, where SRPs rarely or never succeed. For separations above 30 degrees, AIVs significantly outperform all methods with average error of $\{5,\ 7,\ 10,\ 12\}$ degrees respectively for $\{2,\ 3\ ,4\ ,5\}$ sources whereas PIV has average error of $\{8,\ 18,\ 20,\ 23\}$ degrees respectively.

5.2. Computation Complexity

In this section we discuss the number of computations required in each method for a single TF-bin in terms of the number of real multiplications. Note that the multiplication of two complex numbers is counted as four real multiplication while $|\ .\ |^2$ is counted as two real multiplications. We do not include the number of multiplications in (2) as we precalculate and store all the required $Y_{lm}\left(\Omega\right)$.

For PIV, we have 48 $(3 \times 3 \times 4 + 3 \times 4)$ operations where the numbers in parentheses respectively represent the number of axes, harmonic modes and the real multiplications in (6) and the number of axes and the real multiplications in (5).

For AIV using (10), we have $48 + (L+1)^2 \times (2+4+2)$ operations for a single look direction where the numbers respectively correspond to the PIV, number of eigenbeams up to the order L, a real-complex followed by a complex-complex multiplications, and squared magnitude.

For PWD-SRP using (8), we have $(L_d+1)^2 \times 4 + 2$ operations for a single look direction where the numbers respectively correspond to the number of eigenbeams up to the order L_d, a complex-complex multiplication followed by a squared magnitude. For a full-grid of 181×360, PWD-SRP has millions of operations per TF-bin. This results in significantly higher computation complexity for PWD-SRP compared to AIV and PIV.

6. CONCLUSIONS

We presented an evaluation of AIV-based DOA estimation for multiple source scenarios. The results show that AIV clearly outperforms the previous PIV-based method, and the baseline method PWD-SRP in different reverberation times, source separation angles and the number of sources with noticeably improved robustness. For extremely low separation of 15 degrees all methods fail to accurately localize the sources. For angular separation of 30 degrees and above, AIV has average accuracy of 5 to 10 degrees while other methods have highly varying accuracy of worse than 10 degrees. The PWD-SRP method is unable to successfully localize all sources for three or more sources if they are closer than 45 degrees. The results also show that the third-order AIV estimator does not provide a noticeable advantage over the second-order AIV estimator for multiple source scenarios. It is shown that AIV has higher accuracy and robustness compared to a baseline PWD-SRP while using the same eigenbeams.

REFERENCES

[1] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[2] B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.

[3] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[4] E. Fisher and B. Rafaely, "The nearfield spherical microphone array," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 5272–5275.

[5] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 702–714, 2007.

[6] S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 361–371, Feb. 2011.

[7] H. Sun, S. Yan, and U. P. Svensson, "Robust minimum sidelobe beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 1045–1051, May 2011.

[8] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *The Journal of the Acoustical Society of America*, vol. 132, pp. 261, 2012.

[9] H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2005, vol. 3, pp. iii/89–iii/92.

[10] S. Tervo and A. Politis, "Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 10, pp. 1539–1551, October 2015.

[11] J. L. Yuxiang Hu and X. Qiu, "A maximum likelihood direction of arrival estimation method for open-sphere microphone arrays in the spherical harmonic domain," *The Journal of the Acoustical Society of America*, vol. 138, pp. 791, 2015.

[12] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 442–446.

[13] A. H. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2015.

[14] C. Evers, A. H. Moore, and P. A. Naylor, "Multiple source localisation in the spherical harmonic domain," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Nice, France, July 2014.

[15] S. Hafezi, A. H. Moore, and P. A. Naylor, "3D acoustic source localization in the spherical harmonic domain based on optimized grid search," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Shanghai, China, 2016.

[16] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, London, first edition, 1999.

[17] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Acoustics, Speech and Signal Magazine*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[18] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication: Challenges and Perspectives*, I. Cohen, J. Benesty, and S. Gannot, Eds., chapter 11. Springer, Jan. 2010.

[19] D. P. Jarrett, "Spherical Microphone array Impulse Response (SMIR) generator," http://www.ee.ic.ac.uk/sap/smirgen/.

[20] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[21] G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," Technical report, University College London, June 1987.

[22] ITU-T, "Objective measurement of active speech level," Dec. 2011.