REAL-TIME INTERAURAL TIME DELAY ESTIMATION VIA ONSET DETECTION

Elizabeth Ren, Gustavo Cid Ornelas, and Hans-Andrea Loeliger

ETH Zurich, Dept. of Information Technology & Electrical Engineering
{ren, loeliger}@isi.ee.ethz.ch, cgustavo@student.ethz.ch

ABSTRACT

Reliable real-time estimation of the interaural time delay of a sound source is difficult in the presence of noise and reverberation. However, the psychoacoustical precedence effect suggests that accurate estimation is possible by concentrating on the first-arriving sound. This paper introduces a novel real-time estimation method inspired by the precedence effect. First, the arrival of first-arriving sound is detected by performing a hypothesis test based on local approximation of the binaural signal with exponentially decaying sinusoids, which effectively model the shape of a sound onset. After detection, the interaural time delay is directly retrieved from the phase shift of the approximating sinusoids. The local model approximation is done with efficient recursions by parameterization of the model with autonomous linear state-space models, making the algorithm implementable in real-time.

Index Terms— time delay estimation, interaural time delay, sound source localization, local model approximation

1. INTRODUCTION

Given an array of microphones, sound waves propagating through air will arrive at each microphone at a different time depending on the environment and the distance between the sound source and the microphone. Tracking the delay between the time of the sound arrival at the microphones therefore aids in the localization of the sound source.

Mammals in fact use the time delay between the sound waves arriving at the right and left ear as a cue for sound source localization [1]. This cue is referred to as the interaural time delay (ITD) and is commonly estimated for real-time binaural source localization. A high level of precision is needed for accurate estimation of the ITD, since for a typical human head size, the time delay is at most 660 μs, i.e., approximately 30 samples for a sampling frequency of 44.1 kHz [2]. However, many ITD estimation methods such as cross-correlation [3, 4] and similar methods, e.g., the popular generalized cross-correlation method (GCC) [5] only produce estimates from a discrete set. Moreover, reliable estimation without prior knowledge of the head-related impulse response (HRIR) is made difficult by environmental noise and reverberation [3, 6, 7]. Methods such as [8, 9, 10, 3] attempt to mitigate this problem by adaptive estimation of the filter response between the signals of both ears or of the HRIR. Another approach [11] selectively considers estimates based on a interaural coherence measure.

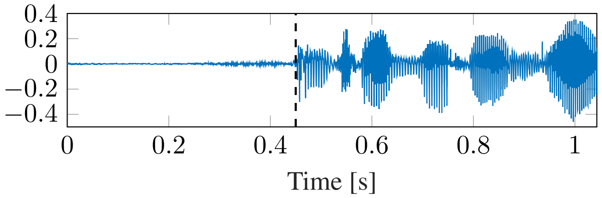

Fig. 1. Sound onset in a speech signal (indicated by line).

There is evidence that humans rely on extracting cues from the first-arriving sound, that is assumed to have travelled the direct path from the source to the ear, for source localization. This phenomenom, known as the precedence effect [12, 13, 14], suggests to restrict ITD estimation to first-arriving sound segments. In an acoustic signal, the first-arriving sound is usually visible as onsets of sound after periods of silence (cf. Fig. 1). This paper introduces a novel method for real-time ITD estimation based on the precedence effect. The main idea is to first determine whether an onset is present in the binaural signal, and estimate the ITD only if an onset was detected. The proposed algorithm can be efficiently implemented in real-time by local approximation of an appropriate autonomous linear state-space model (LSSM) [15, 16, 17] for the onset detection. A precise estimation is achieved, since the ITD is directly computed from the phase shift between the approximated model of each channel. The paper is structured as follows. The necessary theory on local model approximation is summarized in Section 2 before the estimation algorithm is described in Section 3. Experimental results for datasets of speech signals in various environments are discussed in Section 4, while Section 5 concludes the paper.

2. LOCAL MODEL APPROXIMATION

In this section, we summarize the theory of local model approximation needed for our algorithm. More on this subject can be found in [15, 16]. Given the observations $y_k \in \mathbb{R}$ of a discrete-time signal at time steps $k \in \mathbb{N}$, we consider the problem of locally approximating the signal at each time step $k \geq 1$ by a function $f : \mathbb{N}_0 \to \mathbb{R}$, that is the impulse response of an autonomous linear state-space model (LSSM) of the order n with parameters $\{c, A, s\}$, i.e.,

$$f(\ell) = cA^\ell s, \quad \ell \geq 0\,, \tag{1}$$

where the initial state vector $s \in \mathbb{R}^n$ and state transition matrix $A \in \mathbb{R}^{n\times n}$ are fixed. The trajectory of f is determined by the observation vector $c \in \mathbb{R}^{1\times n}$. In [15], it is shown that the set of functions produced by such LSSMs is a vector space consisting of linear combinations and products of exponentials, polynomials and sinusoids. The local approximation of y at each time step k is done by minimizing the cost function

$$\begin{aligned} J_k(c) &= \sum_{i=k-L+1}^{k} w(k-i)(y_i - f(k-i))^2 & (2)\\ &= \sum_{i=k-L+1}^{k} c_w A_w^{k-i} s_w (y_i - cA^{k-i}s)^2\,, & (3)\end{aligned}$$

where the start of (1) is assigned to k, $L \in \mathbb{N} \cup \{\infty\}$ is the length of the local window, and the fixed window weight function $w : \mathbb{N}_0 \to \mathbb{R}$ in (2) is also described by a LSSM (1) of the order n_w and parameters $\{c_w, A_w, s_w\}$. We assume $w(\ell) \propto \gamma^\ell$, where $0 < \gamma < 1$, making the window stable and $w(\ell) \to 0$ for $\ell \to \infty$, thereby minimizing the effect of samples further away from k on the fit. The approximating function value of y_i at time k is thus given by $\hat{c}_k A^{k-i} s$, where

$$\hat{c}_k = \underset{c\in\mathbb{R}^{1\times n}}{\operatorname{argmin}}\, J_k(c)\,. \tag{4}$$

The summation terms in the cost (2)-(3), which refer to unknown observations y_i for $i < 1$, are neglected. Following the derivation in [15, Ch. 6.2.2] (given for the case $L \to \infty$), (3) can be reformulated as

$$J_k(c) = c_w\chi_k - 2(c\otimes c_w)\zeta_k + (c\otimes c_w)s_k c^\mathsf{T}\,, \tag{5}$$

where the operator $\otimes$ is the Kronecker product and

$$\chi_k = \sum_{i=k-L+1}^{k} A_w^{k-i} s_w y_i^2 \in \mathbb{R}^{n_w} \tag{6}$$

$$\zeta_k = \sum_{i=k-L+1}^{k} (A\otimes A_w)^{k-i}(s\otimes s_w)y_i \in \mathbb{R}^{nn_w} \tag{7}$$

$$s_k = \sum_{i=k-L+1}^{k} (A\otimes A_w)^{k-i}(s\otimes s_w)s^\mathsf{T} A^{(k-i)\mathsf{T}} \in \mathbb{R}^{nn_w\times n} \tag{8}$$

which are computed by the recursions

$$\chi_k = A_w\chi_{k-1} + s_w y_k^2 - A_w^L s_w y_{k-L} ^2 \tag{9}$$

$$\begin{aligned}\zeta_k &= (A\otimes A_w)\zeta_{k-1} + (s\otimes s_w)y_k\\ &\quad - (A\otimes A_w)^L(s\otimes s_w)y_{k-L}\end{aligned} \tag{10}$$

initialized by $\chi_0 = 0$ and $\zeta_0 = 0$, while (8) can be shown to be independent of k and the signal y_i, and is thus precomputed. The cost (5) is in fact a quadratic function in c of the form

$$J_k(c) = \kappa_k - 2c\xi_k + cW_k c^\mathsf{T} \tag{11}$$

with quantities $\kappa_k \in \mathbb{R}$, $\xi_k \in \mathbb{R}^n$ and $W_k \in \mathbb{R}^{n\times n}$ that are computed from the parameters (6)-(8) according to

$$\kappa_k = c_w\chi_k \tag{12}$$

$$\{\xi_k\}_{j,p} = (P_{p,j}^{(1,n)} \otimes c_w)\zeta_k \tag{13}$$

$$\{W_k\}_{j',p'} = \operatorname{tr}\left((P_{p',j'}^{(n,n)} \otimes c_w)s_k\right)\,, \tag{14}$$

where $P_{i,j}^{(m,r)}$ is the $m \times r$-matrix with a one at index (i, j) and zero everywhere else. The minimization (4) is thus given by setting the derivative of (11) to zero:

$$\hat{c}_k = (W_k^{-1}\xi_k)^\mathsf{T}\,. \tag{15}$$

Inserting (15) in (11) yields the minimal cost

$$\min_{c\in\mathbb{R}^{1\times n}} J_k(c) = \kappa_k - \operatorname{tr}\left(\xi_k^\mathsf{T} W_k^{-1}\xi_k\right)\,. \tag{16}$$

For an infinite window, i.e., $L \to \infty$, the recursions (9)-(10) reduce to

$$\chi_k = A_w\chi_{k-1} + s_w y_k^2 \tag{17}$$

$$\zeta_k = (A\otimes A_w)\zeta_{k-1} + (s\otimes s_w)y_k \tag{18}$$

while (8) can either be replaced by its steady-state value [15, Ch. 6.2.4] (with some initialization error) or computed by the recursion initialized by $s_0 = 0$,

$$s_k = (A\otimes A_w)s_{k-1}A^\mathsf{T} + (s\otimes s_w)s^\mathsf{T}\,. \tag{19}$$

A measure for how well the assumed LSSM model (1) matches the local signal is the local cost ratio (LCR) [15, Ch. 7.7.2], [16], which compares the cost of local approximation with (1) and with a pure noise model

$$\begin{aligned}\mathrm{LCR}_k &= -\frac{1}{2}\log\left(\frac{\min_{c\in\mathbb{R}^{1\times n}} J_k(c)}{J_k(0)}\right) & (20)\\ &= -\frac{1}{2}\log\left(\frac{\kappa_k - \xi_k^\mathsf{T} W_k^{-1}\xi_k}{\kappa_k}\right) \geq 0\,, & (21)\end{aligned}$$

where we used (11) and (16). When (1) matches well with the signal, the fit cost is lower than that of a pure noise model. In contrast, when (1) does not fit well, the cost is close to that of a pure noise model. Therefore, by searching for segments when the LCR becomes larger, one can detect events in the form of (1) in the signal.

3. ESTIMATION ALGORITHM

The algorithm for ITD estimation is described in this section. We denote the samples of the binaural input signal by $y_k^E \in \mathbb{R}, k \geq 1$, where $E \in \{L, R\}$ refers to the left or right channel. The sampling frequency is denoted by f_s.

3.1. Onset Detection

We search for onsets in y_k by evaluating for each channel, the LCR (20)-(21) for local model approximation with a bank of LSSM signal forms (1) that each resemble an onset of a certain frequency. We choose to use a bank of $Q \in \mathbb{N}$ exponentially decaying sinusoids as LSSM signals (1), where the q-th LSSM signal, $q \in \{1, \dots, Q\}$, is given by

$$o_q(\ell) = \alpha_q \rho_q^\ell \sin(\Omega_q \ell + \phi_q) \quad \ell \geq 0, \tag{22}$$

where the decay $\rho_q \in (0, 1)$ and the normalized frequency Ω_q are fixed, while the amplitude $\alpha_q \in \mathbb{R}$ and phase $\phi_q \in [0, 2\pi]$ are fit to the signal. In order for the relation between the phase shift and the time delay to be unambiguous, the maximal possible time delay $\tau_{\max}$ must not exceed half of the period of the frequency f_q of the onset (which is related to the normalized frequency by $\Omega_q = 2\pi f_q / f_s$), i.e.,

$$f_q < \frac{1}{2\tau_{\max}} . \tag{23}$$

The LSSM of (22) is of order $n_q = 2$ with parameters

$$A = \rho_q \begin{bmatrix} \cos(\Omega_q) & -\sin(\Omega_q) \\ \sin(\Omega_q) & \cos(\Omega_q) \end{bmatrix} \tag{24}$$

$$s = \begin{bmatrix} 1 & 0 \end{bmatrix}^\mathsf{T} , \tag{25}$$

while the relation between the observation vector and the fit parameters is

$$c = \alpha_q \begin{bmatrix} \sin(\phi_q) & \cos(\phi_q) \end{bmatrix} . \tag{26}$$

For the cost (2) we use an infinite gamma window ($L = \infty$)

$$w(\ell) = \gamma^\ell \ell^3, \quad \ell \geq 0, \tag{27}$$

with decay $\gamma \in (0, 1)$, which should be larger than ρ_q of the onset models (22), such that the window properly captures the onset. The LSSM of (27) is of the order $n_w = 4$ and

$$c_w = \begin{bmatrix} 6 & 6 & 1 & 0 \end{bmatrix} \tag{28}$$

$$A_w = \gamma \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{29}$$

$$s_w = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\mathsf{T} . \tag{30}$$

The relevant recursions are (17)-(18) which are different for each onset model and channel, while the recursion (19) is independent of the signal. We denote the fit parameters (15) of the signal channel E and onset model q at time k by $\hat{c}^E_{q,k}$. The corresponding model fit (22) is denoted by $\hat{o}^E_{q,k}$ and the LCR (20) is denoted by $\mathrm{LCR}^E_{q,k}$. Fig. 2 shows an example of the fit $\hat{o}^E_{q,k}$ at the onset of a speech signal located at an azimuth of $-80°$, which is indicated by a peak of $\mathrm{LCR}^E_{q,k}$. One can see that the ITD is represented by the phase shift between the model fits to the signal.

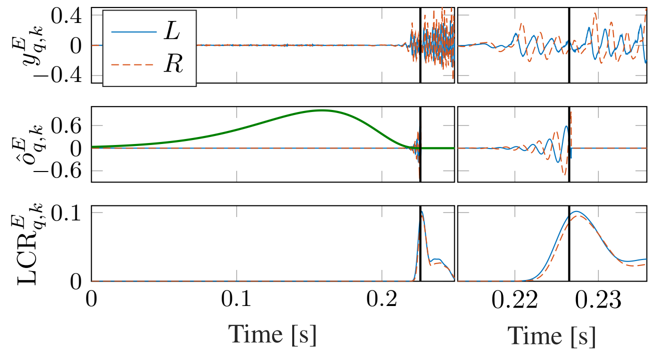

Fig. 2. Top: speech signal at azimuth of $-80°$. Middle: model fit at onset indicated by vertical lines, gamma window of cost in green. Bottom: associated LCR, onset detected at peak. Left: close-up on onset where ITD is more visible.

We detect an onset by searching for simultaneous peaks in the LCR of both channels of a onset model. This is done by local approximation of each $\mathrm{LCR}^E_{q,k}$ with a third degree polynomial

$$p(\ell) = \beta_0 + \beta_1 \ell + \beta_2 \ell^2 + \beta_3 \ell^3, \quad \ell \geq 0, \tag{31}$$

whose corresponding LSSM is of the order $n = 4$ with parameters (cf. (1)),

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{32}$$

$$s = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\mathsf{T} \tag{33}$$

$$c = \begin{bmatrix} \beta_3 & \beta_2 & \beta_1 & \beta_0 \end{bmatrix} . \tag{34}$$

For the cost (2), we use a finite-length exponential window

$$w(\ell) = \gamma^\ell, \quad 0 \leq \ell \leq L - 1, \tag{35}$$

where $\gamma \in (0, 1)$ and whose LSSM is of the order $n_w = 1$ with

$$c_w = s_w = 1 \tag{36}$$

$$A_w = \gamma . \tag{37}$$

The relevant recursions are (9)-(10) while (8) is precomputed. We denote the estimated polynomial coefficients (cf. (15),(34)) of $\mathrm{LCR}^E_{q,k}$ by $\hat{b}^E_{q,k} = (\hat{\beta}^E_{q,k,0}, \dots, \hat{\beta}^E_{q,k,3})$ and the corresponding polynomial (31) is denoted by $\hat{p}^E_{q,k}$.

We aim to detect an onset at time k for onset model q when there is an onset in both channels, which we assume is indicated by a peak in $\mathrm{LCR}^E_{q,k}$. We thus detect an onset when for $E \in \{L, R\}$

$$\hat{p}^E_{q,k}(0) > \hat{p}^E_{q,k}(L - 1) \tag{38}$$

$$\hat{p}^{E\prime}_{q,k}(j) \triangleq \hat{\beta}^E_{q,k,1} + 2\hat{\beta}^E_{q,k,2} j + 3\hat{\beta}^E_{q,k,3} j^2 \geq \delta , \tag{39}$$

where $\delta > 0$ is a threshold on the first derivative of the polynomials evaluated at $j = \frac{L-1}{2}$.

3.2. Time Delay Estimation

We assume we detect an onset at time k for frequencies $q_r \in \{1, \ldots, Q\}$, where $r \in \{1, \ldots, R\}, 1 \leq R \leq Q$. For time delay estimation, we attempt to select the onset model of a frequency that is similarly represented in both channels. For this purpose, we select the onset model based on comparing the first derivative of the polynomial fits (cf. (39)) with

$$\hat{q} = \underset{q_r, r \in \{1,\ldots,R\}}{\operatorname{argmin}} \left| \left| \frac{\hat{p}^{L\prime}_{q,k}(j)}{\hat{p}^{R\prime}_{q,k}(j)} \right| - 1 \right| . \tag{40}$$

The phase estimate of each channel is extracted from the observation vectors $\hat{c}^{E}_{\hat{q},k}$ with the relation (26) according to

$$\hat{\phi}^{E}_{\hat{q},k} = \arctan \left(\frac{\{\hat{c}^{E}_{\hat{q},k}\}_1}{\{\hat{c}^{E}_{\hat{q},k}\}_2} \right) . \tag{41}$$

The ITD estimate $\hat{\tau}_{LR,k}$ is determined from the minimal phase shift between the local sinusoids of both ears. This is found by first finding the phase shift of $\hat{o}^{R}_{\hat{q},k}$ to the closest zero-phase point, which is $\delta_R \triangleq \min(-\hat{\phi}^{R}_{\hat{q},k}, 2\pi - \hat{\phi}^{R}_{\hat{q},k})$. Then, we select the nearest zero-phase point of $\hat{o}^{L}_{\hat{q},k}$ to that of the right ear and compute the phase shift to $\hat{o}^{L}_{\hat{q},k}$. This is $\delta_L \triangleq \operatorname{argmin}_{d \in \mathcal{D}} |\delta_R - d|$, where $\mathcal{D} = \{-\hat{\phi}^{L}_{\hat{q},k}, 2\pi - \hat{\phi}^{L}_{\hat{q},k}\}$. Finally, the ITD estimate is given by

$$\hat{\tau}_{LR,k} = \frac{\delta_R - \delta_L}{\Omega_{\hat{q}}} . \tag{42}$$

4. EXPERIMENTS

To test our algorithm, we used a dataset of speech signals with $f_s = 44.1$ kHz, varying by horizontal azimuth and reverberation of the environment. The dataset was generated by filtering the international speech test signal (ISTS) [18] (of length 60 s) with 25 anechoic HRIRs from the CIPIC database [19] with azimuths $\pm 80°, \pm 65°, \pm 55°$ and from $-45°$ to $45°$ in $5°$ steps. Furthermore, we used 18 HRIRs from the AIR database [20], of which 13 were recorded in a stairway at 1 m distance with azimuths ranging from $-90°$ to $90°$ with $15°$ steps, and 5 were recorded in the Aula Carolina Aachen (a former church) at 3 m distance with azimuths ranging from $-90°$ to $90°$ with $45°$ steps. The stairway represents an everyday environment in terms of reverberation while the Aula is highly reverberant. The HRIRs were downsampled to f_s if necessary.

For the algorithm, we used $Q = 3$ onset models with frequencies $300, 400$ and 500 Hz, and decays $\rho_q = 0.99$. The decay of (27) was $\gamma = 0.999$ and that of (35) was $\gamma = 0.9999$ with window length $L = 201$. See Fig. 2, which shows detection of an onset at 300 Hz in a CIPIC signal, for the effective window length compared to that of the onset model. The threshold for (39) was $\delta = 1\text{e}{-4}$ for the anechoic and stairway data, while for the church data, we used $\delta = 4\text{e}{-6}$. This is to account for lower peak magnitudes of the LCR due to reverberation. A comparison of how the algorithm performs with respect to reverberation is shown in Fig. 3 for a segment of the ISTS signal at $0°$ azimuth where zero delay is expected. The LCR is sparse and correctly indicates the onset positions, where plausible ITD estimates can be attained. Also plotted are the ITD estimates when no onset is detected. Without onset detection, false estimates would increase with the amount of reverberation. The estimation results for the whole dataset are shown in Fig. 4, where a clear trend between the azimuth and the ITD can be seen for the three environments.

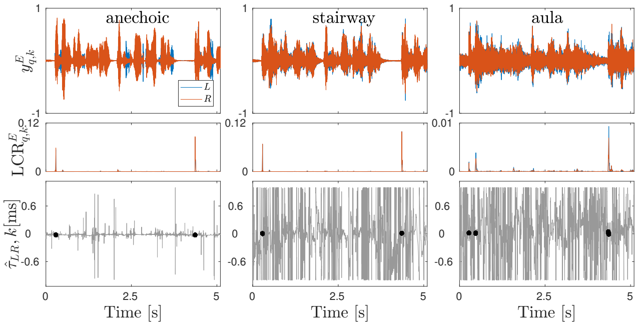

Fig. 3. Top: speech signals at azimuth $0°$. Middle: onset detection via LCRs. Bottom: ITD estimates at each time step, valid estimates at detected onsets are marked in black.

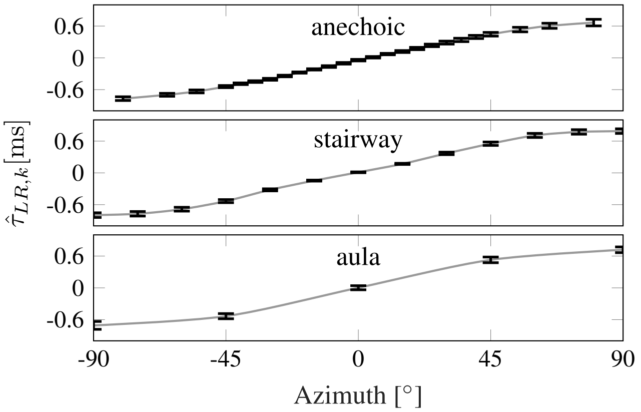

Fig. 4. Mean estimated ITD for the speech signal dataset, the horizontal error bars indicate the standard deviation.

5. CONCLUSION

This paper introduces a novel approach to real-time ITD estimation based on onset detection, which is inspired by the precedence effect of human hearing. Experiments with speech signal datasets show that the algorithm is capable of estimating the ITD even in highly reverberant environments.

6. REFERENCES

[1] B. Nordlund, “Physical factors in angular localization,” *Acta Otolaryngologica*, vol. 54, pp. 75–93, 1962.

[2] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, 2006, ch. Binaural Sound Localization, pp. 147 – 185.

[3] J. Chen, J. Benesty, and Y. Huang, “Time delay estimation in room acoustic environments: An overview,” *EURASIP, J. Adv. Signal. Process.*, vol. 6, pp. 1–19, 2006.

[4] R. F. Lyon, “A computational model of binaural localization and separation,” in *ICASSP ’83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1983, pp. 1148 – 1151.

[5] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[6] J. Chen, J. Benesty, and Y. A. Huang, “Performance of gcc- and amdf-based time-delay estimation in practical reverberant environments,” *EURASIP J. Adv. Signal. Process.*, vol. 1, pp. 25–36, 2005.

[7] B. Champagne, S. Bédard, and A. Stéphenne, “Performance of time-delay estimation in the presence of room reverberation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 2, no. 4, pp. 148 – 152, 1996.

[8] F. A. Reed, P. L. Feintuch, and N. J. Bershad, “Time delay estimation using the lms adaptive filter-static behavior,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 3, pp. 561 – 571, 1981.

[9] S. Doclo and M. Moonen, “Robust time-delay estimation in highly adverse acoustic environments,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001*, 2001, pp. 59–62.

[10] J. Benesty, “Adaptive eigenvalue decomposition algorithm for passiveacoustic source localization,” *J. Acoust. Soc. Am.*, vol. 107, no. 1, pp. 384–391, 2000.

[11] C. Faller and J. Merimaa, “Source localization in complex listening situations: Selection of binaural cues based on interaural coherence,” *J. Acoust. Soc. Am.*, vol. 116, pp. 3075–3089, 2004.

[12] H. Wallach, E. B. Newman, and M. R. Rosenzweig, “The precedence effect in sound localization,” *The American Journal of Psychology*, vol. 62, no. 3, pp. 315–336, 1949.

[13] B. Rakerd and W. M. Hartmann, “Localization of sound in rooms. iii: Onset and duration effects,” *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1695–1706, 1986.

[14] A. D. Brown, G. C. Stecker, and D. J. Tollin, “The precedence effect in sound localization,” *JARO (Journal of the Association for Research in Otolaryngology)*, vol. 16, pp. 1–28, 2015.

[15] N. Zalmai, “A state space world for detecting and estimating events and learning sparse signal decompositions,” Ph.D. dissertation, ETH Zurich, 2017, no. 24360.

[16] R. Wildhaber, N. Zalmai, M. Jacomet, and H.-A. Loeliger, “Windowed state-space filters for signal detection and separation,” *IEEE Trans. Sig. Proc.*, vol. 66, pp. 3768 – 3783, 2018.

[17] R. A. Wildhaber, E. Ren, F. Waldmann, and H.-A. Loeliger, “Signal analysis using local polynomial approximations,” in *EUSIPCO*, 2020.

[18] I. Holube, S. Fredaleke, M. Vlaming, and B. Kollmeier, “Development and analysis of an international speech test signal (ISTS),” *International Journal of Audiology*, vol. 49, pp. 891–903, 2010.

[19] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The cipic hrtf database,” in *Proc. 2001 IEEE Workshop Appl. Signal Process. Audio and Electroacoust.*, 2001, pp. 99–102.

[20] M. Jeub, M. Schäfer, and P. Vary, “A binaural room impulse response database for the evaluation of dereverberation algorithms,” in *Digital Signal Processing*, 2006, database: https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/.