Proceedings of the 5th International Symposium on Communications, Control and Signal Processing,
ISCCSP 2012, Rome, Italy, 2-4 May 2012

GEOMETRIC INFERENCE OF THE ROOM GEOMETRY UNDER TEMPERATURE VARIATIONS

Paolo Annibale[†], *Jason Filos*[§], *Patrick A. Naylor*[§], *Rudolf Rabenstein*[†]

[§] Imperial College London, Exhibition Road, London SW7 2AZ, UK
[†] University Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany

ABSTRACT

Geometric inference is an approach for localizing reflectors in a closed acoustic space. It is based on a simple observation that turns time differences of arrival (TDOA) or time of arrival (TOA) measurements from the signals of a microphone array into a geometric constraint. The reflector localization methodology relies on accurate TDOA which is directly dependent on speed of sound information. Estimating the actual speed of sound at the ambient temperature therefore greatly improves the accuracy of the reflector localization in uncontrolled environments. This manuscript shows how to use the geometric inference jointly with the speed of sound estimation for a more accurate reflector localization. Simulations and experiments show the validity of the proposed approach.

Index Terms— Geometric inference, speed of sound estimation, channel estimation, TDOA, TOA

1. INTRODUCTION

Inferring information about the environment conditions in which space-time processing algorithms operate is an emerging research topic. It has been recently considered by several authors [1, 2, 3, 4, 5, 6, 7]. Estimating the room geometry can be advantageous for applications such as wavefield rendering [1], source localization [8] and dereverberation [9]. Environmental properties, such as the ambient temperature, directly influence the propagation speed of the sound waves. Most acoustic processing algorithms assume a known propagation speed, which is a reliable assumption only under controlled laboratory conditions. Accurately estimating the speed of sound is therefore highly desired.

Geometric inference is an approach for localizing reflectors in a closed acoustic space [2, 3, 4]. It is based on a technique that turns time differences of arrival (TDOA) or time of arrival (TOA) measurements from the signals of a microphone array into geometric constraints. By exploiting such geometric constraints and using the tools of projective geometry, the line parameters of the reflectors can be inferred. As a first step, considering a sound source and a microphone array, the acoustic impulse response (AIR) is estimated for each acoustic channel. The peaks in the AIRs correspond to TOAs that are related to the direct-path propagation and the propagation due to reflections. In many practical cases the TOA of the signal corresponding to the source-receiver path is not available due to lack of synchronization. In this case the TDOAs between different receivers can be exploited to localize the sound source as long as the propagation speed is known. In fact the source localization is commonly carried out by using a standard speed of sound value to convert the TDOAs in source's Range Differences (RDs) [8, 10]. Then the estimated source range together with the propagation speed turns TDOA measurements into TOAs. As a next step the geometric constraint is formed where each reflective TOA is shown to correspond to an ellipse, with foci in source and receiver locations, and major axis equal to the length of the reflective path. Under ideal conditions, combining the ellipses related to the reflector and source-receiver pairs uniquely identifies the desired line parameters describing the reflectors.

It is evident that the reflector localization methodology relies on accurate TDOA or TOA information which is therefore dependent on accurate source localization and consequently accurate knowledge of the actual speed of sound. In fact, in uncontrolled environments, the use of the standard speed value might lead to inaccuracy in the source localization due to temperature variations [11].

This paper presents a geometric inference framework which includes the estimation of the actual speed of sound to improve the accuracy under temperature variations. The paper is structured as follows. Sec. 2 gives a description of the basis for the geometric inference, from the estimation of the acoustic channels to the conversion of TDOAs and TOAs into geometric constraints. Sec. 3 shows how to calibrate the geometric estimation to the actual propagation speed which can be estimated from the measured TDOAs. Sec. 4 is devoted to experimental results and in Sec. 5 conclusions are drawn.

2. ROOM INFERENCE

Let M sensors be distributed in a 2-D plane at positions $\mathbf{a}_i \triangleq [x_i \; y_i]^T$, $i = 0, \ldots, M-1$, and a source at $\mathbf{x} \triangleq [x_s \; y_s]^T$. Each sensor receives the output signal given by

$$x_i(t) = \mathbf{h}_i^T \mathbf{s}(t) + b_i(t), \tag{1}$$

where $\mathbf{h}_i$ is the ith channel acoustic impulse response, $\mathbf{s}(t)$ the source signal at time t and $b_i(t)$ is the additive noise at the ith output. This output is composed of the sum of the direct-path signal and scaled replicas of the source signal. The delay of each replica is determined by the respective positions of reflectors, source and receivers. Accordingly, the AIRs are given by

$$h_i(t) = \sum_{q=0}^{Q} \alpha_{i,q} \delta(t - \tau_{i,q}), \tag{2}$$

where Q is the total number of reflections of all orders, $\alpha_{i,q}$ is an attenuation term and $\tau_{i,q}$ is defined as the TOA associated with the ith sensor and the qth reflection. In vector/matrix form the model is expressed as

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{b}(t), \tag{3}$$

This work was conducted during the project SCENIC (Future and Emerging Technologies programme within the Seventh Framework Programme for Research of the European Commission, FET-Open grant number: 226007).

978-1-4673-0276-0/12/$31.00 ©2012 IEEE

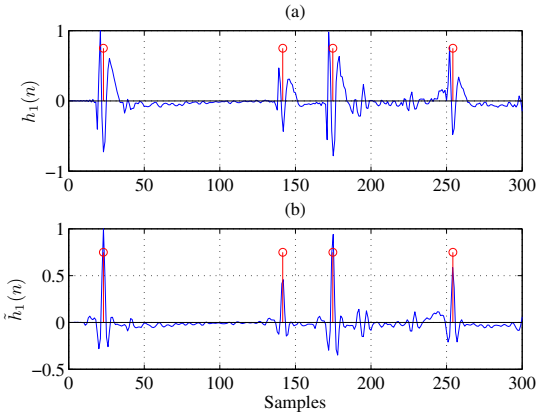

Fig. 1. Direct-path and three first-order reflections for (a) measured impulse response, (b) modified impulse response according to (6). Red '$\circ$' mark the estimated peak locations.

where

$$\begin{aligned}
\mathbf{x}(t) &= [x_0(t)\ x_1(t)\ \cdots\ x_{M-1}(t)]^T, \\
\mathbf{H} &= \begin{bmatrix} h_{0,0} & h_{0,1} & \cdots & h_{0,L-1} \\ h_{1,0} & h_{1,1} & \cdots & h_{1,L-1} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M-1,0} & h_{M-1,1} & \cdots & h_{M-1,L-1} \end{bmatrix}_{M\times L}, \\
\mathbf{s}(t) &= [s(t)\ s(t-1)\ \cdots\ s(t-L+1)]^T, \\
\mathbf{b}(t) &= [b_0(t)\ b_1(t)\ \cdots\ b_{M-1}(t)]^T,
\end{aligned}$$

and L is the length of the longest channel impulse response. For the remainder of this paper, we will not consider all reflections Q but only TOAs that are related to direct-paths and the single first-order reflection in each channel. The single first-order reflection is called the dominant reflection and is assumed to coincide with the first dominant peak after the direct-path peak.

2.1. Peak Detection from Acoustic Impulse Responses

Let $h_i(n)$ be samples of $h_i(t)$ at sample index n. Fractional delays result from path lengths that are not integer multiples of the distance propagated by sound in one sample period. Detection of impulsive events can be achieved to within one sample by considering local centres of energy with algorithms such as the sliding group delay function [12].

AIRs measured in real acoustic environments present a more challenging problem as the loudspeaker impulse response $h^{\mathrm{s}}(n)$ is convolved with the AIR $h_i(n)$. Assuming supervised identification with which estimation error can be ignored, the measured AIR is

$$\hat{h}_i(n) = h^{\mathrm{s}}(n) * h_i(n). \tag{4}$$

An example impulse response for a measured system is seen in Fig. 1 (a), showing respectively the direct-path and three first-order reflections for a single channel. The centres of each event are marked by red '$\circ$', each of which are surrounded by nearby ripples caused by $h^{\mathrm{s}}(n)$. The ripples cause ambiguity in determining the exact time corresponding to the peak and therefore a matched filter is proposed to alleviate this problem.

The length of $h^{\mathrm{s}}(n)$ is usually sufficiently short that it has decayed before the arrival of the first-order reflections [13]. Therefore, $h^{\mathrm{s}}(n)$ can be observed from the first few nonzero taps in $\hat{h}_i(n)$. Let n_i^{DP} be the propagation time of the direct-path signal from the source to microphone i and L^{s} be the approximate length of the loudspeaker impulse response. The loudspeaker impulse response can be estimated by $\hat{h}_i^{\mathrm{s}}(n) = \hat{h}_i(n + n_i^{\mathrm{DP}})w_i(n)$, where

$$w_i(n) = \begin{cases} 1 & \text{if } 0 \le n < L^{\mathrm{s}} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The filter $\hat{h}^{\mathrm{s}}(n)$ can be equalized by finding a filter $g^{\mathrm{s}}(n)$ such that $\hat{h}^{\mathrm{s}}(n)g^{\mathrm{s}}(n) \simeq \delta_a(n)$, i.e. approaching the unit impulse function $\delta_a(n)$, in a least-squares sense. Although this is an optimal solution it has been found unreliable in our experience in most practical situations. A suboptimal but more robust approach is given by the sliding correlation or matched filter [14],

$$\tilde{h}_i(n) = \sum_{j=0}^{L^{\mathrm{s}}-1} \hat{h}_i^{\mathrm{s}}(j)h_i(n+j), \tag{6}$$

that equalizes $\hat{h}_i^{\mathrm{s}}$ to a sinc function as demonstrated in Fig. 1 for a measured AIR. In (b) the mean group delay of $\hat{h}_i^{\mathrm{s}}$ has been compensated. The detected peaks are denoted by $n_{i,k}$ where i and k are the microphone and reflector index respectively. In the case of synchronized measurements, $n_{i,k} = \tau_{i,k}f_{\mathrm{s}}$.

2.2. Geometric Constraint

Under the hypothesis of specular reflection, the angle of reflection and incidence are assumed equal such that the locus of possible solutions for the reflector forms an ellipse. The focal points of the ellipses coincide with the microphone positions $\mathbf{a}_i$ and the source position $\mathbf{x}_{\mathrm{s}}$. The scaling of the minor and major axes of each ellipse are proportional to $\tau_{i,k}$. In homogenous coordinates a conic in two dimensions using parameters $\{a, b, c, d, e, f\}$ can be expressed as

$$\mathcal{C} = \left\{(x, y) \in \mathbb{R}^2 | ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0\right\}. \tag{7}$$

or by setting $\mathbf{v} = [x\ y\ 1]^T$ and $\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}$ as $\mathbf{v}^T\mathbf{C}\mathbf{v} = 0$, which parameterises an ellipse after constraining

$$\det(\mathbf{C}) \neq 0, \quad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \quad \det(\mathbf{C})/(a+c) < 0. \tag{8}$$

The ellipse associated with the ith microphone and the dominant reflector is defined as $\mathbf{C}_i$. The estimation of the unknown parameters of $\mathbf{C}_i$ is outlined in detail in [3]. Furthermore a line in homogeneous coordinates can be expressed as

$$\mathcal{L} = \left\{(x, y) \in \mathbb{R}^2 | l_1x + l_2y + l_3 = 0\right\}, \tag{9}$$

which after setting $\mathbf{l} = [l_1\ l_2\ l_3]^T$ can be written as $\mathbf{l}^T\mathbf{v} = 0$. If we group together M ellipses that are associated with a particular reflector, then the line parameters of this reflector can be estimated using the following cost function:

$$J\left(\mathbf{l}, \{\mathbf{C}_i^*\}_{i=0}^{M-1}\right) = \sum_{i=0}^{M-1} \left\|\mathbf{l}^T\ \mathbf{C}_i^*\ \mathbf{l}\right\|^2, \tag{10}$$

where $M \ge 3$ and $\mathbf{C}_i^* = \det(\mathbf{C}_i)\mathbf{C}_i^{-1}$ is the adjoint of the conic-matrix $\mathbf{C}_i$.

3. GEOMETRIC INFERENCE UNDER TEMPERATURE VARIATIONS

Under temperature variations a standard value for the speed of sound might yield an inaccurate estimation of $\tau_{i,k}$ impairing the reflector localization, in this case an estimate of the actual speed of sound is necessary. The authors proposed in [15] a novel method to estimate the speed of sound. Such a method relies merely on TDOA measurements and therefore it is suited for geometric inference when no synchronization between source and receiver is available.

3.1. Speed of Sound Estimation from TDOAs

Consider the two-dimensional inference problem of the previous section, i.e. an acoustic source lies in an unknown position $\mathbf{x}_\mathrm{s}$ and M sensors are distributed at the known positions $\boldsymbol{a}_i$ with $i = 0, \cdots, N$ and $N = M - 1$. From the M estimated AIRs a *spherical* set of N TDOAs can be obtained as time differences between the direct-path peaks and the direct-path peak of the reference microphone. If the first microphone $\boldsymbol{a}_0$ is chosen as reference such a set may be represented by the following N-vector

$$\boldsymbol{\tau} = \frac{1}{f_s} \begin{bmatrix} n_{1,0} - n_{0,0} \\ n_{2,0} - n_{0,0} \\ \cdots \\ n_{N,0} - n_{0,0} \end{bmatrix} . \tag{11}$$

According to [15] the following scalar function of the assumed signal propagation speed c can be written

$$\delta(c) = ||\boldsymbol{\Gamma}\boldsymbol{b}(c)|| - \frac{1}{c}\boldsymbol{\Theta}\boldsymbol{b}(c) , \tag{12}$$

where the constant matrices $\boldsymbol{\Theta}$, $\boldsymbol{\Gamma}$ and the N-vector $\boldsymbol{b}(c)$ depend only on the microphone positions $\boldsymbol{a}_i$, $, i = 1, \cdots, N$ and the vector $\boldsymbol{\tau}$. The zero of the above function is an estimate of the actual propagation speed, in this case the actual speed of sound.

Unfortunately such a function $\delta(c)$ involving the Euclidean norm of $\boldsymbol{\Gamma}\boldsymbol{b}(c)$ is nonlinear, therefore applying a root-finding algorithm might be intractable. This issue can be overcome by linearizing $\delta(c)$ near to its zero-crossing as it has been shown to be approximately linear in that range. In acoustic applications, the standard value of speed of sound at 20 °C may be chosen as a reliable linearization point $\bar{c}$, leading to the following first order Taylor expansion

$$\delta(c) \approx \delta_\mathrm{lin}(c) = \bar{\delta} + \bar{\delta}'(c - \bar{c}) , \tag{13}$$

with

$$\bar{\delta} = \delta(\bar{c}) \quad \text{and} \quad \bar{\delta}' = \left.\frac{\mathrm{d}\delta(c)}{\mathrm{d}c}\right|_{c=\bar{c}} . \tag{14}$$

Finally, the estimated propagation speed value is given by the zero-crossing point $\hat{c}$ of the linearized function $\delta_\mathrm{lin}(c)$, i.e.

$$\hat{c} = \frac{\bar{\delta} - \bar{\delta}'\bar{c}}{\bar{\delta}'}, \tag{15}$$

where the value of the first order derivative $\bar{\delta}'$ at $\bar{c}$ can be calculated with derivation rules from (12).

3.2. Multiple Sources Approach

As reported in [15], the above speed estimate can be improved in noisy conditions by exploiting the *full* TDOA set. However here the TDOAs are not obtained by signal correlation, rather they are extracted from the M estimated AIRs. As a consequence the construction of the *full* TDOA set will not add any useful information for the speed estimation. Moreover the microphone array used for the geometric inference experiments (see Sec. 4) uses only $M = 5$ microphones, i.e. only one microphone provides redundant information to counteract the effect of the measurement noise since the minimum number of sensors required for the two dimensional speed estimation problem is 4. Nevertheless the robustness of the algorithm can be still improved by assuming that for a small-size array the speed of sound in a reasonable time interval is the same regardless of the source position. Following this idea the scalar function $\delta(c)$ in (12) can be built using TDOA sets generated by different acoustic sources. For a rectangular room, four differently located acoustic sources are used (see Fig. 2). Hence for each source $\mathbf{x}_j$ ($j = 1, 2, 3, 4$) a function $\delta_j(c)$ is derived. In noisy conditions a robust speed of sound estimate can be found from the minimization in the least-squares sense of such functions. The corresponding cost function is given by

$$\sum_{j=1}^{4} \delta_j(c)^2 = \sum_{j=1}^{4} \left(||\boldsymbol{\Gamma}_j\boldsymbol{b}_j(c)|| - \frac{1}{c}\boldsymbol{\Theta}_j\boldsymbol{b}_j(c) \right)^2 , \tag{16}$$

where the index j indicates that the matrices $\boldsymbol{\Theta}$, $\boldsymbol{\Gamma}$ and the vector $\boldsymbol{b}(c)$ are obtained using the vector $\boldsymbol{\tau}_j$ corresponding to the source $\mathbf{x}_j$. Again the linear approximation described in Sec. 3.1 can be applied to perform the minimization efficiently.

The resulting speed of sound estimate can be now used to accurately estimate the TOAs $\tau_{i,k}$ by means of source localization algorithms and perform the reflector localization according to Sec. 2.2.

4. EXPERIMENTAL RESULTS

The effects of temperature variation on the speed of sound within the reflector localization framework have been evaluated in a real conference room measuring $3.31 \times 3.58 \times 3.00$ m, with concrete walls and two flush-mounted wooden doors in the south and east walls. A microphone array consisting of four microphones spaced by 0.5 m in a '+' configuration and a fifth placed in the centre was positioned at (1.75,1.5) m from the south-west corner. A Genelec 8030A loudspeaker was placed at four distinct source positions. The microphone signals were sampled at 96 kHz. At each position, the acoustic impulse response between the source and microphone array was estimated using the MLS method [16].

At first, geometric inference has been performed without knowledge of the true speed of sound. As usual a value for the speed of sound has to be assumed to convert estimated time differences into range differences. Here a value of $c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$ has been adopted corresponding to a temperature of $\vartheta = 72\ ^\circ\mathrm{C}$. This rather high value has been chosen to demonstrate the effect of an erroneous assumption.

Next the same set of TDOAs has been used for geometric inference, but this time the speed of sound at the time of measurement has been infered from these TDOAs. The resulting speed of sound was $\hat{c} = 345\ \frac{\mathrm{m}}{\mathrm{s}}$ corresponding to a temperature of $\vartheta = 23\ ^\circ\mathrm{C}$.

The results in Fig. 2 and Table 1 show not only the increase of the distance error ϵ_d and the angular error ϵ_a with an erroneously assumed temperature resp. speed of sound. Fig. 2 displays also clearly the effect of a temperature increase on the inference of the wall positions: A higher value of the speed of sound virtually increases the size of the room by upscaling all involved distances.

A variation of the room temperature by $\Delta\vartheta \approx 50K$ may seem excessive except in spaces exposed to seasonal variations of the am-

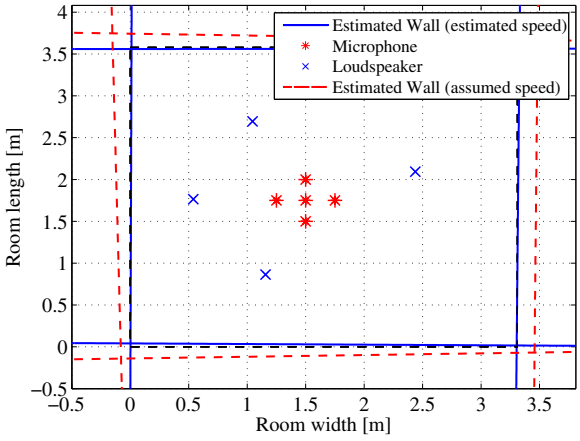

Fig. 2. Room inference results using a microphone array, placed centrally in a small conference room, capturing a MLS sequence from 4 source positions in turn. Red lines: assumed speed of sound $c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$. Blue lines: estimated speed of sound $c = 345\ \frac{\mathrm{m}}{\mathrm{s}}$. Dashed black rectangle: actual geometry of the room.

Table 1. Reflector Localization Results with Real-World Data

$c = 345\ \frac{\mathrm{m}}{\mathrm{s}}$

Wall	ϵ_d [cm]	ϵ_a [°]
North	1.890	0.056
East	0.770	0.388
South	5.220	0.429
West	0.650	0.138

$c = 375\ \frac{\mathrm{m}}{\mathrm{s}}$

Wall	ϵ_d [cm]	ϵ_a [°]
North	19.440	1.253
East	16.150	0.400
South	16.910	1.168
West	11.980	1.101

bient temperature. However the room available for the experiment in Fig. 2 is rather small. Larger rooms with a larger travel time of the reflections are more sensitive to speed of sound variations and exhibit the same absolute errors at smaller temperature variations.

5. CONCLUSIONS

A technique for inferring the geometry of a room under temperature variations has been presented. Since existing methods rely on either synchronized measurements or often inaccurate estimates of the speed of sound, the propagation speed is estimated before the dominant reflectors are localized. In this way the room geometry can be reconstructed even when there are fluctuations in the ambient temperature. Improvements in accuracy are demonstrated in a real conference room that is exposed to strong temperature variations.

6. REFERENCES

[1] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA '09*, 2009.

[2] F Antonacci, A Sarti, and S Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, Tx, Mar. 2010, pp. 2822–2825.

[3] J. Filos, E. A. P. Habets, and P. A. Naylor, "A two-step approach to blindly infer room geometries," in *Proc. of Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Sept. 2010.

[4] J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti, and P. A. Naylor, "Robust inference of room geometry from acoustic measurements using the Hough transform," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011, pp. 161–165.

[5] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "A methodology for evaluating the accuracy of wave field rendering techniques," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 69–72.

[6] E. Mabande, Haohai Sun, K. Kowalczyk, and W. Kellermann, "On 2D localization of reflectors using robust beamforming techniques," in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Prag, Czech Republic, May 2011, pp. 153–156.

[7] I. Dokmanic, Y.M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-D polygonal case," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 321 –324.

[8] Yiteng Huang, J. Benesty, G. W. Elko, and R. M. Mersereati, "Real-time passive source localization: a practical linear-correction least-squares approach," *Trans. on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, August 2002.

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

[10] P. Stoica and Jian Li, "Lecture notes - source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, November 2006.

[11] P. Annibale and R. Rabenstein, "Accuracy of time-difference-of-arrival based source localization algorithms under temperature variations," in *Proc. of Int. Symposium on Communications, Control and Signal Processing, (ISCCSP), Limassol, Cyprus*. IEEE, 2010.

[12] Mike Brookes, Patrick A. Naylor, and Jon Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. on Speech Audio Process.*, vol. 14, 2006.

[13] L. R. Fincham, "Refinements in the impulse testing of loudspeakers," *Journal Audio Eng. Soc.*, vol. 33, no. 3, pp. 133–140, Mar. 1985.

[14] G. Turin, "An introduction to matched filters," *IRE Trans. on Information Theory*, vol. 6, no. 3, pp. 311–329, June 1960.

[15] P. Annibale and R. Rabenstein, "Acoustic source localization and speed estimation based on time-differences-of-arrival under temperature variations," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, August 2010.

[16] J. Vanderkooy, "Aspects of MLS measuring systems," *Journal Audio Eng. Soc.*, vol. 42, pp. 219–231, 1994.