On Loudspeakers as Recording Devices

Roest, Tobias; Martinez, Jorge; Oosterom, Han; Hendriks, Richard

Publication date
2022

Document Version
Final published version

Published in
AES Europe Spring 2022 - 152nd Audio Engineering Society Convention 2022

Citation (APA)
Roest, T., Martinez, J., Oosterom, H., & Hendriks, R. (2022). On Loudspeakers as Recording Devices. In *AES Europe Spring 2022 - 152nd Audio Engineering Society Convention 2022* (pp. 361-369). Article 10586 Audio Engineering Society. https://www.aes.org/e-lib/browse.cfm?elib=21699

Audio Engineering Society

Convention Paper 10586

Presented at the 152nd Convention
2022 May, In-Person and Online

On Loudspeakers as Recording Devices

Tobias Roest[1], Jorge Martinez[1], Han Oosterom[2], and Richard C. Hendriks[1]

[1] *TU Delft*
[2] *Bosch Security Systems B.V.*

Correspondence should be addressed to Tobias Roest (tobiasroest@gmail.com)

ABSTRACT

Moving-coil electrodynamic loudspeakers and dynamic microphones use the same linear actuator technology at the core of their operation. Utilising this similarity, loudspeakers have a possible use as recording devices in cases where using dedicated microphones is not feasible. Such a use case exists in public address and voice alarm systems. This paper evaluates the feasibility of using the loudspeakers already in place in these systems as recording devices to provide information back to the system. A system using a single loudspeaker as both a playback and recording device simultaneously is analysed, modelled and simulated. The results show that using a current measuring set-up with an analogue-to-digital converter capable of detecting a range of roughly 120 dB, a speech signal incident at 46 dBSPL in a cone of 150° from a loudspeaker can be successfully estimated in an office room with an announcement playing at 88 dBSPL and background interference present at the same time. As the estimated signal is unknown to the system, the solution generalises to other signal types as well.

1 Introduction

Moving-coil electrodynamic loudspeakers and dynamic microphones have existed side-by-side for decades, while having the same linear actuator technology at the core of their operation. This paper provides a novel evaluation of the use of a loudspeaker as a microphone, to bridge the gap between these devices and possibly find a new use for loudspeakers. With microphones being cheap and easy to implement nowadays, only specific use cases would benefit from using a loudspeaker as a microphone. Such a use case exists in public address (PA) and voice alarm systems, which are integrated in buildings during construction. For example, recording capabilities in such systems could aid rescue workers when evacuating buildings by listening for people calling for help. The ambient noise level in an area could also be estimated from a recording, which can then be used to adjust the playback level of announcements for better intelligibility. However, adding a network of microphones to such a system is a costly affair, and is thus often not done. If the loudspeakers already in place could be used for recording, the system could be significantly improved with minimal adjustments. To effectively evaluate the use of loudspeakers for simultaneous playback and recording, a single loudspeaker case is considered in this paper.

Section 2 discusses the considerations for the electronic system, and provides three possible circuit topologies for implementation. Section 3 discusses the signal model of a loudspeaker playing and recording in a room. Section 4 considers the identification

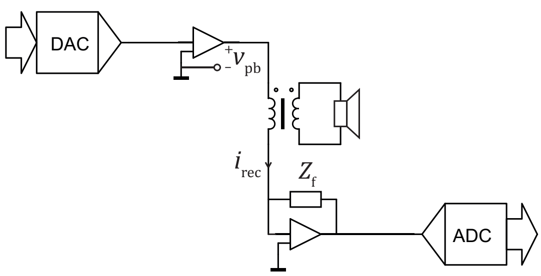


Fig. 1: A measurement setup based on a transimpedance amplifier to measure the current. Feedback impedance Z_f needs to be matched to the frequency dependent loudspeaker impedance.

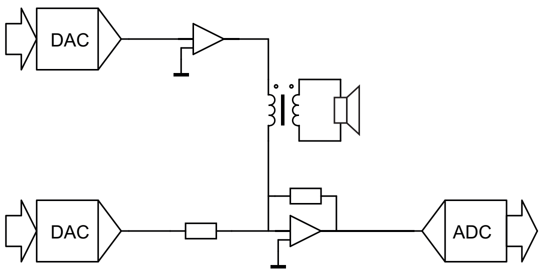


Fig. 2: A measurement setup which uses a secondary amplifier channel to reduce the effect of the frequency dependent loudspeaker impedance on the current measurement.

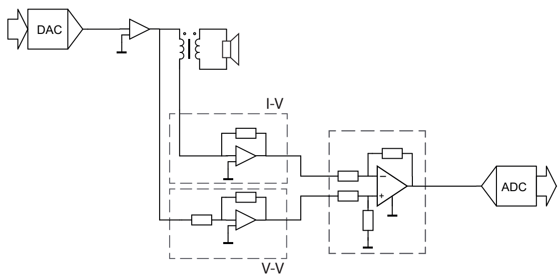


Fig. 3: A measurement setup which uses a differential amplifier to reduce the dynamic range between the recording signal and the digital-to-analogue converter output. The playback signal voltage is placed on the common mode of a differential amplifier, where it is rejected by the differential nature of the system. The current measurement will also contain the recording signal, which is amplified by the differential amplifier.

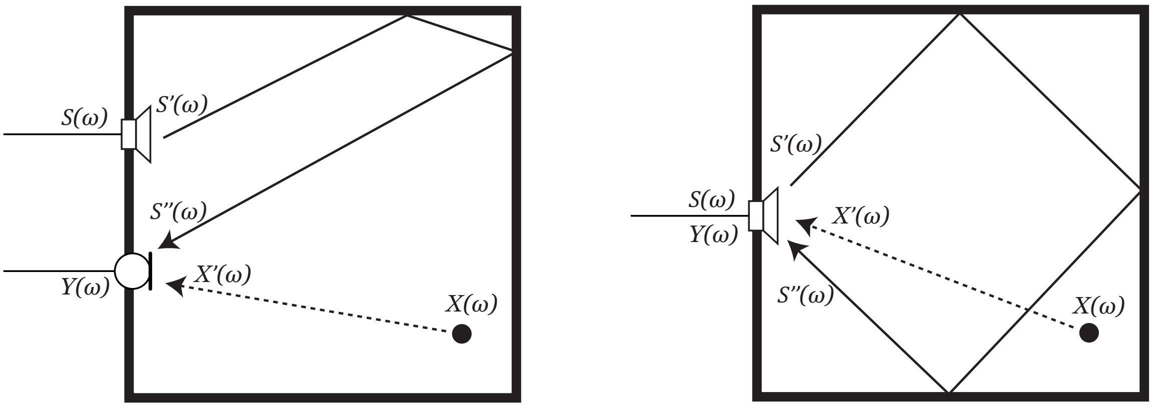
of loudspeaker recording characteristics through measurements. The results presented in sections 3 and 4 are then used in Section 5, where a simulation is discussed as a proof-of-concept for a system that can play and record audio simultaneously using a single loudspeaker. The results from this are presented in Section 6. Section 7 concludes the paper with a statement of feasibility, and Section 8 closes with suggestions for continuation of the research.

2 Electronics for recording during playback

Let us evaluate a system using digital communication between devices, as this is the case in many PA systems [1]. At the amplifier, the digital signal is sent to a digital-to-analogue converter and an amplifier. Assuming a system uses a voltage amplifier to drive the loudspeaker [2], the measurement for the recording needs to be the current i_{rec} through the loudspeaker. Measuring the voltage across the output of the amplifier will only result in a reading of the amplifier output v_{pb} without any recorded signal. Using a transimpedance amplifier design, the current can be measured at the amplifier side. Such a design is shown in Fig. 1.

As the loudspeaker impedance is significantly frequency dependent, this will impact the measurement (discussed more in-depth in Section 4.1), and this needs to be compensated for. This could be done by using another amplifier channel as shown in Fig. 2, where a digital feedback loop can be used to calibrate the system during the recording of a known signal.

The dynamic range of the recorded signal can be quite large, due to the low output of the loudspeaker as a microphone compared to the playback signal. The dynamic range for a noise floor of 46 dBSPL, derived from the incident SPL in the simulation that will be shown in Section 5, is between 105 – 120 dB depending on loudspeaker model and playback level for the measured devices. Selecting an analogue to digital converter capable of this dynamic range is essential. As devices with these capabilities do exist [3], recording during playback is deemed possible.



(a) The conventional case, where the loudspeaker and microphone are separate devices

(b) The case of interest, where the loudspeaker and the microphone are one device

Fig. 4: Schematic drawings of the cases. In these figures, only one path per source is drawn for clarity, but the acoustic transfer function in the models is assumed to contain all paths. The playback signal is represented by different alterations of $S(\omega)$. $X(\omega)$ represents the signal from the point source. $Y(\omega)$ represents the signal output from the recording device.

To reduce the dynamic range, a differential amplifier configuration, as shown in Fig. 3, is a possible solution. Using the current measurement from the transimpedance amplifier on one input, and taking the amplifier voltage as the other input, the amplifier output voltage v_{pb} will be a common-mode input for the differential amplifier. Combining this with an amplifier with a high common-mode rejection ratio can significantly improve the recorded signal quality.

Evaluation of a physical system will show additional noise sources introduced by the system, such as e.g. device self-noise and quantisation noise. Additional measures to reduce the dynamic range of the recording signal or cancel noise from system components could be needed.

3 Signal Model

Let us first describe the signal model for a conventional setup of a separate loudspeaker and microphone in a room as shown in Fig. 1a. For convenience we define all notation in the discrete time frequency domain. Taking N equally spaced samples on the frequency range $[-Fs/2, F_s/2]$, with F_s the sampling frequency, we can describe the frequency bin ω_k as

$$\omega_k = \frac{2\pi F_s}{N}k, \quad k = \begin{cases} [-\frac{N}{2}, \frac{N}{2}-1](N \text{ even}), \\ [-\frac{N-1}{2}, \frac{N-1}{2}](N \text{ odd}), \end{cases} \tag{1}$$

where the bin index k will be omitted for brevity in this paper. Let $S(\omega)$ denote a signal that will be played by the loudspeaker, and let $X(\omega)$ denote a point source in the room. Let the loudspeaker transfer function be $H_{\mathrm{l}}(\omega)$, the microphone transfer $H_{\mathrm{m}}(\omega, \theta_i)$, and let the acoustic transfer function (ATF) from source location $\mathbf{x_s}$ to receiver location $\mathbf{x_r}$ be $H_{\mathrm{r}}(\omega, \mathbf{x_s}, \mathbf{x_r})$. Here, θ_i represents the incident angle of the sound. Adding uncorrelated device self-noise sources $N_{\mathrm{l}}(\omega)$ and $N_{\mathrm{m}}(\omega)$ and an uncorrelated noise source in the room $N_{\mathrm{r}}(\omega)$ to the model, we write the microphone measurement $Y(\omega)$ as

$$\begin{aligned} Y(\omega) =& H_{\mathrm{m}}(\omega, \theta_{\mathrm{s}})H_{\mathrm{r}}(\omega, \mathbf{x}_{\mathrm{l}}, \mathrm{x}_{\mathrm{m}})(H_{\mathrm{l}}(\omega)S(\omega) + N_{\mathrm{l}}(\omega)) \\ &+ H_{\mathrm{m}}(\omega, \theta_{\mathrm{x}})(H_{\mathrm{r}}(\omega, \mathbf{x}_{\mathrm{x}}, \mathbf{x}_{\mathrm{m}})X(\omega)) \\ &+ H_{\mathrm{m}}(\omega, \theta_{\mathrm{n}})(H_{\mathrm{r}}(\omega, \mathbf{x}_{\mathrm{n}}, \mathbf{x}_{\mathrm{m}})N_{\mathrm{r}}(\omega)) + N_{\mathrm{m}}(\omega), \end{aligned} \tag{2}$$

where subscript $_{\mathrm{l}}$ indicates the loudspeaker, subscript $_{\mathrm{m}}$ indicates the microphone, subscript $_n$ indicates the noise source, and subscript $_{\mathrm{r}}$ indicates the room. We can adapt this model to the case where the loudspeaker is also used as a microphone. To do so, we change the position of the receiver in the ATF to the loudspeaker position and use the loudspeaker recording response for $H_{\mathrm{m}}(\omega, \theta_s)$. The signal model for this case, visualised in Fig. 4b, is

$$\begin{aligned}Y(\omega) = &H_\mathrm{m}(\omega,\theta_s)H'_\mathrm{r}(\omega,\mathbf{x}_\mathrm{l},\mathbf{x}_\mathrm{l})(H_\mathrm{l}(\omega)S(\omega)+N_\mathrm{l}(\omega))\\ &+H_\mathrm{m}(\omega,\theta_\mathrm{x})(H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{x},\mathbf{x}_\mathrm{l})X(\omega))\\ &+H_\mathrm{m}(\omega,\theta_\mathrm{n})(H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{n},\mathbf{x}_\mathrm{l})N_\mathrm{r}(\omega))+N_\mathrm{m}(\omega).\end{aligned} \tag{3}$$

Notice here that we use H'_r instead of the ATF H_r as will be explained below. The ATF H_r can be modelled approximately by using the mirror image source method (MISM) described in [4]. The MISM is a simulation method based on geometrical acoustics [5], and is valid for a small rectangular room. It can be assumed to be fairly correct for broadband signals [6], besides when the loudspeaker and microphone are colocated as in Fig. 4b. This can lead to problems as the distance equals zero, leading to undefined values when calculating the ATF. Let H'_r therefore denote a modified version of the ATF calculation where the elements with a source-receiver distance of zero are removed from the summation, that is,

$$H'_\mathrm{r}(\omega,\mathbf{x_s},\mathbf{x_r}) = \begin{cases}\displaystyle\sum_{p=1}^{8}\sum_{\mathbf{r}=-\infty}^{\infty}\frac{\exp(j\frac{\omega}{c}\mathbf{d})}{4\pi\mathbf{d}}\exp(-j\omega t), & \\ & \text{for } \mathbf{d}\neq 0,\\ 0, & \text{for } \mathbf{d}=0,\end{cases} \tag{4}$$

with $\mathbf{d} = |\mathbf{R_p}+\mathbf{R_r}|$. Here, t represents the time, c represents the speed of sound, and $j=\sqrt{-1}$. $\mathbf{R_p}$ represents the eight vectors given by the eight permutations over $\pm$ of

$$\mathbf{R_p} = (x_s \pm x_r, y_s \pm y_r, z_s \pm z_r), \tag{5}$$

$\mathbf{r}$ is the integer vector triplet (n,l,m), and

$$\mathbf{R_r} = 2(nL_x, lL_y, mL_z), \tag{6}$$

where (L_x, L_y, L_z) are the room dimensions [4].

4 Loudspeaker measurements

To assess the performance of loudspeakers as recording devices, the recording transfer function of the loudspeaker needs to be measured. A measurement was set up in an anechoic chamber at the acoustic lab of Bosch Security Systems B.V. in Eindhoven. Fig. 5 shows the set up. A loudspeaker plays a frequency sweep, and the device under test (DUT), another loudspeaker mounted in a ceiling panel inside a fire dome records the signal. A reference microphone is used to calibrate the measurement. As the reference microphone is placed in front of the DUT, some compensation is done on the sensitivity measurements. The DUTs used for these measurements consist of the LC1 ceiling loudspeaker range from Bosch Security Systems B.V. [7]. In the set-up shown, three characteristics were measured:

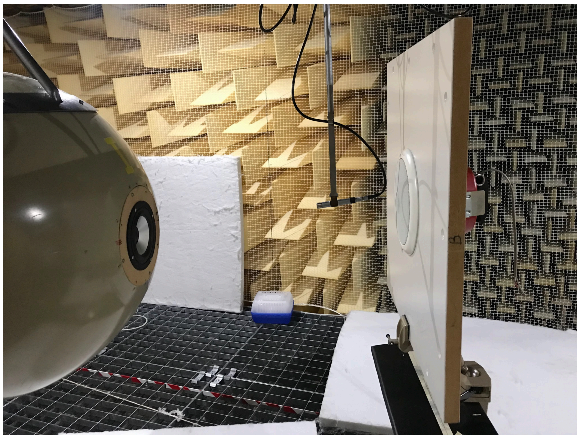


Fig. 5: A close-up of the setup used for measuring the recording response of loudspeakers.

1. The frequency response along the principal axis
2. The directional response
3. The output sensitivity.

4.1 Frequency response along the principal axis

Fig. 6 shows two of the measured loudspeaker recording frequency responses, compared to the response when used as a loudspeaker. These are the recording responses from two devices of the same type, the LC1-WM06E8 ceiling loudspeaker. Other types of speakers from this range showed slight differences in the responses, but the overall shape is very similar. Comparing these recording responses, two main trends can be seen:

- Between roughly 150 Hz and 750 Hz a significant increase in the response is seen compared to the higher frequencies. This is most probably the result of the DUTs' resonant frequency.

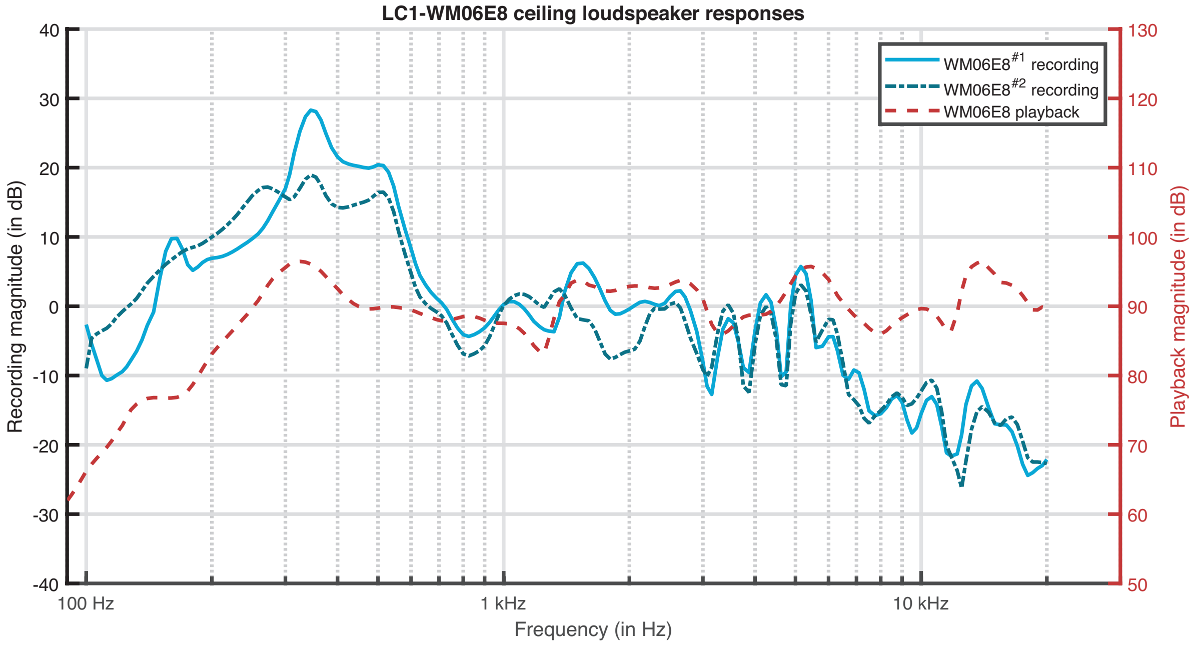


Fig. 6: The on-axis measured recording response of the WM06E8 loudspeaker, compared with the playback response. The blue solid and dash-dotted curves represent the recording response of two different loudspeakers. The red dashed curve is the playback response, as given by the loudspeaker datasheet.

The resonance of microphones is usually significantly damped. Using the fire dome to mount the DUT means it has a small enclosure, which causes the resonant frequency of the system to be higher than the free air resonant frequency of the driver [1]. The effect of the resonant frequency is clearly seen here because the measurement system used to identify the recording transfer has a high-impedance input [8]. Using a load resistor with a much smaller, frequency dependent, resistance R_L than the DC resistance of the voice coil R_0 ($R_L \ll R_0$) can significantly reduce the effect of the resonance in the recording measurement. However, this is not a suitable solution when also using the loudspeaker for playback.

- From 5kHz upwards, the recording response starts to drop off significantly. This is most likely due to cone breakup effect at these higher frequencies, where the cone stops vibrating like a piston [8]. High frequency compensation methods in the construction of the loudspeaker could help to compensate for this effect.

4.2 Directional recording response

In Fig. 7 the polar plots for recording at 1 kHz are shown for the same two devices as before. The measurements were done at intervals of 15° in one quadrant and mirrored along the principal (0°) axis. The polar plot shows a large peak at 45°. Looking at the loudspeaker construction, the peak is probably introduced by the whizzer cones attached to the loudspeaker in this model, which sit at a 45° angle. Across all measured loudspeakers and angles, the lowest output sensitivity seen was 18 mV/Pa. This is still within a usable range compared to the on-axis sensitivity.

4.3 Output sensitivity

The on-axis recording output sensitivity for all the measured loudspeakers is given in Table 1. This output sensitivity was measured at the transformer terminals. The results are also referenced to the speaker terminals. These results are comparable to the ratings of microphones (2-50 mV/Pa depending on the intended source) [9]. One should take into account, however,

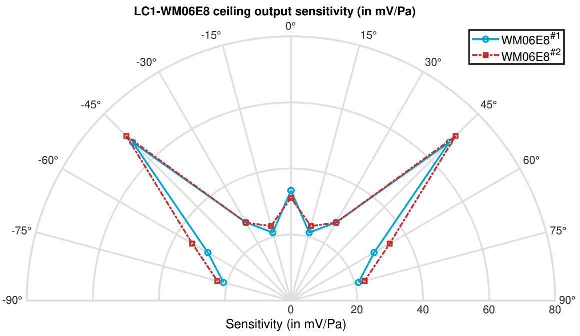


Fig. 7: Polar sensitivity plot of 2 loudspeakers as microphones as seen at the transformer connection. The plot was generated using data from one quadrant, mirrored on the 0° axis.

Table 1: Recording output sensitivity of DUTs

Loudspeaker model / device	Sensitivity at transformer (in mV/Pa)	Sensitivity at loudspeaker (in mV/Pa)
WM06E8 #1	33.34	1.77
WM06E8 #2	31.13	1.65
WC06E8	36.68	1.95
UM06E8 #1	25.02	1.33
UM06E8 #2	25.11	1.33
UM12E8	31.33	1.66
UM24E8	30.71	1.63

that a loudspeaker is not designed nor made to have a specific or optimised recording sensitivity.

5 Simulation

To evaluate the use of loudspeakers as recording devices, a comparison with an ideal case is useful. To this end, a simulation is set up, based on the theoretical signal model from Section 3 and the measurements from Section 4. Using the room impulse response generator described in [10], a room is modelled with dimensions of 5 m x 4 m x 3 m. The recording device (a loudspeaker or ideal microphone) is placed in the ceiling of this room at $\mathbf{x}_\mathrm{l} = (4,2,3)$. A speech signal from the TIMIT database [11] is used as output signal from a point source placed at location $\mathbf{x}_\mathrm{x} = (3,2,1.8)$. Lastly, a point source acting as interference is placed at the edge of the room at $\mathbf{x}_\mathrm{n} = (0,1,1)$.

Case 1: Recording without playback

This section discusses a simplified simulation where the loudspeaker does not play a signal, as to compare the recording capabilities of a loudspeaker with an ideal case. The sound pressure levels are set to be 50 dBSPL for the speech signal at location $\mathbf{x_x}$ and 40 dBSPL for the noise signal at location $\mathbf{x_n}$. The best case recording is taken as

$$\begin{aligned} Y_\mathrm{m}(\omega) =& H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{x},\mathbf{x}_\mathrm{m})X(\omega) \\ &+ H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{n},\mathbf{x}_\mathrm{m})N_\mathrm{r}(\omega). \end{aligned} \tag{7}$$

For the loudspeaker recording, a linear phase filter is derived from the on-axis recording transfer presented in Section 4. The signal recorded with the loudspeaker is then given by

$$\begin{aligned} Y_\mathrm{l}(\omega) =& H_\mathrm{m}(\omega)H_\mathrm{r}(\omega,\mathbf{x_x},\mathbf{x_l})X(\omega) \\ &+ H_\mathrm{m}(\omega)H_\mathrm{r}(\omega,\mathbf{x_n},\mathbf{x_l})N_\mathrm{r}(\omega). \end{aligned} \tag{8}$$

Case 2: Recording during playback

In the second case, a signal is also played through the loudspeaker. This is implemented by using another excerpt from the TIMIT database. The loudspeaker is simulated to play at a power of 1W, which gives an SPL of 88 dB. On the loudspeaker cable, this signal is also present. While playing at 1W, the recorded point source signal level is -105.6 dB compared to the playback signal level on the cable. When a separate microphone is used for the recording, the signal model is described as

$$\begin{aligned} Y_\mathrm{m}(\omega) =& H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{l},\mathbf{x}_\mathrm{m})H_l(\omega)S(\omega) \\ &+ H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{x},\mathbf{x}_\mathrm{m})X(\omega) \\ &+ H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{n},\mathbf{x}_\mathrm{m})N_\mathrm{r}(\omega). \end{aligned} \tag{9}$$

For the loudspeaker recording, the signal model is described as

$$\begin{aligned} Y_\mathrm{l}(\omega) =& S(\omega) \\ &+ H_\mathrm{r}'(\omega,\mathbf{x}_\mathrm{l},\mathbf{x}_\mathrm{l})H_\mathrm{l}(\omega)S(\omega) \\ &+ H_\mathrm{m}(\omega)H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{x},\mathbf{x}_\mathrm{l})X(\omega) \\ &+ H_\mathrm{m}(\omega)H_\mathrm{r}(\omega,\mathbf{x}_\mathrm{n},\mathbf{x}_\mathrm{l})N_\mathrm{r}(\omega). \end{aligned} \tag{10}$$

Table 2: Intelligibility of the simulated recording when only recording

	SIIB	STOI	Intelligibility
y_{m} (recorded with a microphone)	355	0.89	$\approx 100\%$
y_{l} (recorded with a loudspeaker)	239	0.84	$\approx 100\%$

Where $H_{\mathrm{r}}'(\omega, \mathbf{x}_{\mathrm{l}}, \mathbf{x}_{\mathrm{l}})$ is the modified version of the ATF described by (4).

The unwanted component on the loudspeaker cable is removed using a least mean square (LMS) filter [12]. To simulate the case where (an estimation of) the acoustic transfer function is known, the announcement signal as incident at the recording device is also removed using the LMS filter. This gives a best case scenario for the estimation of a point source signal.

Signal evaluation

The evaluation of the signals is done by determining the predicted intelligibility using two different instrumental measures. For this, speech intelligibility in bits (SIIB) [13] and the short-time objective intelligibility measure (STOI) [14] are used. These measures have relatively high correlation to actual intelligibility [15], and will thus allow for quick interpretation of the results. Both these measures are non-linear, and are asymptotic to 100% intelligibility. SIIB typically produces a result of 0-150 bits, where 0 bits corresponds to zero intelligibility, and more than 150 bits corresponds to high intelligibility. The range of results for STOI is typically a scalar ranging from 0.2-0.9, where 0.2 corresponds to zero intelligibility, and more than 0.8 corresponds to high intelligibility.

6 Results

The results from the simulation of Case 1, only recording, are shown in Table 2. While some information is lost by using a loudspeaker instead of an ideal microphone, this has a minimal impact on the intelligibility of the signal, as the measures are asymptotic to 100% in this range. This means that using a loudspeaker for this purpose is indeed feasible.

Table 3: Intelligibility of the simulated recording during playback

	SIIB	STOI	Intelligibility
y_{l} (full recording)	8	0.15	$\approx 0\%$
y_{m} (microphone recording)	10	0.09	$\approx 0\%$
y_{l} (line signal removed)	9	0.18	$\approx 0\%$
y_{m} (room signal removed)	307	0.80	$\approx 100\%$
y_{l} (line and room signal removed)	80	0.54	$\approx 60\%$

Table 3 shows the results of the simulation of Case 2, recording during playback. Here the impact of the playback signal is evident. The full loudspeaker recording is dominated by the playback signal on the line, and hardly any information is present. When using a separate microphone, or removing just the playback signal from the loudspeaker recording with an LMS filter, still no intelligible signal of the point source is present in the result. This can be attributed to the loud playback level of the loudspeaker compared to the point source signal. After removing the playback signal as it is simulated to be in the room as well, a significant increase in intelligibility is seen with both recording devices. Here, the difference between loudspeaker and microphone becomes more clear. However, at 60% intelligibility the loudspeaker is still deemed to be a feasible recording device.

The LMS filter in the simulation only removes known, unwanted components from the recording. No prior information about the signal of interest is needed. The presented solution is therefore expected to generalise to other signal types, e.g. noise signals for ambient noise level estimation.

7 Conclusion

This paper set out to prove the feasibility of using loudspeakers as recording devices. To this end, four areas of a system recording with one loudspeaker have been presented. First, current-measuring design concepts for the electronics were shown. Secondly, a signal model is described to provide a basis for signal processing and simulations. Thirdly, measurements are performed to identify the recording characteristics of a ceiling mounted loudspeaker range. After this, a simulation is performed to evaluate the recording capabilities together with an LMS filter. The results show that using a ceiling mounted loudspeaker, a speech signal of 50 dBSPL can be successfully estimated with reasonable intelligibility, even when the loudspeaker is simultaneously used for playback at 1W (88 dBSPL). As the filter used in the simulation does not require prior information on the signal of interest, the solution is expected to generalise to different signal types.

From the results, we conclude that using a loudspeaker as a recording device is a feasible solution in cases where adding microphones is not an option. At slightly lower signal quality, using loudspeakers instead of microphones could significantly reduce cost in a system, such as e.g. public address and voice alarm systems.

8 Discussion

As this paper set out to primarily prove the feasibility of simultaneous playback and recording with one loudspeaker, not all involved topics can be discussed in-depth. Before loudspeakers can be readily implemented as recording devices, more research needs to be done. Suggestions for continuation of the work focus on two main areas: increasing the quality and range of the recording, and improving the ease of implementation for a system as described. In the first area, some suggestions are:

- Designing a differential recording amplifier to improve the dynamic range of the signal before going to the analogue-to-digital converter.
- Improving the signal processing with more involved digital filters on the signals recorded by loudspeakers. Using case-dependent signal processing techniques are expected to significantly improve the results.
- Improving the construction of a loudspeaker to have a more flat recording response without significantly altering the playback response.

Some recommendations for further research into the second improvement area, practical implementation of the system, are:

- Evaluate and test the topology-level designs shown in Section 2.
- Evaluate compensation methods for the loudspeaker recording transfer. Equalisation based on estimates of the loudspeaker recording transfer can be evaluated for implementation over a wide range of devices.
- Extending the single loudspeaker case to the case of multiple connected loudspeakers. This is relevant in public address systems, where often multiple loudspeakers are connected to one line.

With the results presented here, using loudspeakers as microphones is also deemed to be a security sensitive issue. The feasibility shown here signifies the technique is also suitable for espionage. Security in public address and voice alarm systems is thus also very important to prevent eavesdropping.

References

[1] Kim, J.-S., "Digital public address system with central control using operational MICOM," in *The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems*, pp. 1308–1311, 2012, doi:10.1109/SCIS-ISIS.2012.6505357.

[2] Borwick, J., *Loudspeaker and headphone handbook*, Focal Press, Oxford, 2001, ISBN 9781136120855.

[3] Prasad, A., Chokhawala, A., Thompson, K., and Melanson, J., "A 120dB 300mW stereo audio A/D converter with 110dB THD+N," in *Proceedings of the 30th European Solid-State Circuits Conference*, pp. 191–194, 2004, doi:10.1109/ESSCIR.2004.1356650.

[4] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979, ISSN NA, doi:10.1121/1.382599.

[5] Kuttruff, H., *Room Acoustics*, CRC Press, 6th edition, 2016, doi:10.1201/9781315372150.

[6] Crocker, M. J., *Handbook of acoustics*, Wiley, New York, 1998, ISBN 9781523117895 1523117893.

[7] "LC1 Modular Ceiling Loudspeaker Range," https://resources-boschsecurity-cdn.azureedge.net/public/documents/LC_1_Modular_Ceeilin_Data_sheet_enUS_1924358923.pdf, 2017.

[8] Merhaut, J., "Low-Frequency Measurement of Loudspeakers By the Reciprocity Method." *AES: Journal of the Audio Engineering Society*, 30(12), pp. 882–888, 1982, ISSN 00047554.

[9] Eargle, J., "Microphone Measurerments, Standards and Specifications," in *The Microphone Book*, chapter 7, Focal Press,, Oxford, 2nd edition, 2005, ISBN 0240519612, doi:10.4324/9780080473468.

[10] Habets, E., "Room impulse response generator," https://www.researchgate.net/publication/259991276_Room_Impulse_Response_Generator, 2006.

[11] Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., and Zue, V., "TIMIT Acoustic-Phonetic Continuous Speech Corpus," 1993, doi:hdl:11272.1/AB2/SWVENO.

[12] Haykin, S., "Least-Mean-Square Algorithm," in *Adaptive Filter Theory*, Prentice-Hall, 3rd ed. NJ, USA, 1995, ch. 9, ISBN 013322760X.

[13] Van Kuyk, S., Bastiaan Kleijn, W., and Hendriks, R. C., "An instrumental intelligibility metric based on information theory," *IEEE Signal Processing Letters*, 25(1), pp. 115–119, 2018, ISSN 10709908, doi:10.1109/LSP.2017.2774250.

[14] Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J., "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Transactions on Audio, Speech and Language Processing*, 19(7), pp. 2125–2136, 2011, ISSN 15587916, doi:10.1109/TASL.2011.2114881.

[15] Van Kuyk, S., Bastiaan Kleijn, W., and Hendriks, R. C., "An Evaluation of Intrusive Instrumental Intelligibility Metrics," *IEEE/ACM Transactions on Audio Speech and Language Processing*, 26(11), pp. 2153–2166, 2018, ISSN 23299290, doi:10.1109/TASLP.2018.2856374.