ICA 2022
24th INTERNATIONAL CONGRESS ON ACOUSTICS
October 24(Mon) - 28(Fri), 2022 Gyeongju, Korea

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

!"#

Portable Spatial Super-Hearing Technology For Ultrasound Range

Ville PULKKI[(1*)], Ossi MIIKKULAINEN[(1)], Leo MCCORMACK[(1)]

[(1)]Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Finland
[(*)]Correspondence: Ville.Pulkki@aalto.fi

ABSTRACT
The spatial super-hearing technology originally proposed by the present research group brings ultrasonic signals into the audible range and auralises them using headphones, in such a manner that the listener is also able to localise the sources through spatial hearing. The signals are captured using a microphone array, and the direction-of-arrival and diffuseness parameters are analysed in the time-frequency domain. This work describes the methods and hardware employed in a fully portable implementation of the system, for a frequency range from 20 kHz to 90 kHz. Six ultrasonic digital MEMS microphones are mounted onto a 3D-printed cube with rounded edges resembling a standard dice, with face-to-face distance of 25 mm. The directivity caused by the dice was simulated with COMSOL, and the resulting pattern is plotted and discussed. In the device, the microphone outputs are multiplexed using a custom-designed converter, which also houses a 2-channel D/A-converter for sound output. The microphone signals are delivered using USB to a pocket-sized standard computer, where the processing is conducted, and the binaural signal is sent back to the converter for DA conversion. Changes made to the processing, compared to the earlier implementation, are also discussed.

Keywords: late reverberation model, sound energy decay, coupled rooms, multi-exponential decay

1 INTRODUCTION

Although the sense of hearing is sensitive to a wide range of frequencies, there is an upper frequency limit. This limit is approximately 20 kHz for young human subjects and is gradually lowered with increasing age. The frequencies above 20 kHz are commonly referred to as ultrasonic frequencies. Many animals, such as bats, rodents, insects, reptiles and amphibians produce strong vocalisations in the ultrasonic range[1], and man-made devices may also generate ultrasonic sounds in their normal or abnormal operation; such as gas leaks in pipes[2]. Ultrasonic signals can be brought to audible frequencies using signal processing techniques, for example, bats are often monitored using specific detectors [3], which can play back the down-shifted sound through a miniature loudspeaker. However, while the sounds they produce are audible to the listener, such devices do not permit the perception of the direction of ultrasonic sound sources.

The present group has recently developed a technology to render ultrasonic frequencies audible within the range of human hearing, while simultaneously allowing the directions of the ultrasonic sources to be perceived by the listener in a real acoustic environment. The first published device [4] utilizes a miniature head-mounted ultrasonic microphone array, accompanied by parametric spatial audio reproduction of the down-shifted sounds over headphones. This article discusses an improved version of the device, which has been designed for mobile use, and some changes in the microphone array design are also reported.

2 Ultrasonic super-hearing technology

The ultrasonic super-hearing technology is based on the application of time-frequency-domain parametric spatial audio techniques, which have been previously developed for the enhancement of sound-field reproduction and compression of spatial sound scenes [5]. The methods first employ spatial analysis techniques over time and frequency, in order to extract spatial parameters describing the input sound-field as captured by the employed

Figure 1. Left: Ultrasonic super-hearing device, where the microphone array and audio interface are mounted to headphones and signal processing is performed in miniature computer inside a black case seen on floor with its power bank. Right: Close-up figures of headphone-mounted parts of the device.

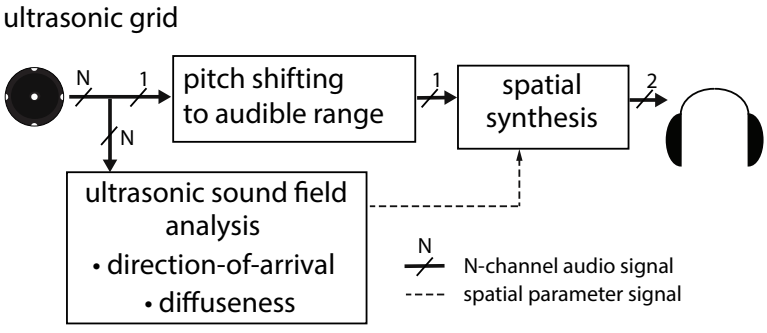

Figure 2. Signal processing chain for ultrasonic super-hearing.

microphone array. These analysed parameters are then subsequently used to synthesise the audio signals for the target reproduction setup in an adaptive and more informed manner. The processing stages of the proposed technique are depicted in Fig. 2. Note that each signal from the microphone array is first transformed into the short-time frequency domain, and the spatial analysis is conducted for each time-frequency tile independently. The parameter representing the most prominent direction-of-arrival is defined as a unit vector pointing to the direction of arriving sound $\hat{\mathbf{r}}_{\mathrm{DoA}}(t,f)$, where t and f are time and frequency parameters, respectively. The second analyzed parameter corresponds to the diffuseness of the sound-field, defined as a real-valued number $\hat{\psi}(t,f) \in [0,1]$, where ideally $\hat{\psi} = 0$ is obtained for sound-fields comprising a single plane wave, and $\hat{\psi} = 1$ is obtained for a purely diffuse-field or for fields where several sources are active at the same time.

One of the sensor signals is then also modified via a pitch-shifting or down-modulating method, in order to bring captured ultrasonic sounds down to the audible frequency range. This modified signal is first attenuated depending on analysed diffuseness, and subsequently spatialised for playback over headphones, by using the appropriate binaural filters corresponding to the analysed direction-of-arrival parameters. In practice, the system reproduces sound when a single source is dominant, and de-emphasises all other portions of signal. Non-individualised head-related transfer function (HRTF) digital filters [6] were employed for the spatialisation for the developed device.

In summary, the proposed processing approach permits frequency-modified signals to be synthesised with plausible binaural and monaural cues, which may subsequently be delivered to the listener to enable the localisation of ultrasonic sound sources. Furthermore, since the super-hearing device turns with the head of the listener, and the processing latency of the device was constrained to 44 ms, much of the dynamic cues should also be preserved. Note that the effect of processing latency has been previously studied in the context of head-tracked binaural reproduction systems, where it has been found that a system latency above 50-100 ms can impair the spatial perception [7, 8]. Therefore, it should be noted that a trade-off must be made between: attaining high spatial image and audio quality (which are improved through longer temporal windows and a higher level of overlapping) and having low processing latency (which relies on shorter windows and reduced overlapping). The current processing latency has been engineered so that both the spatial image and audio quality after pitch-shifting, as determined based on informal listening, remain reasonably high.

3 Design of the device

A mobile version of the super-hearing device built for the first tests [4] was targeted in the current project. The planning was based around miniature computers available, and on the options how to bring multichannel audio with high sampling rate for real-time processing and immediate playback.

3.1 Hardware and audio software

A miniature computer running windows operating system (Lattepanda 4GB/64GB) was selected to run the super-hearing application, primarily because it was able to import 6 channels of audio at a 192 kHz sampling rate. Many of other miniature computers investigated by the present authors did not have sufficiently fast buses for this task. A portable computer was also preferred over a dedicated microprocessor, since the software already developed by the research group could be easily utilized in the device. In practice, the REAPER audio production tool is hosted on the device, and the super-hearing audio processing is applied therein through use of an open-source VST plugin[1].

Digital MEMS microphones were chosen (Knowles SPH0641LU4H-1), and a 8 x 12 mm board housing them was designed and manufactured. The microphones were organized into groups of two, and stereophonic audio from each pair in pulse density modulation (PDM) format is conveyed to the custom converter board. On the converter board, dedicated format conversion IC's (Texas Instruments PCMD3180) are used to convert the PDM formatted signals to inter-IC sound (I^2S) format, which is supported by the USB multichannel audio interface (MiniDSP MCHStreamer). The format conversion board and USB audio interface were installed into a 3D-printed box mounted on the top of the headphones used in the device. Six audio channels with a 192 kHz sampling rate are then fed using a standard USB cord to the miniature computer.

[1] The VST audio plugin, and related MATLAB scripts, may be found here: https://github.com/leomccormack/Super-Hearing

After processing, a stereophonic audio channel (also in I^2S format with sampling frequency of 192 kHz) is routed back to the format conversion board, where a two-channel digital-to-analog converter (Texas Instruments PCM1795) with 96 kHz signal bandwidth and high current output driver (JRC NJM4556) is implemented. The analog signals are then played over standard stereophonic headphones. For demonstration purposes, a pair of headphones with active noise cancellation and inbuilt headphone amplifier is used.

3.2 Sensor array

The previous version of the device utilized analog microphones flush-mounted to a 3D-printed spherical housing. Although a sphere seems to be a natural choice for geometry, it causes some issues in the direction-of-arrival analysis method utilized in the system. In the analysis stage, the alias-free short-time Fourier transform (afSTFT) filterbank described in [9] was employed to first divide the input signals into 512 uniformly spaced frequency bands, which are then analysed independently. These time-frequency domain signals $\mathbf{x}(t,f) = [x(t,f),...,x_Q(t,f)] \in \mathbb{C}^{Q\times 1}$ are denoted with t and f to represent the down-sampled time and frequency indices, respectively. Given that the intended operating range of the system is above the spatial aliasing frequency of the array, the direction-of-arrival (DoA) unit vector, $\hat{\mathbf{r}}_{\mathrm{DoA}}(t,f)$, is estimated using the sensor-amplitude driven space-domain approach proposed in [10]. This relies on first determining the instantaneous DoA estimates as:

$$\hat{\mathbf{r}}_{\mathrm{DoA}}(t,f) = \sum_{q=1}^{Q} |x_q(t,f)|\mathbf{n}(\Omega_q), \tag{1}$$

where $\mathbf{n}(\Omega_q) \in \mathbb{R}^{3\times 1}$ are Cartesian unit vectors describing the direction of each sensor, q. Note that the array causes prominent acoustical shadowing of sound waves and the amplitude $|x_q(t,f)|$ is highest on the side of arrival, and lowest on the opposite shadowed side. When the direction vectors of the sensors are weighted with the amplitude values and summed, the resulting vector points to the most prominent direction-of-arrival of sound. Note that since these DoA estimates do not rely on inter-sensor phase relations, they are unaffected by spatial-aliasing.

The method is based on analysis of the shadowing effect by the rigid body, and optimally the directional pattern of the microphone should be a unidirectional cardioid pattern. However, the physical size of the array exceeds the spatial aliasing limit, and instead of unidirectional pattern, strong sidelobes are obtained, depending on frequency.

The digital MEMS microphones used in the device are mounted on small electronic boards, which calls for cubical arrangement of them, differing from spherical arrangement used before. The cubical arrangement was thought of as an ideal geometry for the array, and it was postulated that a combination of spherical and cubical geometries would provide more even shadowing patterns compared to those obtained by a spherical geometry. This proposed geometry was therefore first studied, where a sphere with a radius of 15 mm was cut on six sides of a cube with edge length of 25 mm. The geometry is therefore reminiscent of a dice, as shown in right side of Fig. 3. The geometry is not regular, and diffraction effects should be less systematic than with sphere. This was thought to produce pattern that would have less salient sidelobes. The shadowing effect was verified using COMSOL simulation. The simulation was run by placing an ideal pressure source on the position of the microphone mounted on the rigid bodies, and by computing the sound pressure level at a distance of 100 mm from the center of the body. The results of simulation for both spherical and dice-shaped arrays are shown in Fig. 3, where it can be observed, that on the contralateral side of the source the sidelobes are much less prominent, and more irregular than with spherical rigid body.

The dice-shaped body is thus practical in usage with microphones mounted on small circuit boards, and the more-diverse shawdowing effects produced by the body can be used by the direction-of-arrival estimator employed by the system.

3.3 Signal processing

The signal processing methods adopted in the revised version were modified only slightly in the current article. The pitch shifting method in earlier version was based on the use of phase vocoder [11, 12]. Phase vocoders are used typically to make moderate shifts in pitch, and the up-to three-octave shift required for the present super-hearing application did not always produce pitch-shifted signals of high signal quality. In the current version,

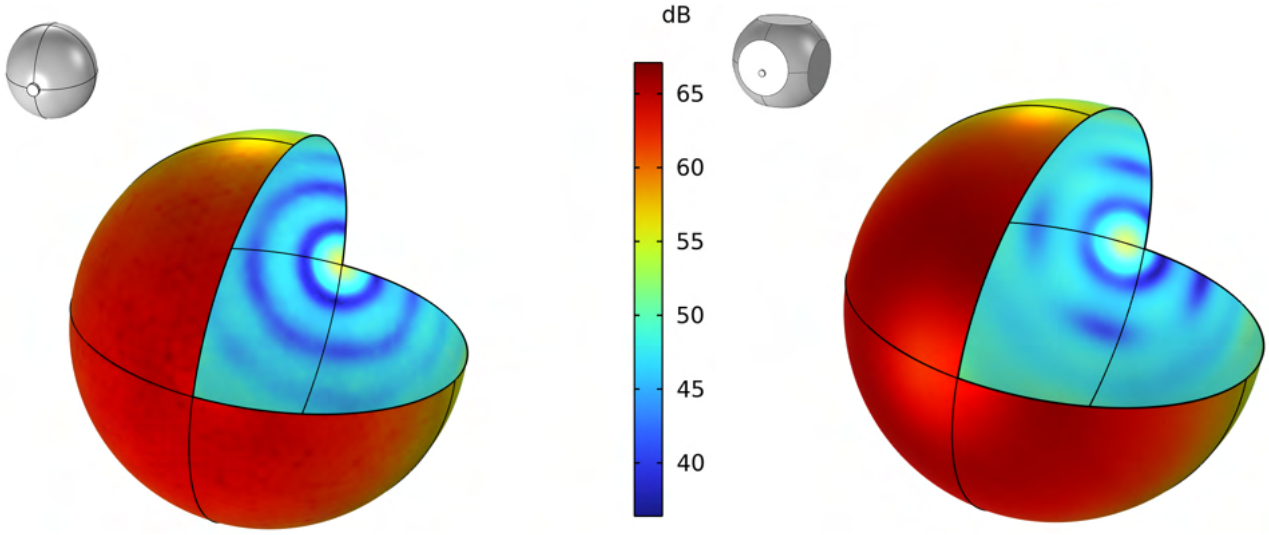

Figure 3. Sound pressure level simulated on a spherical surface at a distance of 100 mm from a rigid body, given a point source emanating from one of the microphones attached to the surface of the rigid body. Left: Spherical rigid body with 15-mm radius. Right: "dice" rigid body composed as intersection between a 15mm-radius sphere and a cube with 25-mm edge length. Due to reciprocity, the SPL surface also represents the directional pattern of the microphone.

a simpler resampling method operating based on the time-domain signals was employed. In practice, in each time window with length of 2 ms, each microphone signal is time-stretched by a factor of 8, and overlap-added with the tails of earlier time windows. The pitch-shifted signal is then reproduced using the same methods as in the first version of the technology. The benefit of the simple resampling method is that the shift of each frequency is deterministic, although in some cases some frequency components can be heavily attenuated, and time structure can be smeared, due to destructive interference in the overlap-add processing.

4 Summary

This article describes the development a mobile version of an ultrasonic super-hearing device, which was previously proposed by the present audio research group in [4]. The device features an ultrasonic 6-microphone array, which is mounted onto a pair of headphones. The most prominent direction-of-arrival of sound in the region between 20 kHz and 60 kHz is then estimated, based on the array signals. The sum of all 6 signals is then pitch shifted to the audible frequency range, and subsequently auralised in the analysed direction. The device and proposed processing therefore allows the wearer to both hear and localise ultrasonic sound sources, such as bats or leaks in pressurised gas pipes. The paper details the development of a mobile/portable version of this device, and discusses the modifications made to the algorithms compared to the first implementation of the technology. Namely, the most significant changes were: the usage of digital MEMS microphones, as apposed to analog microphones; the development of a processing chain featuring a miniature portable computer, multichannel USB card and a dedicated processing board for input audio format conversion and output D/A conversion; and the usage of a dice-shaped rigid array, which housed the MEMS microphones.

ACKNOWLEDGMENTS

This research has received funding from the Aalto University Doctoral School of Electrical Engineering and the Academy of Finland, project no. 317341.

REFERENCES

[1] G. Sales, *Ultrasonic communication by animals*. Springer Science & Business Media, 2012.

[2] W. Tao, W. Dongying, P. Yu, and F. Wei, "Gas leak localization and detection method based on a multi-point ultrasonic sensor array with TDOA algorithm," *Measurement Science and Technology*, vol. 26, no. 9, p. 095002, 2015.

[3] M. Barataud, "Acoustic ecology of European bats," *Species, identification, study of their habitats and foraging behaviour. Biotope, Mèze*, 2015.

[4] V. Pulkki, L. McCormack, and R. Gonzalez, "Superhuman spatial hearing technology for ultrasonic frequencies," *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.

[5] V. Pulkki, S. Delikaris-Manias, and A. Politis, *Parametric time-frequency domain spatial audio*. Wiley Online Library, 2018.

[6] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, 1995.

[7] A. Lindau and S. Weinzierl, "Assessing the plausibility of virtual acoustic environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810, 2012.

[8] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. Katz, and C. De Boishéraud, "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 2011–2023, 2017.

[9] J. Vilkamo and T. Bäckström, "Time-frequency processing: Methods and tools," *Parametric Time-Frequency Domain Spatial Audio*, p. 3, 2017.

[10] A. Politis, S. Delikaris-Manias, and V. Pulkki, "Direction-of-arrival and diffuseness estimation above spatial aliasing for symmetrical directional microphone arrays," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6–10, IEEE, 2015.

[11] S. M. Bernsee, "Pitch shifting using the Fourier transform," *The DSP Dimension, http://blogs. zynaptiz. com/bernsee/pitch-shifting-using-the-ft*, 1999.

[12] U. Zölzer, *DAFX: digital audio effects*. John Wiley & Sons, 2011.