Journal of Neuroscience Methods 309 (2018) 60–70

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Acoustic calibration in an echoic environment

Alexander Kazakov[a], Israel Nelken[a,b,*]

[a] Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel
[b] Dept. of Neurobiology, The Alexander Silberman Inst. of Life Sciences, The Hebrew University of Jerusalem, Israel

ARTICLE INFO

Keywords:
Acoustic calibration
Sound distortions
Golay complementary sequences
Echoic environments
Freely moving animals
3D localization

ABSTRACT

Background: The sound fed to a loudspeaker may significantly differ from that reaching the ear of the listener. The transformation from one to the other consists of spectral distortions with strong dependence on the relative locations of the speaker and the listener as well as on the geometry of the environment. With the increased importance of research in awake, freely-moving animals in large arenas, it becomes important to understand how animal location influences the corresponding spectral distortions.
New method: We describe a full calibration pipeline that includes spatial sampling and estimation of the spectral distortions. We estimated the impulse responses of the environment using Golay complementary sequences.
Using those sequences, we also describe an acoustic 3D localization method for freely moving animals.
Results: In our arena, the impulse responses are dominated by a small number of strong reflections. We use this understanding to provide guidelines for designing the geometry of the environment as well as the presented sounds, in order to provide more uniform sound levels throughout the environment. Our 3D localization method achieves a 1.5 cm accuracy through the utilization of sound cues only.
Comparison with existing methods: To our knowledge, this is the first description of a large-scale acoustic calibration pipeline with acoustic localization for neuroscience studies.
Conclusions: Principled sampling of large arena allows for better design and control of the acoustic information provided to freely-moving animals.

1. Introduction

A growing number of studies in brain sciences use freely moving animals, a process that is driven by technological advances such as the possibility to perform extracellular recordings with large electrode arrays using telemetry (Fan et al., 2011; Jow et al., 2012). However, less constrained behavior comes with a cost in terms of stimulus control. In audition, for example, sounds reproduced by loudspeaker are distorted as they propagate through the environment, so that the sound reaching the ear is different from that initially produced by the loudspeaker (Fig. 1A.). These distortions are usually estimated through acoustic calibration – the comparison between the sound at the input to the loudspeaker with that actually reaching the ears. For freely-moving animals, the need to calibrate sounds in multiple locations is an especially hard task.

Since auditory communication occurs at sound levels where nonlinear effects are negligible, acoustic distortions are fully characterized by measuring the responses to an impulse (the 'impulse response'). Evaluation of the acoustic properties using an impulse response is called 'Impulse response analysis' and is thoroughly described by Warren (Warren, 2014). In short, the straightforward method to measure the impulse response is to record the responses to a single, ideal impulse. However, practical implementations of ideal impulses have low acoustic power because of their short duration, leading to low signal to noise ratio (SNR) in the resulting estimates. Instead of a single impulse, it is possible to use a pseudo-random noise, which is later cross-correlated with the recorded response, resulting in the impulse response of the system. For this to work well, the test signal has to have an autocorrelation consisting of a single peak at zero delay and zero correlation at all other delays. To maximize the quality of the estimates the test signal should also be of maximal power. For example, the use of maximal length sequences has been proposed by Schroeder (Schroeder, 1979). Maximal length sequences, however, are not exactly white and thus can't be used for estimating a perfect impulse response. To solve this problem, Golay complementary sequences were proposed (Foster, 1986) and later applied for acoustic measurement in a number of studies of the auditory system (Jacobson et al., 2001; Mrsic-flogel et al., 2005; Zhou et al., 1992). Even though Golay complementary sequences require playback and recording of two different sequences, the test signal is shorter than its pseudo-random equivalents, resulting in faster

* Corresponding author.
E-mail address: Israel.nelken@mail.huji.ac.il (I. Nelken).

https://doi.org/10.1016/j.jneumeth.2018.08.025
Received 25 May 2018; Received in revised form 21 August 2018; Accepted 21 August 2018
Available online 27 August 2018
0165-0270/

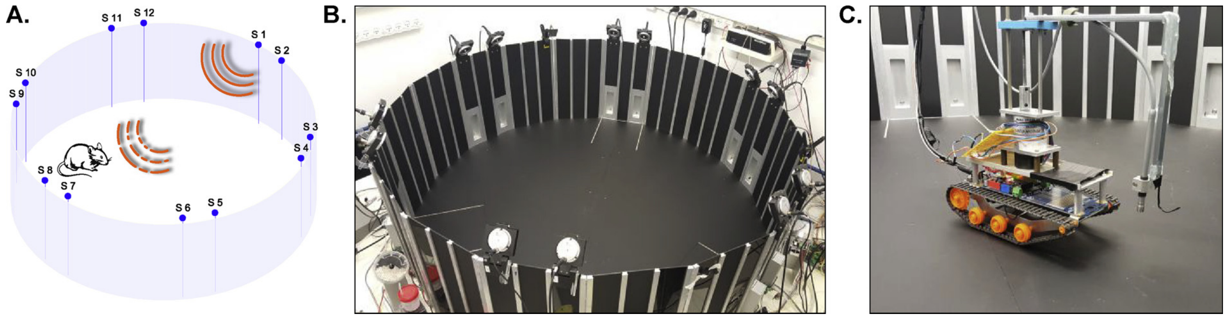

Fig. 1. The echoic environment and the sound sampling robot.
A. Simplified model of the arena, depicting the twelve speakers arranged in pairs (blue points along the circumferential wall). The arena is designed for experiments with freely behaving rats, driven by auditory stimuli (solid orange lines). Acoustic calibration compares the sound at the rat's location to that transduced by the computer, and reflects both the speaker and the arena effects on the sound. **B.** A photograph of the real arena. **C.** The acoustic calibration was performed in hundreds of locations along the arena, using a robot (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

estimation of the impulse response (Foster, 1986). Müller and Massarani (Müller and Massarani, 2001) proposed the use of frequency sweeps as an alternative way for sampling the transfer function. But while the SNR is indeed improved, the sequential sweeping along a wide frequency range can be time-consuming. We therefore used Golay complementary sequences to achieve a high throughput calibration pipeline.

In this paper, we describe an automated pipeline for high throughput sound calibration in a large arena used in our own research. We demonstrate the use of acoustic calibration for understanding the geometrical origin of the resulting acoustic distortions, and illustrate some approaches to alleviate them. We then suggest a second use for the estimated impulse responses: we exploit them for rapid 3D localization of the animal. Fast and accurate localization of freely moving animals is an open challenge. Although many generic approaches have been proposed for object tracking, only few are applicable for studies of freely moving rodents. Tracking an animal in three dimensions (3D, volumetric tracking) is even harder. Camera based approaches require line of sight with the animal and installation of multiple cameras (Lai et al., 2011; Yovel et al., 2010). Solutions based on magnetic tracking require installation of dedicated sensors and may not provide sufficient spatial resolution (Jow et al., 2012). Here, the same hardware used for acoustic calibration (multiple loudspeakers with a microphone on the head of the animal) is used also for high-resolution, fast 3D tracking. While the localization problem can be solved by the inverse method – locating a sound source positioned on the animal using a microphone array (Strobel and Rabenstein, 2000) – this would have required hardware that was not available in our setup.

2. Materials and methods

2.1. The arena

The arena is roughly circular (the circle is approximated by 18 straight segments, with diameter of 160 cm and wall height of 50 cm, see Fig. 1A and B). Six pairs of speakers (MF1, TDT) are evenly spread around the upper rim of the wall, with distance of 17 cm between speakers of each pair and central angle spread of 60° between successive pairs of speaker. All speakers are tilted by 35° below the horizontal axis such that their acoustic axes are pointed towards the center of the arena (Fig. 2 in Kazakov and Nelken, submitted).

2.2. Acoustic modeling of the arena

We idealize the acoustics of the arena as a linear, time-invariant system. The system is therefore fully specified by the impulse responses (IRs) from each speaker to each point in the arena. While the linearity assumption is very good at the sound levels considered here, the time invariance is clearly an approximation, since the presence of an animal in the arena modifies the acoustic properties in a way that depends on animal location. Nevertheless, the structure of the impulse responses is primarily determined by geometric factors that are independent of animal location. This point is further addressed in the Results.

The main challenge here is the need to sample the impulse response from multiple speakers to densely spaced points in the arena. Full reconstruction of the pressure field at all relevant frequencies requires the distance between spatially sampled points to be a fraction of the wavelength at the highest frequency of interest. However, this is practically impossible: rats hear up to 75 kHz, where the wavelength is 4 mm, requiring a sampling density of about 1 mm (Kelly and Masterton, 1977). In practice, we sampled the IRs with spatial resolution of about 1 cm, using a custom built robot (Fig. 1C). While this resolution limits the ability to fully reconstruct the resulting pressure field to frequencies below 10 kHz, we will show that useful information can nevertheless be extracted at higher frequencies as well.

The Fourier transform of the IR (Fig. 2A) is called the transfer function (TF, Fig. 2B). The transfer function describes the steady-state amplitude gain and phase shift of sinusoidal signals as a function of frequency, although when plotting TFs we invariably plotted their amplitudes only. As discussed later, we were mostly interested in the changes in gain as a function of frequency, rather than in the absolute values of those gains. The TFs are therefore plotted in dB scale relative to an arbitrary reference. The TFs varied with the relative location of the microphone and the speaker that produced the IRs. We demonstrated this both by repositioning the microphone inside the arena, while estimating the IRs from a single speaker (Fig. 2C), or by measuring the IRs from different speakers at a microphone in a fixed location (Fig. 2D). The frequency-dependent variations are not produced by the noise in the measurement, since these were highly reproducible for a fixed location (Supp. Fig. 1).

2.3. Impulse response measurements using Golay sequences

The IRs were measured by playing two complementary Golay sequences, as suggested by Zhou (Zhou et al., 1992) and illustrated in Fig. 3 in Kazakov and Nelken (submitted). Starting from $a_1 = [1\,1]$ and $b_1 = [1 -1]$, Golay sequences of length 2^{i+1} were constructed from Golay sequences of length 2^i following a recursive concatenation: $a_{i+1} = (a_i\,|b_i)$, $b_{i+1} = (a_i\,|-b_i)$ where $(\cdot|\cdot)$ denotes concatenation and $(-)$ denotes elementwise negation.

These two sequences were typically played with a time interval of 0.2 s between them, to let most of the echoes decay (this interval is referred to below as the 'silence buffer'; its duration was varied and reduced to 10 ms when we studied 3D localization). The IR was recovered by cross-correlating the recorded sounds with the played sequences and summing the results obtained for the two sequences. The