Design of Pseudo-Spherical Microphone Array with Extended Frequency Range for Robot Audition

Vladimir Tourbabin and Boaz Rafaely
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Shava, Israel
Email: vtourbabin@gmail.com, br@bgu.ac.il

Abstract

Microphone arrays constitute the front end for sound acquisition in the auditory system of most humanoid robots. Their design and performance therefore play a central role in robot audition. Although some previous studies are concerned with the optimization of microphone placement for robot audition, spatial aliasing, constituting a major challenge in array design, has not been studied extensively for this application. Spatial aliasing is approached in this paper using a spherical harmonics (SH) framework. A method for microphone positioning that extends the aliasing-free frequency range of low SH orders is developed, and validated by simulation for the design of a microphone array on the head of the humanoid NAO. Aliasing level is significantly reduced compared to the efficient nearly-uniform microphone distribution. The proposed method was employed in the implementation of a 12-microphone array for NAO.

Introduction

Microphone arrays are becoming increasingly important for robot audition, and their design around the head or the body of robots becomes a great challenge. Recent publications study the optimization of microphone placement for enriching the spatial information acquired by the array [1] and for improving the sound localization performance [2]. However, spatial aliasing, being a major problem in array design, has not been studied to a great extent in the humanoid-robot audition literature. Due to the sphere-like shape of a humanoid robot head on which the microphones are usually distributed, the aliasing problem is approached here using the spherical harmonics (SH) framework [3]. A method for microphone positioning that extends the aliasing-free frequency range of low SH orders is presented. The reduced aliasing is achieved by choosing the most appropriate microphone positioning. The method can be used to complement the existing techniques for aliasing cancellation by signal processing [4].

The proposed method is studied by computer simulations of a microphone array design around the head of the humanoid robot NAO. Performance of the resulting optimal configuration is evaluated against the efficient nearly-uniform microphone distribution, illustrating the advantage of the proposed method in providing an extended operating frequency range.

In addition, the proposed method is applied to the design of a new 12 microphone array for NAO. This design is subject to the real constraints on microphone positioning due to cameras, loudspeakers, and other components inside the robot head, therefore validating the application of the method for real robots.

Design of spherical microphone arrays

Spherical microphone arrays are typically composed of a set of microphones positioned on the surface of a sphere. In this case the sound pressure at the microphones can be written as [3]

$$p(k, r_q, \theta_q, \phi_q) = \sum_{n=0}^{N} \sum_{m=-n}^{n} a_{nm}(k) b_n(kr_q) Y_n^m(\theta_q, \phi_q) \tag{1}$$

where p is the sound pressure, k is the wave number, (r_q, θ_q, ϕ_q) is the position of microphone q in spherical coordinates, $a_{nm}(k)$ is the plane wave density of the sound field around the microphone array, represented in the spherical harmonics domain, $b_n(kr)$ is the radial function, dependent on the sphere configuration, and $Y_n^m(\theta, \phi)$ is the spherical harmonic of order n and degree m. It is assumed that the array is composed of Q microphones, and that the sound pressure is order limited in the spherical harmonics domain, i.e. $n \leq N$. This assumption typically holds at frequencies that satisfy $kr < N$, due to the decay of the radial functions [3].

A good engineering design will typically ensure that Q is sufficiently large, i.e. $Q \geq (N+1)^2$, to allow the computation of $(N+1)^2$ spherical harmonics coefficients using Q microphones. Furthermore, the array radius is chosen in accordance with the operating frequency range to satisfy $kr < N$. An example of such a design is the eigenmike [5], of radius $r = 4.2$ cm, with $Q = 32$ microphones, and a maximum order of $N = 4$ designed for the speech frequency range. Under these assumptions, Eq. (1) can be written in a matrix form as

$$\mathbf{p} = \mathbf{B}\mathbf{a} \tag{2}$$

where $Q \times (N+1)^2$ matrix $\mathbf{B}$ holds functions $b_n(kr_q)Y_n^m(\theta_q, \phi_q)$, $Q \times 1$ vector $\mathbf{p}$ holds the pressure values, and $(N+1)^2 \times 1$ vector $\mathbf{a}$ holds $a_{nm}(k)$. Now, a_{nm} can be computed using

$$\hat{\mathbf{a}} = \mathbf{B}^{\dagger}\mathbf{p} = \mathbf{B}^{\dagger}\mathbf{B}\mathbf{a}, \tag{3}$$

in which case $\mathbf{B}^\dagger\mathbf{B} \approx \mathbf{I}$ for a well designed system, with $\mathbf{B}^\dagger$ representing the Moore-Penrose pseudo inverse of matrix $\mathbf{B}$.

However, some designs impose constraints, in which case the good design practice cannot be maintained. One example is the design of a microphone array on a robot head. The sphere-like shape is suitable for the design and processing in the spherical harmonics domain. However, the number of microphones may be constrained by hardware limitations, while the head radius may not be subject to modification. In this case, the desired frequency operating range, e.g. for speech, may not satisfy $kr < N$. At high frequencies, for which $kr > N$, higher-order coefficients become significant. In this case $\mathbf{a}$ is composed of high-order harmonics, up to say $\tilde{N} > N$, with dimensions $(\tilde{N}+1)^2 \times 1$, leading to the matrix $\tilde{\mathbf{B}}$ of dimensions $Q \times (\tilde{N}+1)^2$. In this case, $\mathbf{B}^\dagger\tilde{\mathbf{B}}$ is no longer square and may not satisfy $\mathbf{B}^\dagger\tilde{\mathbf{B}} \approx [\mathbf{I}\ \mathbf{0}]^T$. The estimation of the plane-wave density, $\hat{\mathbf{a}}$, will therefore suffer form spatial aliasing error, and array performance may degrade significantly [3].

Design with reduced spatial aliasing

In this section a new design approach is presented that aims to reduce the effect of spatial aliasing, typically when the operating frequency range satisfies $kr > N$. The aliasing matrix is denoted by

$$\mathbf{D} = \mathbf{B}^\dagger\tilde{\mathbf{B}}. \tag{4}$$

Matrix $\mathbf{D}$ is of dimensions $(N+1)^2 \times (\tilde{N}+1)^2$, with N denoting the desired array operating order, and $\tilde{N}$ represents the actual order of the sound field as measured by the array. When $\tilde{N} > N$, spatial aliasing may occur. Furthermore, matrix $\mathbf{D}$ is typically frequency dependent, and so the effect of aliasing may also be frequency dependent.

In this paper, design with reduced aliasing is approached by optimal positioning of microphones. Microphone positions are defined in matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ and therefore affect aliasing in matrix $\mathbf{D}$. A general scalar cost function is first defined, to quantify the magnitude of the aliasing error in matrix $\mathbf{D}$, as a function of orders and frequency

$$\rho = g(\mathbf{D}, \mathcal{S}) \tag{5}$$

with $\mathcal{S}$ denoting the set of orders, n, $\tilde{n}$, and frequencies, of interest, and g denoting the cost function, representing, for example, maximum value or average value of the off-diagonal elements in $\mathbf{D}$.

One possible cost function could include all array orders $0 \le n \le N$, all sound field orders, $0 \le \tilde{n} \le \tilde{N}$, and the given operating frequency range, $f_{low} \le f \le f_{high}$, with a cost function that represents the element in matrix $\mathbf{D}$ with the maximum magnitude over the set $\mathcal{S}$.

Another possible cost function could include the same set $\mathcal{S}$ but compute the *average* values of the off-diagonal elements in matrix $\mathbf{D}$ in the defined set. A third possibility could focus on specific orders in n and $\tilde{n}$, and specific frequencies.

Array design for NAO

A design example of a microphone array around the head of the humanoid robot NAO [6] is presented in this section. An illustration of the head of Nao is presented in Fig. 1. The head has a sphere-like shape with a radius of around 6.25 cm. In this design example, 12 microphones are positioned on the head of NAO. With 12 microphones, the expected array order is $N = 2$. However, if $\tilde{N} > 2$, aliasing may occur. Using a sphere model for the head of NAO, the equality $kr = 2$ occurs at 1.7 kHz. This is well below the desired maximum frequency, which may be around 4 kHz or even higher for speech signals.

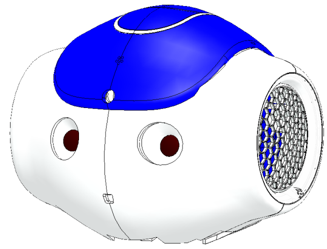

Abbildung 1: The head of the humanoid robot NAO.

As an initial design, 12 microphone are positioned in a nearly uniform manner [3] around NAO's head. Matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ are calculated from the simulated steering vectors of a boundary element model of the robot [7], and matrix $\mathbf{D}$ is calculated from Eq. (4) as a function of frequency, and for $N = 1$ and $\tilde{N} = 4$. Note that $N = 1$ is chosen to ensure a sufficiently low level of aliasing. In principle $N = 2$ could be selected, but the overall aliasing error in this example is expected to be too high.

With the aim of reducing the effect of spatial aliasing, an objective function has been defined according to Eq. (5), with g realizing the maximum magnitude over the elements of matrix $(\mathbf{D} - \mathbf{I})$, i.e. off-diagonal elements of $\mathbf{D}$, for $0 \le n \le 1$ and $0 \le \tilde{n} \le 4$, and for a frequency of 3.5 kHz. It has been observed that optimization at a high frequency reduces aliasing at the lower frequency range as well. The Matlab interior-point optimization solver `fmincon()` has been employed for computing microphone positions.

The values of $\max|\mathbf{D} - \mathbf{I}|$ at a range of frequencies is presented in Fig. 2 for the two designs. The figure clearly shows reduced aliasing errors over a wide frequency range for the minimum-aliasing design compared to nearly uniform sampling. A reduction of up to about 7 dB is achieved at the middle frequency range.

The magnitude of the elements of matrix $\mathbf{D}$ for the two designs at frequencies 1 kHz and 3.5 kHz are plotted in Figs. 3 and 4, respectively. At 1 kHz, both designs show a very small magnitude for the off-diagonal elements. This is because this frequency is within the aliasing-free operating range. However, at 3.5 kHz, the off-diagonal elements are more significant, and aliasing error is expected. Nevertheless, the minimum-aliasing design shows a lower maximal error compared to the nearly uniform design.

The proposed design method has been implemented to produce a set of microphone positions for the head of NAO, for an array composed of 12 microphones. Due to practical constraints on the allowed microphone positions, the minimum-aliasing design presented above could not be implemented directly. Instead, the regions on the robot head at which microphones could not be positioned, covering about two-thirds of the head area, were first defined. Then, a semi-exhaustive search was performed to allocate the best 12 positions out of possible 327 positions, which were uniformly distributed over the non-restricted area. The search was performed by exhaustively selecting sets of the best 3 microphones at a time. Figure 5 shows the aliasing error achieved for this case. The error is higher compared to Fig. 2. This is an example of a real-world design - constraints on microphone positioning may come at the expense of performance.

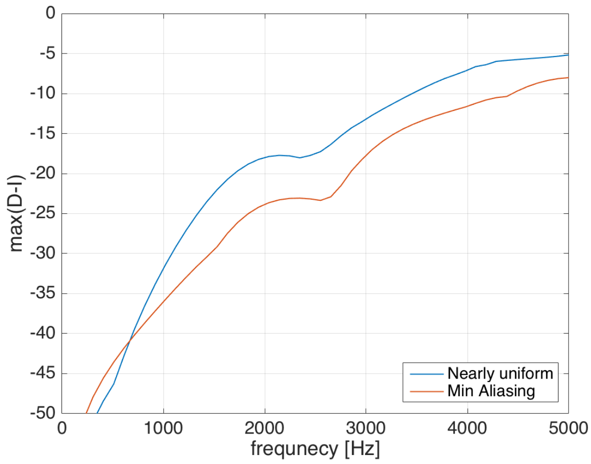

Abbildung 2: The aliasing error, defined as the maximum value of the off-diagonal elements of matrix **D**, for the nearly uniform design and the minimum aliasing design, as a function of frequency.

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465.

Conclusion

This paper presented a new method for microphone placement on the head of humanoid robots. The method minimizes spatial aliasing, therefore extending the useful operating frequency range of the array. The application of the method to the design of an array for the humanoid NAO has been presented.

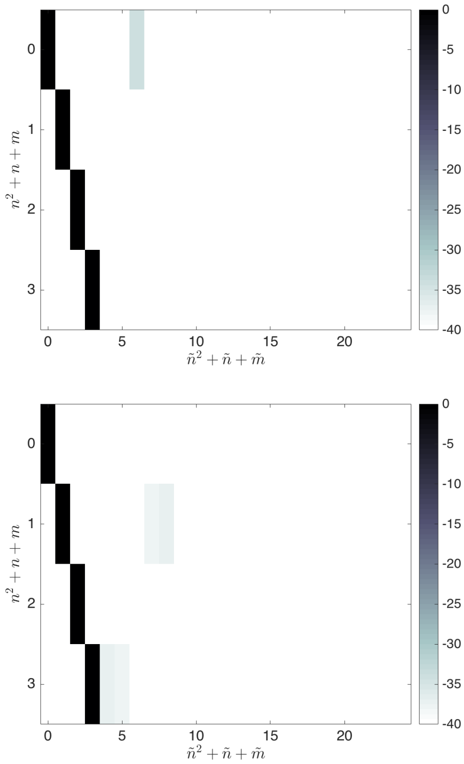

Abbildung 3: The magnitude of the elements of matrix **D**, for the nearly uniform design (top) and the minimum aliasing design (bottom), at 1 kHz.

Literatur

[1] Tourbabin, V. and Rafaely, B.: Theoretical Framework for the Optimization of Microphone Array Configuration for Humanoid Robot Audition. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(12) (2014), 1803-1814.

[2] Skaf, A. and Danes, P.: Optimal positioning of a binaural sensor on a humanoid head for sound source localization. IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2011, 165-170.

[3] Rafaely, B.: Fundamentals of spherical array processing. Springer 2015.

[4] Alon, D.L. and Rafaely, B.: Beamforming with Optimal Aliasing Cancellation in Spherical Microphone Arrays. IEEE/ACM Transactions on Audio, Speech, and Language Processing 24(1) (2016), 196-210.

[5] The eigenmike, by mhacoustics. `http://www.mhacoustics.com`.

[6] The humanoid robot Nao, by Aldebaran Robotics. `https://www.aldebaran.com/en/cool-robots/nao`.

[7] Tourbabin, V. and Rafaely, B.: Direction of Arrival Estimation Using Microphone Array Processing for Moving Humanoid Robots. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(11), (2015).

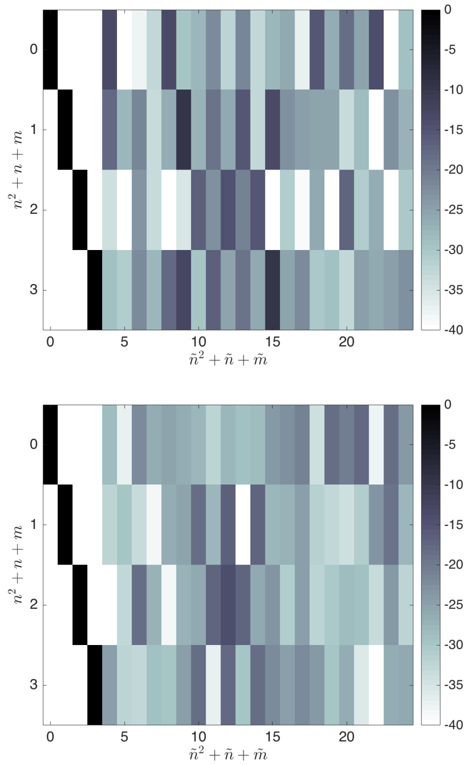

Abbildung 4: The magnitude of the elements of matrix $\mathbf{D}$, for the nearly uniform design (top) and the minimum aliasing design (bottom), at 3.5 kHz.

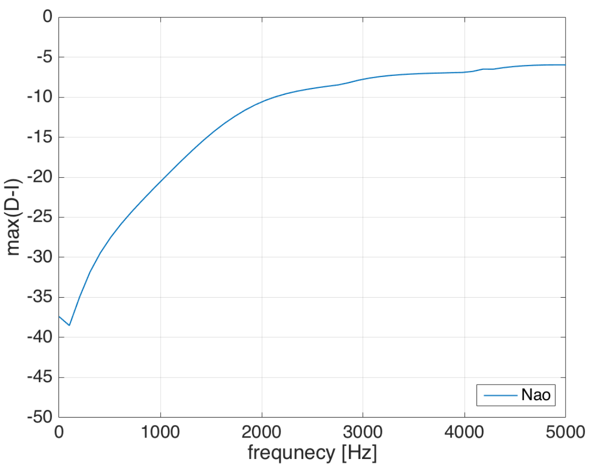

Abbildung 5: The aliasing error, defined as the maximum value of the off-diagonal elements of matrix $\mathbf{D}$, for the constrained Nao design, denoted benchmark I, as a function of frequency.