Sound Sketch:
Shaping sound in space and time using loudspeaker arrays

Jung-Woo Choi[1]

[1] Korea Advanced Institute of Science and Technology, Korea

ABSTRACT

The control of the spatial distribution of noise and sound has been widely investigated for achieving passive and active noise control. By extending the concepts of active noise control and zones of quiet, one can sketch sound, i.e., synthesize sound fields of various shapes using loudspeaker arrays to achieve the desired sound quality in a finite zone of interest. Sound sketching involves reproducing existing sound fields as well as creating new shapes of sound fields that can scarcely be observed in common acoustical spaces. In this lecture, two representative problems related to sound shaping are presented: the synthesis of a personal sound zone, and the manipulation of virtual sound sources. Through the shaping of sound in space, sound fields can be focused over a selected area to form a personal sound zone within which a person hears only a specified sound program without being disturbed by other unwanted programs. In addition to focusing sound fields, it is possible to control a sound wavefront to produce virtual sound sources with various radiation patterns in a source-free space. The underlying array signal processing theories, developed since 2000, are introduced in regard to practical systems implemented for various applications. The novel paradigm for sound field control systems, the Sound of Things, is introduced with a sound sketch interface devised to facilitate gesture-based interaction between a person and a sound field.

Keywords: sound field control, sound manipulation, sound field reproduction, acoustic contrast, sound of things, sound sketch. I-INCE Classification of Subjects Number(s): 23.5

1. INTRODUCTION

Throughout the last century, researchers in sound and vibration control have produced different kinds of sound images. For noise control engineers, the sound picture needs to be erased for the region in which the listener is located; this is the concept of “zone of quiet” in active noise control. In the area of sound field reproduction, people want to replicate the scenery of an entire sound field, such as that of a famous opera house or stadium. Even when we enjoy a movie, we hear changes in sound picture that have been carefully designed by audio engineers from scene to scene.

The control of sound fields using loudspeaker arrays, i.e., sound sketching, is a topic of considerable breath that covers noise control, spatial audio, and acoustic communications. Although their purposes and desired shapes of sound fields are all different, sound field control techniques based on linear acoustics have one thing in common: utilizing interferences of sound fields generated by multiple acoustic transducers or loudspeakers. This mechanism has been known since the discovery of Huygens’ principle and the Kirchhoff-Helmholtz integral equation and looks quite simple in nature, but the control strategy can be diverse, depending on the objectives of sketching.

As noise and sound control technologies evolve, the sound field or sound picture we want to draw becomes complex. Various temporal and spatial impressions perceived in the sound field are being evaluated by objective and subjective measures, and real-time maneuverability of such measures is becoming important. For the generation of a sound field in a desired shape, therefore, the multi-dimensional aspects of a sound field should be considered in both the analysis and synthesis stages.

Generally, the sketching of a sound field can be categorized into two problems: what to draw and how to draw it? The first problem has to do with a question on how to define the “desired shape” of the

[1] khepera@kaist.ac.kr

sound field. The target or desired sound field should be mathematically defined as an object function that can exactly reflect what we want to achieve. Assigning appropriate constraints of a given problem can also help to define the target. In sound field control, constraints to the problem are required to control only selected acoustic quantities without affecting others, as well as to prescribe physical limitations due to finite number of loudspeakers or limited zone size.

The second problem involves finding the best way to achieve the desired shape of a sound field. The target sound field should be expressed in terms of sound fields of loudspeakers and their contribution, i.e., excitation signals have to be within the physically realizable range. The optimal solution to the defined object function can be derived in a closed form or can be found via optimization techniques.

The aim of this article is to show how sound fields can be designed and synthesized differently for various applications. To this end, sound field control techniques of two representative areas—sound field reproduction and manipulation(Fig. 1)—are introduced to demonstrate differences in their object functions and optimization strategies.

Apart from these two theoretical problems, practical issues on the construction of the loudspeaker array system and the interactive sound field authoring interface are needed to be resolved in order to spread array-based technologies into the market. Instead of making an array system with pre-determined loudspeaker positions, for example, we can consider combining multiple sounding objects distributed in space as a single-array system. In this way, the complexity in constructing array systems can be overcome through the intelligent networking between smart sounding objects, which are automatically reconfigurable, flexible in layout construction, and can even sense the acoustic environment. In this regard, a new paradigm called the Sound of Things(SoT) is introduced, and it is explained how the technologies developed for noise and vibration control problems can be utilized for its implementation. A pilot study is presented in order to realize the SoT, using commercial devices and overcoming a series of technical challenges.

(a) Original

(b) Reproduction

(c) Manipulation: the Houses of Parliament (1899~1903, Claude Monet(1840~1926))

Figure 1 – Concept of sound field reproduction and manipulation: their analogy to paintings
(a) Original image (b) reproduced picture (c) manipulated pictures with different impressions

2. Sketching the sound field using loudspeaker arrays

2.1 Problem statement

In sound field control problems, we have multiple loudspeakers, such as shown in Fig. 2, driven by multichannel control signals, i.e., a source signal filtered by multichannel filters. Sound fields produced by multiple loudspeakers interfere in space and time to form a certain shape of resultant sound field. The spatial distribution of a sound field can be controlled over a finite zone of interest(V) through the tuning of relative magnitudes and weights of multichannel filters.

For ease of description, assume that a sound field is generated from a source signal of unit amplitude and single frequency ω. Multiple loudspeakers are positioned at $\vec{r}_s^{(n)}$ ($n = 1, \cdots, N$), and their sound fields are measured at discrete locations $\vec{r}^{(m)}$ ($m = 1, \cdots, M$) sampling the zone of interest(Fig. 3). By denoting the transfer function of each loudspeaker measured at $\vec{r}^{(m)}$ as $h(\vec{r}^{(m)} \mid \vec{r}_s^{(n)})$, we can construct a $M \times N$ matrix $\mathbf{H}$ ($[\mathbf{H}]^{(m,n)} = h(\vec{r}^{(m)} \mid \vec{r}_s^{(n)}) / \sqrt{M}$) relating acoustic responses between multiple loudspeakers and multiple microphones. Note that, for brevity, the transfer function matrix is normalized by the square root of the number of microphones, and its frequency dependency is omitted. Then, the sound pressure field vector, also normalized by the square root of M ($\mathbf{p} = [p(\vec{r}^{(1)}), \cdots, p(\vec{r}^{(M)})]^T / \sqrt{M}$), can be described in terms of the matrix $\mathbf{H}$ and multichannel filter coefficients $\mathbf{q} = [q(\vec{r}_s^{(1)}), \cdots, q(\vec{r}_s^{(N)})]^T$.

$$\mathbf{p} = \mathbf{Hq} \tag{1}$$

If our aim is to draw a picture of the sound pressure field, then the problem is to find the multichannel filter coefficients $\mathbf{q}$ that produce a target pressure field $\mathbf{p}_d$. The sound field control problem addressing the reproduction of the pre-defined target field is called the *sound field reproduction* problem(Fig. 1(b)). As already mentioned, however, the target sound field can be defined in many different ways, depending on the objective. The pressure field is the most popular choice and has been extensively studied, but there can be many different types of target fields of particle velocity, acoustic intensity, or energy.

Figure 2 – Loudspeaker array systems for the sound field control

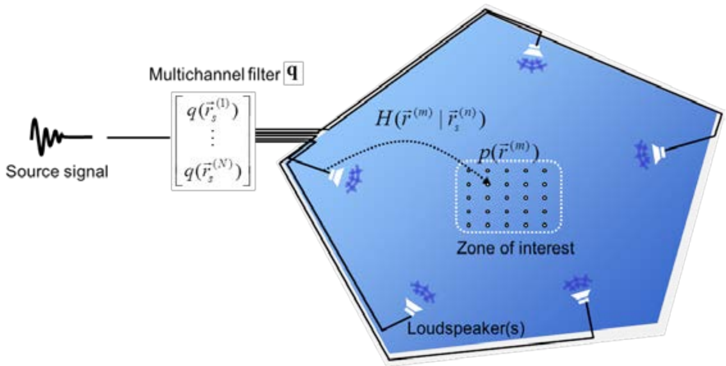


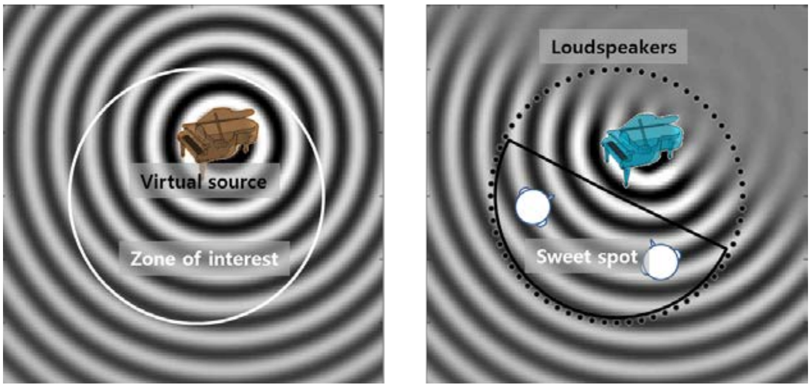
Figure 3 – Structure of a sound field control system

2.2 Sound field reproduction

The concept of sound field reproduction is useful when we try to offer an immersive auditory experience as if the listener is located in another environment. In this scenario, the sensation of the all auditory impressions should be consistent at multiple listener positions; the perceived location of sound sources, perceived stage width, even the sense of ambience should be identical to those experienced in the target sound field. However, both the accurate evaluation and reproduction of multiple auditory impressions at multiple locations is nearly impossible. For this reason, sound field reproduction approaches attempt to replicate the physical wave field using objective evaluation measures(Fig. 4).

The early concept of replicating a sound pressure field from one room to another is described in a study by Camras(1). In this work, recordings from multiple microphones were played back using loudspeakers installed at the same locations as microphones in another room. This principle is somewhat rudimentary compared with the Kirchhoff–Helmholtz integral, which can be used for perfect reproduction irrespective of differences in boundary conditions, but well explains the basic concept of the sound field reproduction problem—matching the target field by superposing secondary fields from multiple loudspeakers.

The aim of this paper, however, is not on the review of the extensive literature(see (2) for example) and long history of sound field reproduction. Rather, it is focused on discussing the limitations of conventional methods as well as possible future approaches, to draw the sound picture we want.



(a) Target sound field (b) Reproduced sound field

Figure 4 – Concept of sound field reproduction

2.2.1 Pressure-matching problem

Suppose that we have a target field to reproduce ($\mathbf{p}_d$), and loudspeakers are driven to minimize the difference or error between the target $\mathbf{p}_d$ and the actual reproduced field $\mathbf{p}$. Then the error minimization problem can be formulated as an optimization problem

$$\underset{\mathbf{q}}{\text{Minimize}} \;\; \varepsilon = \left\|\mathbf{p}_d - \mathbf{p}\right\|^2 = \left\|\mathbf{p}_d - \mathbf{H}\mathbf{q}\right\|^2, \tag{2}$$

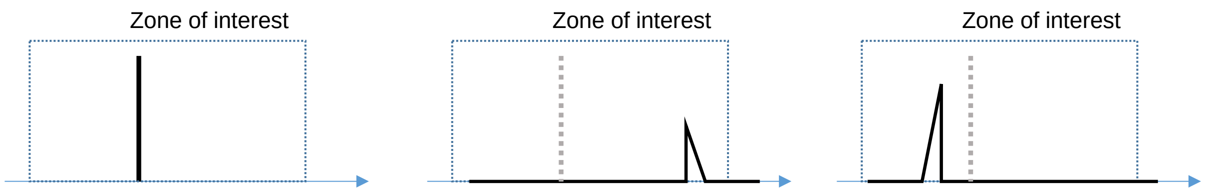
where $\|\;\|$ represents the 2-norm of a vector, and ε is the squared error between the target and reproduced fields. In this case, the reproduction error is represented by mean-squared error(MSE), which becomes the object function of an optimization problem. The final goal of this approach is to find loudspeaker excitation signals $\mathbf{q}$ that minimize the difference between two pressure fields, and hence, the minimum mean-squared-error (MMSE) approach is a typical inverse problem, which can suffer from non-uniqueness, non-existence, and stability issues(e.g.,(3)). This approach requires the target field to be defined before solving the problem, and the non-existence problem can arise when the realization of the target field is physically impossible, as described in detail in Sec. 2.3. The non-uniqueness problem, which implies that we have multiple solutions to produce a target field, is not harmful for the sound field reproduction. It can be problematic, however, when measurement points are limited and the control at those points cannot manifest the reproduction of sound field over a zone of interest. The non-uniqueness and stability problems can be resolved by exploiting an additional constraints or penalties. One popular type of penalty is the input-power penalty, which alters the optimization problem as

$$\mathbf{q}_{opt} = \underset{\mathbf{q}}{\text{argmin}} \; \left\|\mathbf{p}_d - \mathbf{H}\mathbf{q}\right\|^2 + \lambda\left\|\mathbf{q}\right\|^2. \tag{3}$$

This type of problem is often called the least squares problem with Tikhonov regularization(4). The tuning parameter λ controls the relative weight between the MSE and input power, and there exists a clear trade-off between the two. Techniques such as Morozov's discrepancy principle(5)(6), L-curve method(3)(7), or generalized cross-validation(GCV)(8) have been utilized to determine its optimal value.

2.2.2 Error shaping

Mean squared error is a straightforward measure that depends on the size and location of the zone of interest. One simple way to define MSE is to sum up all the errors in the pressure field at points sampling the given zone of interest. In this way, only the total amount of squared error is required to be minimized, and hence, errors at different sampling positions are counted with the same weight. One downside of using MSE as an object function is that the distribution of the error cannot be controlled because of the uniform weight, i.e., it cannot discriminate the “different but useful” fields from the “similar but useless” field. For example, suppose that the target field is given by a spatial impulse, as depicted in Fig. 5(a), and we have two candidate solutions(Fig. 5(b)(c)). Which of these two solutions is better? In the MSE sense, the first solution is the better candidate, but as far as its shape is concerned, the second one might be preferred because the location of pulse is closer to that of the original impulse.



(a) target field (b) solution 1 (c) solution 2

Figure 5– Two candidate solutions; which sound field is better?

Despite its simplicity, MSE does not always reflect what we want. The example case, such as Fig. 5, requires some sort of spatial shaping of the reproduction error to choose the better solution. The shaping of the reproduction error can either be accomplished by adding extra spatial weighting to the object function or by transforming pressure fields into another domain. The former approach, for example, can suppress the reproduction error at the zone center by applying stronger weighting to that area. Nevertheless, defining a proper spatial weighting is a heuristic process requiring much trial-and-error.

In contrast, the transform approach(e.g., (9–11)) resolves this ambiguity by decomposing a sound field into a sum of orthogonal basis functions, each of which has a different spatial contribution. In an ideal situation, where one can use an infinite number of basis functions, the decomposition would not produce any new results, because the total amount of MSE would be preserved before and after the transform. In practice, however, we can control only a finite number of points and basis functions, which can lead to significantly altered reproduction results. To make this statement clear, let us consider the spherical harmonic expansion(SHE) of finite orders:

$$\mathbf{c}_d = \mathbf{T}\mathbf{p}_d \quad , \quad \mathbf{c} = \mathbf{T}\mathbf{p} \,, \tag{4}$$

where $\mathbf{T}$ is the transform matrix corresponding to SHE, and $\mathbf{c}$ denotes the vector of the spherical harmonic coefficients. Without expansion, the MSE defined over a geometrical grid sampling the zone of interest becomes the object function to be minimized. For SHE, we solve the same MSE problem

$$\underset{\mathbf{q}}{\text{Minimize}} \;\; \varepsilon = \left\| \mathbf{c}_d - \mathbf{c} \right\|^2 , \tag{5}$$

but in this case the pressure vectors ($\mathbf{c}_d$ and $\mathbf{c}$) are comprised of spherical harmonic coefficients. For this reason, the transform approach is also called the *mode-matching approach*(10). Since the radial dependence of spherical harmonics, i.e., spherical Hankel or Bessel functions of low orders has more contributions at the coordinate origin than at the outer region, the minimization of the error defined in terms of spherical harmonic coefficients quickly eliminates the reproduction error from the center of the spherical coordinate, leaving most of the error in the outer region.

2.3 Non-existence problem

When defining a target sound field or objective function, the existence of a solution should be checked in advance. The target sound field should be the solution of a wave equation and expressed in terms of the loudspeakers’ transfer functions. Therefore, any arbitrary shape cannot comprise the target field. Although one can find the optimal solution minimizing MSE even when the target sound field is physically unrealizable, in some cases the amount of reproduction errors is consistently high, such that it is meaningless to find a solution with minimal error. For example, consider a problem of reproducing a virtual sound source inside a volume enclosed by loudspeakers(Fig. 6(a)). This problem is a typical non-existence problem, because the sound field from a virtual source inside a volume V ($\vec{r}_{v_0} \in V$) driven by a source signal $q(\vec{r}_{v_0})$ satisfies the inhomogeneous wave equation

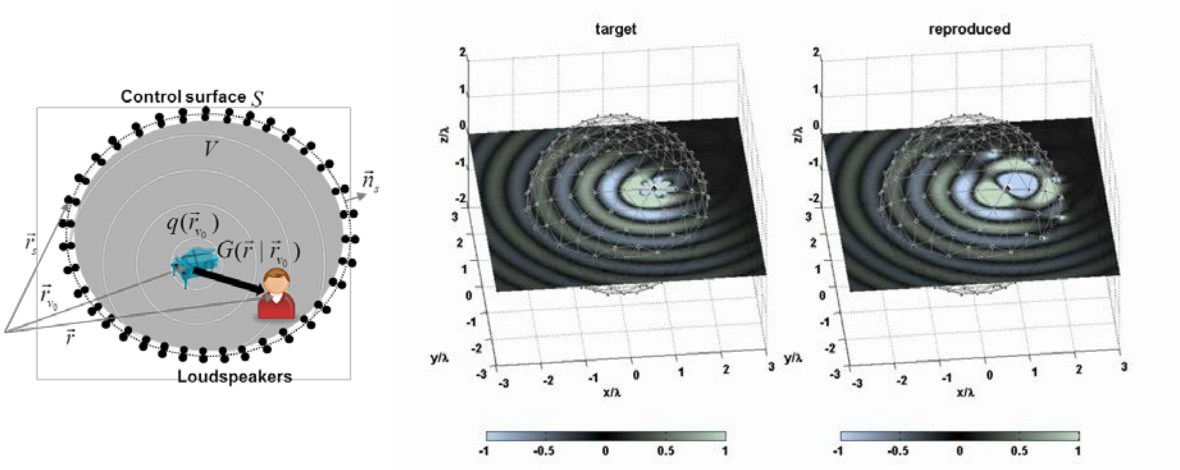
$$\nabla^2 p_d(\vec{r}) + k^2 p_d(\vec{r}) = -q(\vec{r}_{v_0})\delta(\vec{r} - \vec{r}_{v_0}), \tag{6}$$

whereas the sound field reproduced by loudspeakers positioned outside of the volume satisfies the homogeneous wave equation

$$\nabla^2 p_s(\vec{r}) + k^2 p_s(\vec{r}) = 0. \tag{7}$$

Accordingly, it is impossible to generate the singularity of the target field at $\vec{r}_{v_0}$ by superposing sound fields without singularity. If MSE is calculated over the volume V, the reproduced sound field would always yield a reproduction error of infinite magnitude. In this case, the amount of MSE cannot be the measure for searching a better sound field. One may infer that this singularity problem can be avoided by defining a zone of interest to exclude the singular point, but this will create another problem: what kind of shape and size of the zone can avoid the non-existence problem?

To answer these questions, we first need to identify what kinds of sound fields within a volume are reproducible in terms of loudspeakers. Although the Kirchhoff–Helmholtz integral states that there is a direct formula to reproduce an arbitrary sound field from a *virtual source outside* V by the combination of monopole and dipole sources, the KH integral cannot be applied to a virtual source located *inside* the volume.



(a) (b) (c)

Figure 6 – Reproduction of the sound field from a virtual source inside of a loudspeaker array(12)
(a) configuration of virtual sources and 194 loudspeakers (b) (alternative) target field consisting of multipole sources' field (c) reproduced field

This issue, however, has been addressed using several indirect methods. Daniel(13) showed that one can still find a solution from the mode-matching approach with the virtual source inside the volume, albeit with limited zone size. Boone and Verheijen(14,15) derived a formula for a line array by approximating the Rayleigh integral using the stationary phase approximation. Even with these useful formulas, the best achievable zone size and shape for a reproducible sound field is still unclear.

One possible solution involves the elementary functions, i.e., alternative fields, that satisfy the homogeneous differential equation within V but still resemble the target field except near the singular point $\vec{r}_{v_0}$. The alternative field can be found by subtracting the target field from its time reverse(12). If we express the target sound field $p_d(\vec{r})$ using a Green's function($p_d(\vec{r}) = G(\vec{r}\,|\,\vec{r}_{v_0})q(\vec{r}_{v_0})$), then the basic form of the alternative field is given by

$$\begin{aligned} p_a(\vec{r}) &= p(\vec{r}) - p_{tr}(\vec{r}) \\ &= G(\vec{r}\,|\,\vec{r}_{v_0})q(\vec{r}_{v_0}) - G(\vec{r}\,|\,\vec{r}_{v_0})^* q(\vec{r}_{v_0}). \end{aligned} \tag{8}$$

The second term(p_{tr}) of Eq. (8) is denoted as the time-reversed radiation(12), since only the Green's function representing the wave propagation is conjugated without modifying the source signal $q(\vec{r}_{v_0})$. Then, the alternative field p_a simply represents the subtraction of the diverging and converging sound fields. It can be easily shown that this sound field is the solution of a homogeneous wave equation, because the time reversal of the Green's function still satisfies Eq.(6):

$$\nabla^2 p_{tr}(\vec{r}) + k^2 p_{tr}(\vec{r}) = -q(\vec{r}_{v_0})\delta(\vec{r}-\vec{r}_{v_0}), \qquad \vec{r}, \vec{r}_{v_0} \in V. \tag{9}$$

Accordingly, the subtraction of Eq. (9) from Eq. (6) satisfies the homogeneous wave equation.

$$\nabla^2 p_a(\vec{r}) + k^2 p_a(\vec{r}) = 0, \qquad \vec{r} \in V. \tag{10}$$

The integral equation for reproducing distributed virtual sources($q(\vec{r}_{v_0})$) has become known as the Porter–Bojarski integral(16,17) or the generalized holographic-imaging equation(18).

$$\int_S \left[p_{tr}(\vec{r}_s)\frac{\partial G(\vec{r}\,|\,\vec{r}_s)}{\partial n_s} - \frac{\partial p_{tr}(\vec{r}_s)}{\partial n_s} G(\vec{r}\,|\,\vec{r}_s) \right] dS(\vec{r}_s) = \int_V (G(\vec{r}\,|\,\vec{r}_{v_0}) - G(\vec{r}\,|\,\vec{r}_{v_0})^*) q(\vec{r}_{v_0}) dV \tag{11}$$

Although this well-known relation states that one can reproduce converging and diverging waves at the virtual source location, the converging wave has to be minimized to reproduce the sound field, which only radiates outwardly from the source. To tackle this problem, Choi and Kim(12) formulated a combined virtual source distributions($q(\vec{r}_{v_0})$) that can separate the converging and diverging waves in space. Specifically, it was shown that the converging and diverging wave fields have opposite directional patterns in the far-field when a virtual multipole source is reproduced. The directional shape of the converging wavefront is the exact mirror image of the diverging wavefront, and thus, the directivity of the multipole virtual source can be manipulated to separate the regions where the converging and diverging wavefronts dominate(Fig. 6(b, c)). This practice could also provide another source of reproduction error: the directional pattern requiring high-order harmonics moves the sweet spot far from the location of the virtual source(12).

As can be seen from this example, the physical interpretation of a given sound field reproduction problem favors using a great deal of prior knowledge over directly solving with the optimization technique. Without an understanding of the target and reproduced sound fields, the optimal solution calculated from the optimization code is still one of many possible candidates obtained from imperfect constraints or objective functions.

3. Sound field manipulation

The sound field reproduction is a powerful tool, as long as the target field and evaluation measure can be clearly defined. In many situations, however, what we need is not just simple reproduction. Sometimes, we want to have a "controllability" of specific acoustic quantities. For instance, even if we can perfectly replicate the sound field of Carnegie Hall in our home, not everyone would be satisfied with the result. What should we do if the listener wants to modify the sound field according to his or her preferences, such as paintings shown in Fig. 1(c)? This kind of problem, denoted as the *sound manipulation problem*(19) in this article, can only be answered by providing a means to selectively manipulate a desired sound quantity or quality. Although the sound field reproduction can be a useful tool for manipulating a sound field, the reproduction and manipulation problems have completely different goals.

The sound field manipulation problem can be more complex than simple reproduction, as there are many different kinds of sound impressions involved with listening, and only the desired sound impression should be changed without altering anything else. This challenging goal can be approached in either a direct or indirect sense. The direct approach is to find a single variable that only alters the target sound quantity or quality and controls it; the indirect approach selects key acoustic variables associated with multiple sound qualities and then carries out joint optimization to control the relative weight between them. The former is ideal in the mathematical sense, but sound impressions are mutually coupled to each other and it is often impossible to find a single acoustic variable relating a single sound impression. Therefore, it seems realistic to manipulate multiple acoustic variables until we can find the best compromise between multiple sound impressions. To tackle this complex joint optimization problem, we first need to solve the simple problem of optimizing a single sound quantity. In what follows, it is explained how some basic acoustic quantities can be enhanced and suppressed in space simultaneously. As simple examples, research on manipulating acoustic potential energy and planarity is discussed.

3.1 Drawing the acoustic potential energy: personal sound zone

A personal sound zone refers to the zone of focused sound energy in which only the listener can hear loud sound. The purpose of generating a personal sound zone can be either for minimizing the distraction of other listeners(20)(21) or for constructing personal audio systems(22)(23) that allow listeners located in different regions to enjoy different sound content at the same time. The great benefit of producing a personal sound zone is that the user does not need to wear or bring a headset for enjoying personal audio. Upon this premise, personal audio systems(Fig. 7) have been developed for PC monitors(24), mobiles(25), stereo systems(26) and even for car interior sound(e.g., Fig. 9 or (27)).

The same physics as sound field reproduction—the interference of sound fields generated by multiple loudspeakers—is utilized to produce a personal sound zone. Under this principle, various control strategies can be applied; the frequency range of concern, loudspeaker arrangement, and type of loudspeakers are all factors affecting the choice of an adequate control strategy. For example, Druyvestyn and Garas(22) proposed the use of active noise control, beamforming, and loudspeaker directivity, depending on the frequency range of the input signal.

For simple loudspeaker geometries such as linear or circular arrays, a basic method one can utilize is the beamforming technique. By forming a sound beam directed at the listener, a personal sound zone can be generated along a line of sound-beam propagation.In many practical problems, however, such simple array geometries are not possible due to the limitations on the possible loudspeaker positions. Moreover, reflections, diffraction, and scattering of wavefronts in the listening environment can seriously degrade the sound focusing performance of the beamforming system.

A generalized theory that can account for these complex acoustic propagations has been in development since 2002, under the concept of acoustically bright and dark zones(20,21). This concept, in essence, is simply the extension of the zone of quiet in active noise control(ANC), for multiple zones with different characteristics. In the bright zone, the spatial average of acoustic potential energy is controlled to be maximal, while the energy is attenuated within the dark zone. This approach has a completely different object function from that of sound field reproduction, in that there is no target sound field defined. The energy ratio between two different zones, termed acoustic contrast, is then maximized through several optimization techniques(Fig.8).

Figure 7–Various loudspeaker arrays for personal audio systems: (a) TV (b) Monitor (c) Mobile.

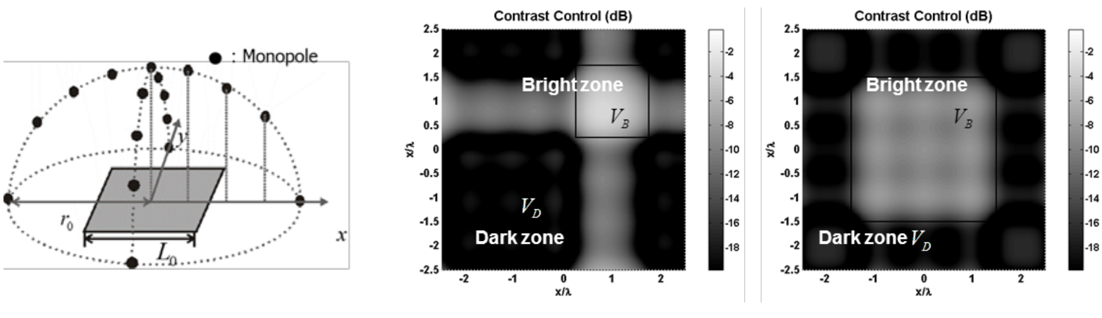


Figure 8– Generation of acoustically bright and dark zones using acoustic contrast optimization(28) (17 monopole sources, $L_0 = 5\lambda$, $r_0 = 20\lambda$)

The optimization technique itself is not especially different from the directivity or array gain optimization techniques in beamforming(29)(30). For example, when the bright zone shrinks to a point, the solution of maximum acoustic contrast is simply given by the solution of maximum array gain in beamforming. The relation between the acoustic contrast optimization and super-directive beamforming for the conventional line array configuration has also been studied(19)(31). Therefore, it should be noted that the "zone control" and "optimization with transfer functions of loudspeakers with arbitrary geometries" are two distinct factors discriminating acoustic contrast control from its predecessors.

To describe the acoustic contrast problem, consider two different zones V_B and V_D. The sound fields in these zones can be denoted by the vectors $\mathbf{p}_B = [p(\vec{r}_B^{(1)}),\cdots,p(\vec{r}_B^{(M_B)})]^T / \sqrt{M_B}$ and $\mathbf{p}_D = [p(\vec{r}_D^{(1)}),\cdots,p(\vec{r}_D^{(M_D)})]^T / \sqrt{M_D}$, where M_B and M_D denote the number of measurement positions in V_B and V_D, respectively. By using the multichannel filter coefficients $\mathbf{q} = [q^{(1)},\cdots,q^{(N)}]^T$ and denoting transfer functions for V_B and V_D as $\mathbf{H}_B$ and $\mathbf{H}_D$, respectively, the pressure vectors can be expanded to

$$\mathbf{p}_B = \mathbf{H}_B\mathbf{q} \ \ , \ \ \mathbf{p}_D = \mathbf{H}_D\mathbf{q}. \tag{12}$$

Acoustic contrast control pursues the maximization of the energy ratio between two different zones V_B and V_D. To incorporate the acoustic potential energy, acoustic contrast control employs the spatial average of acoustic potential energy as a measure representing the energy of a zone. Two energy measures for the bright and dark zones can be described as

$$e_B = \|\mathbf{p}_B\|^2 = \mathbf{q}^H\mathbf{R}_B\mathbf{q}, \quad e_D = \|\mathbf{p}_D\|^2 = \mathbf{q}^H\mathbf{R}_D\mathbf{q}, \tag{13}$$

where each element of the spatial correlation matrix $\mathbf{R} = \mathbf{H}^H\mathbf{H}$ expresses the correlation of transfer functions of different loudspeakers. Then the acoustic contrast is defined as a ratio of these two energies. That is,

$$\alpha = \frac{\|\mathbf{p}_B\|^2}{\|\mathbf{p}_D\|^2} = \frac{\mathbf{q}^H\mathbf{R}_B\mathbf{q}}{\mathbf{q}^H\mathbf{R}_D\mathbf{q}}. \tag{14}$$

It is well-known that the maximization of this Rayleigh quotient form shares a solution with the generalized eigenvalue problem $\mathbf{R}_B\mathbf{q} = \alpha\mathbf{R}_D\mathbf{q}$, which can be solved by iterative techniques such as the QZ algorithm(32). The acoustic contrast can alternatively be defined as

$$\alpha = \frac{\|\mathbf{p}_B\|^2}{\|\mathbf{p}_T\|^2} = \frac{\mathbf{q}^H\mathbf{R}_B\mathbf{q}}{\mathbf{q}^H\mathbf{R}_T\mathbf{q}}, \tag{15}$$

where $\mathbf{q}^H\mathbf{R}_T\mathbf{q} = \mathbf{p}_T^H\mathbf{p}$ is the energy over the total zone of interest calculated from the pressure vector $\mathbf{p}_T = [\sqrt{M_B}\mathbf{p}_B^T \ \sqrt{M_D}\mathbf{p}_D^T]^T / \sqrt{M_B + M_D}$. It has been shown that both forms have the same optimal solution(20).

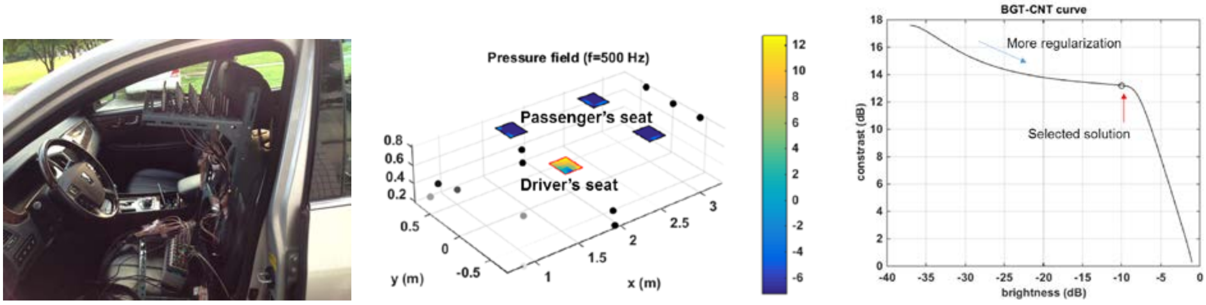
Depending on the constraints of a given system, the acoustic contrast problem can be modified into various forms. For example, when the singularity of matrix $\mathbf{R}_D$ is of concern, the acoustic contrast problem can be modified into a regularized form:

$$\tilde{\alpha} = \frac{\mathbf{q}^H\mathbf{R}_B\mathbf{q}}{\mathbf{q}^H(\mathbf{R}_D + \gamma\mathbf{I})\mathbf{q}} = \frac{\mathbf{q}^H\mathbf{R}_B\mathbf{q}}{\mathbf{q}^H\mathbf{R}_{Dr}\mathbf{q}}. \tag{16}$$

The regularization parameter γ prevents the abrupt increase of the acoustic contrast near the null space of $\mathbf{R}_D$, by incorporating the total power of the multichannel filter coefficient $\mathbf{q}^H\mathbf{q}$. This concept stems from the input power penalty of the MMSE problem(Eq. (3)) and denoted as an acoustic brightness penalty in the zone control problem. Similarly to the acoustic contrast of Eq.(14), the acoustic brightness is defined as a ratio of the acoustic potential energy of the bright zone to the total power of the filters:

$$\beta = \frac{\mathbf{q}^H\mathbf{R}_B\mathbf{q}}{\mathbf{q}^H\mathbf{q}}. \tag{17}$$

Therefore, the modified form of Eq. (16) can be viewed as a hybrid form of the acoustic contrast and brightness problem. The regularization parameter γ plays a similar role as in the regularized least squares problem of Eq. (3), so the optimal parameter γ is determined from the acoustic brightness-to-contrast curve analogous to the L-curve(Fig. 9(c)).



(a) (b) (c)

Figure 9 – Acoustic contrast optimization for the sound field manipulation in a car cabin. (a) Transfer function measurement using a microphone array (b) Sound field focused for the driver's seat (c) Brightness-Contrast curve (at 500Hz)

To illustrate this mathematical similarity, let us reformulate the general acoustic contrast problem as the following optimization problem: we want to find an excitation signal $\mathbf{q}_{opt}$ that maximizes the acoustic potential energy of the bright zone(V_B) while that of the dark zone(V_D) is constrained to a constant value e_D. This statement can be written in a mathematical form as

$$\mathbf{q}_{opt} = \arg\max_{\mathbf{q}} \mathcal{L}(\mathbf{q}), \quad \text{where } \mathcal{L} = \|\mathbf{p}_B\|^2 - \alpha(\|\mathbf{p}_D\|^2 - e_D) - \beta\|\mathbf{q}\|^2 . \tag{18}$$

The difference of this optimization problem and the MMSE problem is that the squared error of Eq. (3) is replaced with the energy difference $\|\mathbf{p}_B\|^2 - \alpha(\|\mathbf{p}_D\|^2 - e_D)$ and the minimization problem is converted to the maximization problem with the negative input power penalty $-\beta\|\mathbf{q}\|^2$.

The variable α is a Lagrange multiplier to describe the dark zone constraint as a penalty function, and β is a kind of tuning parameter to prevent the divergence of input power $\|\mathbf{q}\|^2$. By taking the derivative of $\mathcal{L}$ with respect to $\mathbf{q}$ and α, the optimal solution can be found at

$$\frac{\partial\mathcal{L}}{\partial\mathbf{q}} = 2\left(\mathbf{R}_B\mathbf{q} - \alpha(\mathbf{R}_D + \gamma\mathbf{I})\mathbf{q}\right) = 0, \quad \frac{\partial\mathcal{L}}{\partial\alpha} = \mathbf{q}^H\mathbf{R}_D\mathbf{q} - e_D = 0, \tag{19}$$

where $\gamma = \beta/\alpha$. Accordingly, the problem defined in Eq. (18) is equivalent to the eigenvalue problem given by

$$\mathbf{R}_B\mathbf{q} = \alpha(\mathbf{R}_D + \gamma\mathbf{I})\mathbf{q} \qquad \text{subject to } \mathbf{q}^H\mathbf{R}_D\mathbf{q} = e_D . \tag{20}$$

From Eq.(20), it can be seen that the optimal solution $\mathbf{q}$ is one of the eigenvectors of the generalized eigenvalue problem. The resultant value of the objective function $\mathcal{L} = \alpha e_D$ is maximized when the eigenvalue α is maximum, so $\mathbf{q}_{opt}$ is the eigenvector corresponding to the maximum eigenvalue, scaled such that $\mathbf{q}^H\mathbf{R}_D\mathbf{q} = e_D$. In practical applications, the optimal solution can be scaled differently to yield a constant output energy in the bright zone or constant pressure at a reference position. Otherwise, one can take β as the Lagrangian and tune the variable α to a fixed value. That is,

$$\mathbf{q}_{opt} = \arg\max_{\mathbf{q}} \left[\|\mathbf{p}_B\|^2 - \alpha\|\mathbf{p}_D\|^2 + \beta(J - \|\mathbf{q}\|^2)\right], \tag{21}$$

which is called the energy difference maximization(33). The solutions of these two methods are on the same acoustic contrast–brightness curve, as their objective functions are only different in the choice of the tuning factor.

Although acoustic contrast problems can be approached by simple eigenvalue analysis, there are a couple of drawbacks with this simple object function. First, the acoustic contrast only maximizes the averaged potential energy and cannot control the precise energy distribution. In many practical applications, this drawback does not seriously degrade the usefulness of the method in the low-frequency region where the size of the wavelength is comparable to or larger than the size of the zone. However, in the mid-to-high frequency region, there can be abrupt amplification or suppression of the acoustic energy within the zone of interest. Such a problem can be tackled by setting up a different object function for the high frequency region(27) or by introducing a hybrid object function that combines the acoustic contrast with another measure for reducing sound field variations(34,35).

Examples of combining acoustic contrast with other object functions can be found in various works. Since 2005, hybrid forms of the acoustic contrast and brightness have been proposed(19,23,28,31). Møller *et al.*(36) introduced a joint optimization technique that can achieve sound field reproduction and sound isolation at the same time. To this end, a target sound field $\mathbf{p}_d$ was defined over a bright zone as a plane wave propagating in a single direction, and the weighted sum of the dark zone energy and the reproduction error was minimized:

$$\mathbf{q}_{opt} = \arg\min_{\mathbf{q}} \left[\alpha \|\mathbf{p}_D\|^2 - \|\mathbf{p}_B\|^2 + \zeta \|\mathbf{p}_d - \mathbf{H}_B\mathbf{q}\|^2) \right]. \tag{22}$$

The first two terms of the objective function are simply equivalent to the energy difference of Eq. (18), but inverted in sign. In this hybrid form, therefore, both the negative energy difference and MSE to the target sound field are minimized simultaneously. It is noteworthy that both parameters α and ζ cannot be the Lagrangian multiplier, in cases where the problem is over-determined due to the constraints on both the dark zone energy and the MSE. Accordingly, a proper combination of tuning parameters α, ζ must be found manually to obtain the desired level of acoustic contrast. The search process can be simplified by defining quadratic constraints and solving the QCQP problem, as described in the work of Betleham(37).

The generation of a personal sound zone can also be accomplished by pressure-matching technique. If we consider a target sound field that has a zero pressure field over a dark zone and $\mathbf{p}_d$ over a bright zone, the MMSE problem can be written simply as

$$\begin{aligned} \mathbf{q}_{opt} &= \arg\min_{\mathbf{q}} \left[\|\mathbf{0} - \mathbf{H}_D\mathbf{q}\|^2 + \|\mathbf{p}_d - \mathbf{H}_B\mathbf{q}\|^2) \right] \\ &= \arg\min_{\mathbf{q}} \left[\|\mathbf{p}_D\|^2 + \|\mathbf{p}_d - \mathbf{H}_B\mathbf{q}\|^2) \right]. \end{aligned} \tag{23}$$

From the comparison of Eq. (23) to Eqns. (3) and (22), it can be seen that the differences between these MMSE optimizations are in their penalty functions. More specifically, the energy difference penalty of Eq. (22) can be used to give more weight to the contribution of the dark zone or bright zone energy, whereas Eq. (23) always puts equal weight to the MSE and the dark zone energy.

As in the pressure matching problem, however, the MMSE-based object function also suffers from the error-shaping problem. The same kinds of error-shaping techniques as discussed in Sec. 2.2.2 can be applied for better error shaping. Wu and Abhayapala(38,39) used the spherical or circular harmonics expansion with translation operators to produce spherical or circular sound zones using MMSE approach. Jacobsen *et al.*(40) compared the optimizations of acoustic contrast and MSE, in which the acoustic contrast optimization requires fewer loudspeakers to achieve higher energy difference, but the variation of magnitude and phase within the bright zone cannot be controlled. This conclusion also highlights the need for a more precisely defined objective function to realize the desired sound field.

By introducing a more complex penalty function, one can define various optimization problems to obtain more desirable solutions. Even with new methods and object functions, acoustic contrast is still being used for finding the upper bound of the energy ratio for a given array and zone configuration.

3.2 Shaping sound fields by focusing

The benefit of using acoustic contrast as an objective function is that sound energy can be concentrated over a finite region of interest rather than a point. If we extend the concept of "finite region" into another domain, acoustic contrast optimization can be used to control various acoustic quantities, such as mean active intensity(41), or propagating direction of a wave front(42,43). The basic idea behind this is to enhance the energy of a certain group of sound field components or basis functions while suppressing others.

One interesting extension of the acoustic contrast optimization is the plane wave control technique, known as wavenumber domain focusing. What we want to accomplish with this approach is to produce a plane wave propagating in a desired direction. This objective can be, of course, accomplished by using the pressure matching technique described in Sec. 2.2.1, but the regional focusing concept of the acoustic contrast optimization brings more flexibility in finding the desired sound field.

In particular, the control of propagating direction is possible by transforming a sound field into another domain. In the wavenumber-domain focusing technique, the focusing of sound energy is carried out over a finite region in a wavenumber domain(Fig. 10(a)). To this end, the pressure vector $\mathbf{p}$ of a spatial region V is converted into a wavenumber spectrum vector $\mathbf{w}$ by taking the spatial Fourier transform:

$$\mathbf{w} = \mathbf{F}\mathbf{p}, \tag{24}$$

where the spatial Fourier transform matrix $\mathbf{F}$ is defined by $[\mathbf{F}]_{(\ell,m)} = \exp(i\vec{k}^{(\ell)} \cdot \vec{r}^{(m)})$, and the wavenumber position vectors $\vec{k}^{(\ell)}$ sample the zone of interest V_w in the wavenumber domain($\vec{k}^{(\ell)} \in V_w$). As a consequence, one can set up an acoustic contrast optimization problem by considering pressure vectors $\mathbf{w}_B$ and $\mathbf{w}_D$ for the bright zone(V_{wB}) or dark zone(V_{wD}) defined in the wavenumber domain and corresponding Fourier matrices $\mathbf{F}_B$ and $\mathbf{F}_D$, such that

$$\mathbf{w}_B = \mathbf{F}_B\mathbf{p}, \quad \mathbf{w}_D = \mathbf{F}_D\mathbf{p}. \tag{25}$$

The acoustic potential energy of V_{wB} in the wavenumber domain can then be described as

$$\begin{aligned} e_{kB} &= \|\mathbf{w}_B\|^2 \\ &= \mathbf{q}^H\mathbf{R}_{wB}\mathbf{q} \end{aligned} \quad \text{and} \quad \begin{aligned} e_{kB} &= \|\mathbf{w}_D\|^2 \\ &= \mathbf{q}^H\mathbf{R}_{wD}\mathbf{q}, \end{aligned} \tag{26}$$

where $\mathbf{R}_{wB} = \mathbf{H}^H\mathbf{F}_B^H\mathbf{F}_B\mathbf{H}$ expresses the spatial correlation within the wavenumber zone V_{wB}. This formulation in the wavenumber domain allows us to apply all the techniques discussed in Sec. 3.1 to focus energy onto V_{wB}. Focusing in the wavenumber domain makes most of the energy concentrated into a group of plane waves propagating in a single direction, while the others to be reduced. As a result, the focusing technique can shape the propagating direction of a sound field(Fig. 10).

The fundamental difference between the reproduction and manipulation approaches can also be observed in the wavenumber-domain focusing problem. If the focusing area in the wavenumber domain is given by a point, there would be no difference between two separate approaches(28). As the focusing area is enlarged, however, the wavenumber domain focusing attempts to enhance the energy of the *group of plane waves* propagating in similar directions, instead of a single plane wave. This transform approach enables the selective enhancement of acoustic quantities over a group of sound field components and has many advantages as compared to the MMSE-based reproduction. In 2010, Chang *et al.*(43) used a simple focusing technique(time-reversal) in the wavenumber domain for reducing the computational complexity of a plane wave generation problem. Jackson *et al.*(44) further generalized the wavenumber domain representation using a planarity measure, which represents the degree to which the wavenumber spectrum is focused in a single direction. Specifically, the *planarity* is the ratio of wavenumber spectrums weighted by a cosine function:

$$planarity_A = \frac{\|\mathbf{D}\mathbf{w}\|^2}{\|\mathbf{w}\|^2}, \; [\mathbf{D}]_{(\ell,\ell')} = \begin{cases} \sqrt{\cos\theta^{(\ell)}} & \text{if } \ell = \ell' \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where $\theta^{(\ell)}$ in the diagonal of the weighting matrix represents the angle of the propagating direction of the ℓ^{th} wavenumber component(plane wave) from that of the plane wave of the greatest amplitude. If the diagonal elements of the weighting matrix $\mathbf{D}$ consist of ones or zeroes such that $\|\mathbf{D}\mathbf{w}\| = \|\mathbf{w}_B\|$, the planarity would simply be equivalent to the acoustic contrast of Eq. (15) in wavenumber domain. Therefore, it is the weighting matrix $\mathbf{D}$ that provides more flexibility than the simple geometric selection of a bright zone. Planarity was originally used as a means to evaluate how much the controlled sound field resembles a plane wave, and later it has been extended as a part of the objective function. For stereo reproduction within the personal sound zone, Coleman *et al.*(34) modified the cosine weighting into a general spatial window and used planarity to create two virtual plane waves with high acoustic contrast.

Wavenumber domain focusing and planarity control have demonstrated the potential of sound field shaping via spatial transform and selective focusing. In order to produce more precise or complex shape of sound field, one may attempt to utilize other types of transform or basis function. In developing new techniques with different transforms, the selection of appropriate basis functions will be determined by the ability to discriminate a group of desired shapes from others.

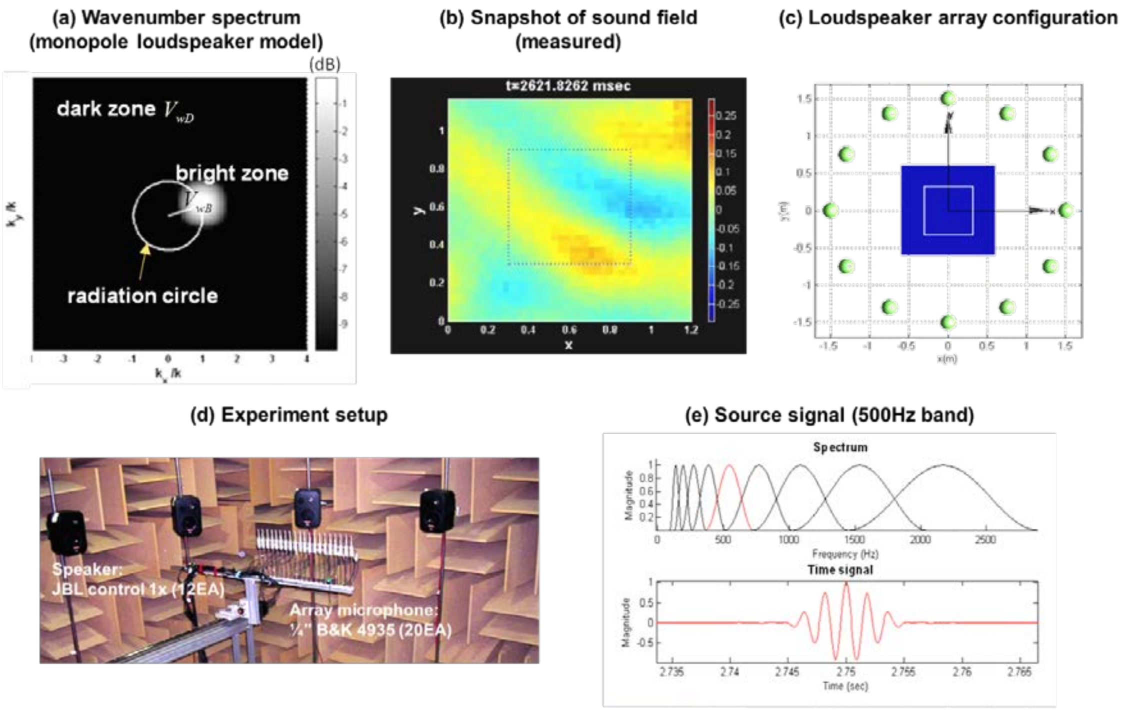


Figure 10 – Example of wavenumber domain focusing: free-field simulation and experiment(28). (a) wavenumber spectrum after focusing (predicted from monopole model) (b) measured sound field from real loudspeakers (c)~(d) array configuration with 12 loudspeakers (e) source signal

3.3 Filter design issues

The multichannel filter design discussed thus far assumes single-frequency excitation. In practice, music or speech signals fed into a multichannel filter have a wideband and time-varying spectrum, and hence, the filter coefficients $\mathbf{q}$, determined in the frequency domain, can produce a lot of problems. A representative example is the causality problem. The inverse Fourier transform of the filter coefficients designed at discrete frequencies with interval $\Delta\omega$ work properly only with cyclic convolution, which yields the causality issue in the real systems producing sound through the linear convolution. This issue is similar to what has been extensively discussed for the frequency domain design of sound field reproduction filters(45–47), and can be circumvented using similar techniques. In particular, the input power penalty of Eqns. (3) and (18) acts as a regularization that prevents the excessive amplification of filter coefficients.

Basically, relative weights between filter coefficients of different frequencies need to be aligned in both magnitude and phase, to ensure the minimal distortion of temporal pressure signals at multiple listening positions. However, spatial optimization techniques, such as acoustic contrast optimization, predetermine the spatial distribution of sound, so the modification of temporal characteristics over multiple positions is inevitably limited. For the alignment of multiple frequency filter coefficients, Choi and Kim(28)(48) used a single-channel post-filter to recover impulse invariance at a single listener position within the bright zone. Although it can reduce temporal artifacts to some extent by rearranging multichannel filters designed in frequency domain, the post-filter cannot fully resolve the causality problem in principle. A time domain optimization technique(49) proposed by Cheer and Elliott controls the time-average of the acoustic potential energy and hence can avoid the causality issue. Nevertheless, owing to the fact that the time averaged potential energy has no consideration of its spectral distribution, time domain optimization often leads to non-equalized responses in frequency domain. To minimize the pressure variation over the bright zone, Cai(50) introduced a differential constraint to the frequency domain formula. Although none of these techniques can perfectly control both the temporal and spatial distribution of the resultant sound field, this can be understood as the downside of the sound field manipulation approach, which enhances some acoustic quantities with less constraints on the magnitude and phase distribution in space and time.

3.4 Consideration of non-ideal conditions

The multichannel filter constructed from computer simulation yields several problems with the loudspeaker array implemented in reality. Manufacturing variance of loudspeakers produces mismatch in loudspeaker gains and phases, and their positioning errors can significantly degrade the performance predicted in the simulation. The robustness of a personal audio system is a critical issue that cannot be overlooked in the design of array hardware and multichannel filters.

Signal-processing techniques for obtaining a multichannel filter robust to the change of acoustic transfer functions have been discussed at length in the literature(23)(31)(35)(51)(52). Apart from signal-processing issues, it is also important to find what kinds of hardware design parameters are sensitive enough to cause changes in performance. To investigate performance degradation due to various implementation errors, Park *et al.*(53) formulated acoustic contrast sensitivity against the first order perturbation of loudspeaker gain, phase, and position. It was shown that the partial covariance of the pressure fields in the bright and dark zones can be used to predict the acoustic contrast change caused by the perturbation of loudspeaker gain and phase, and the reactive intensity can estimate the acoustic contrast change due to the loudspeaker position mismatch. Based on this analysis, an *acoustic contrast sensitivity map* was proposed to guide the robust design of personal audio systems. The map illustrates, for example, which loudspeaker in an array is more susceptible to gain or positioning error(Fig. 11) and should be more carefully manufactured in the array-design stage.

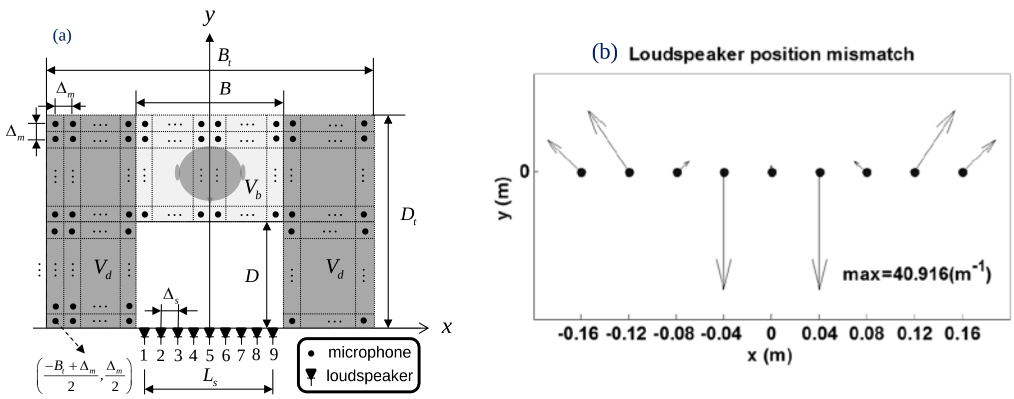


Figure 11 – Example of acoustic contrast sensitivity map

(a) A personal audio system with a bright zone configured near the user location ($B_t = 0.8\text{m}$, $B = 0.4\text{m}$, $D_t = 0.6\text{m}$, $D = 0.2\text{m}$ and $\Delta = 0.02\text{m}$) and control loudspeakers ($L_s = 0.32\text{m}$, $\Delta_s = 0.04\text{m}$).

(b) Contrast sensitivity map (at 3 kHz) for the loudspeaker position mismatch.

4. Paradigm for array sound systems of the future: Sound of Things

So far, various array systems and promising applications that can be realized by employing multiple loudspeakers and sensors have been discussed. As the number of loudspeakers increases, however, the installation and wiring of loudspeakers will be a substantial hurdle in implementing a practical system. Even if the wiring issue can be mitigated by using a wireless network, the construction of an array system is still challenging, since we have a limited installation space and we require information on the exact positions of loudspeakers to synthesize a sound field as desired.

Then, how can we naturally construct and integrate an array system in various environments? Imagine that objects or gadgets used in our ordinary life become smart devices that can produce and measure the sound, identify their own locations, and are connected to each other wirelessly through a unified network(Fig. 12). The floor lamp, coffee table, and light bulbs in a living room can cooperate to act as an active noise control system; a vase, mug, or even window can work together to produce the 3D sound around us. These smart, networked, cooperative objects producing sound can realize the sound field we dreamed of producing from loudspeaker array systems. The concept of a unified sound system implemented by combining smart sounding objects via a unified sound network will be termed as the Sound of Things(SoT), from an analogy to the paradigm: the Internet of Things(IoT)(54,55). The key requirements for such a sound system and sounding objects can be summarized as follows:

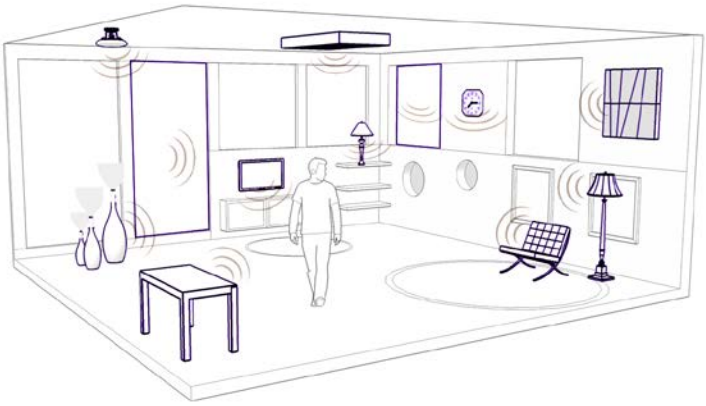


Figure 12 – Illustration of the concept: Sound of Things.
(Smart sounding objects acting as a single loudspeaker array system)

1. Location awareness

Each device should be able to detect its location with respect to the others. The information of the whole audio system is needed to implement a flexible-layout sound system that can adapt to various geometric configuration of sounding objects. The locations can be identified by various means and sensors, but also can be measured acoustically through the exchange of sound signals between devices.

2. Environment monitoring

Sensing the acoustic information of the environment is also an important issue in performing adaptive sound field reproduction. The acoustic transfer function of each device can be directly utilized for generating a target sound field, but more advanced systems might be able to extract essential parameters such as the location of reflecting boundaries, absorption coefficients, as well as the position of the listener.

3. Accurate wireless synchronization

Multiple devices working as an array device need to be strictly synchronized. In sound reproduction systems, the time of arrival at the location of listener plays a crucial role in localizing the direction of a sound source. The human auditory system identifies the time difference of arrival at two ear signals to find the location of a sound source, and the inter-aural time difference(ITD) changes between ±1 ms range depending on the direction of sound arrival. Therefore, the clock of each device should be synchronized to within a few microseconds accuracy. This is no problem with a wired network, but wirelessly connected devices have different clock sources, which result in the misalignment of time on the order of milliseconds(Fig. 13(a)).

Each sounding object has to be a smart and independent device that can fulfill these requirements. Although there can be several issues involved with these requirements, most of them can be resolved by well-known technologies developed for noise and vibration control problems. For example, the identification of multiple devices and listeners can be accomplished by source localization techniques like time-difference-of-arrival(TDOA)(56,57), and for environment monitoring, one can utilize the on-line secondary-path identification technique(58,59) of the active noise control system. The major missing links to realizing the SoT might be the accurate synchronization between multiple objects and an adaptive rendering technique that can cope with flexible layouts of sounding objects.

In general, accurate time synchronization between multiple sounding objects can be accomplished by embedding time stamps in the communication signals or by utilizing Global Positioning System(GPS) signals, but the synchronization through the network time protocol(NTP)(60) is not accurate enough(on the order of tens of milliseconds) when there are non-identical propagation delays between multiple sounding objects, and synchronization via GPS is not applicable for the indoor situations.

One viable option is to synchronize using acoustic signals, which can secure tight synchronization on a time scale of several tens of microseconds at 44.1~96 kHz sampling rates. However, the propagation delay between sounding objects would also be unknown, so the synchronization would have to be made with the propagation delay or distance estimation. Researchers such as Atkinson and Blank(61,62) developed acoustic delay estimation methods for multiple audio systems synchronized via an IP-based network. More recently, BeepBeep(63) introduced the first use of acoustic signals for object-to-object distance measurement with wireless synchronization.

The use of acoustic signals also makes it possible to compensate non-identical radiation characteristics of multiple sounding objects. In a pilot project at KAIST, called Mobile Maestro(64), this functionality was tested with multiple mobile phones. The purpose of this project was to realize an adaptive, flexible-layout audio coordination system using multiple mobile phones. Non-identical mobile phones were arbitrarily distributed in space, and then controlled to synchronize their clocks, identify the layout, and equalize early acoustic responses.

The mobile phone is an ideal platform for SoT, because it has a loudspeaker producing sound, embedded microphones capturing sound, wireless connections, and a processing unit required for de-centralized signal processing. The technology, termed as the adaptive mobile audio coordination(AMAC; Fig. 14), connects multiple mobile devices using a unified sound network through Android APIs(Sound Pool) designed to identify locations of individual mobile phones and make tight synchronization for multichannel sound reproduction. At the initial connection stage, each mobile phone emits and receives maximum length sequence(MLS) signals for synchronizing and measuring acoustic transfer functions simultaneously. This is possible through the bi-directional transmission and reception of acoustic signals with the microphones embedded in the phones.

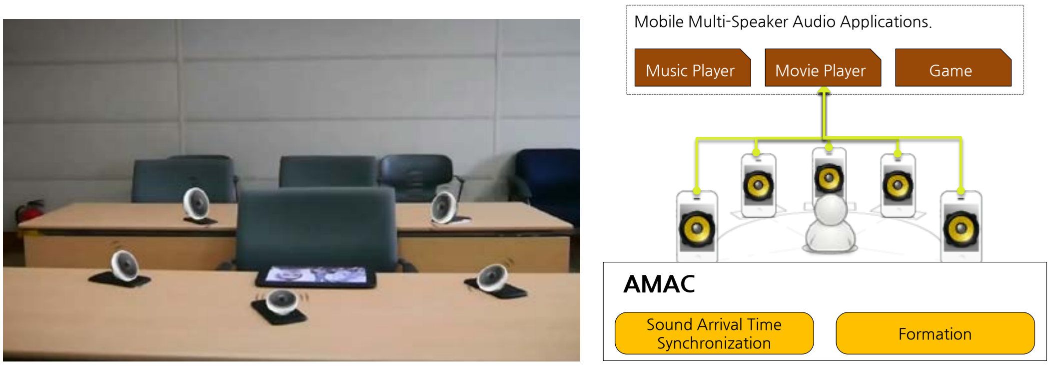
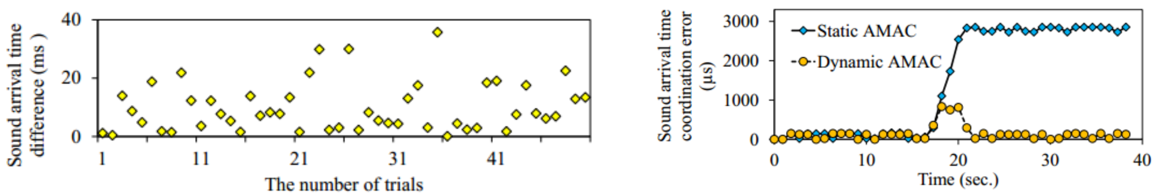


Figure 13 – Overview of the adaptive mobile audio coordination; multiple mobile devices arbitrarily positioned in space are controlled to identify their own locations and to produce the 5.1 channel sound.



(a) (b)

Figure 14 –Synchronization between multiple mobile phones (before and after the audio coordination). (a) Sound arrival time difference on the Android mobile phone (b) Sound arrival time estimation accuracy for dynamically changing loudspeaker layouts

After an initial stage, the change in acoustic environments and speaker coordinates is traced during audio playback using inaudible, frequency-modulated high frequency signals. This is also to prevent the continuous clock-drift during the playback and to adapt to the change in layout in a continuous and seamless fashion without distracting users(Fig. 14(b)).

The adaptive rendering of multiple objects in dynamic layouts is another important field of study. At the current stage, only dynamic equalization is tested, but this work is a starting point for a number of research directions. With the continuous tracking of a dynamic layout, various adaptive rendering techniques, such as automatic sound equalization(65), adaptive HRTF-based rendering(66), or Ambisonics rendering for flexible layouts(67) can be employed.

The SoT paradigm can deliver a rich, immersive listening experience in a natural way. This entails the construction of smart devices that can perform acoustic sensing, calibration, and synchronization continuously and seamlessly. This paradigm may offer great benefits in installation and operation, compared to conventional arrays with non-flexible configuration.

5. Sound sketching interface

The user interface(UI) for a spatial sound rendering system could be beyond the scope of the sound field control, but in practice, it is an indispensable part of the sound rendering system for authoring spatial sound content and promoting the use of new spatial sound rendering technologies. The authoring tool should be intuitive, natural, and more importantly, be able to express newly developed auditory illusions in a straightforward manner. Nevertheless, most of the conventional UIs are based on the track or timeline concept, which is designed for editing the temporal variation of audio files. When both the spatial and temporal characteristics of sound fields are concerned, the current interface has limitations in describing spatial characteristics of sound field. The major problem is that we need to handle acoustic quantities changing in a 4D space(3D for space, 1D for time) using a 2D screen-based interface.

There have been many attempts to manipulate spatial sound impressions more easily. Researchers at Fraunhofer IDMT(68) and TU Berlin(69) have combined the track-based interface with a 2D or 3D geometry navigation panel. In the interface of (68), the tracks are organized to include the 3D position data, as well as various sound quality data with respect to time. Melchior(70) has demonstrated a 3D joystick navigation interface with the haptic feedback, and nowadays many gesture-based editing environments with the Microsoft Kinect sensor are being introduced(e.g., Fig. 15). These are all good examples of offering spatial and temporal navigation simultaneously.

The SoundSketch project on-going at KAIST(Fig. 16) is also a part of the effort to design an interface suitable for editing spatial sound effects. It attempts to resolve the problem of editing in space and time domain by using a 2D sketching interface. The heart of the 2D sketch interface for drawing a 3D curve follows the mechanism of the “I love sketch” interface, designed by Bae(71) in 2007. In the sound sketch interface, a 2D curve is drawn first on a 2D canvas(tablet) with a pen, and then, the user changes the viewpoint by dragging the frame edge across the canvas(Fig. 16(a)~(c)). If the user draws another curve from the new viewpoint, these two curves remove the ambiguity of the first 2D curve and confirm the 3D path of a sound object.

Through 2D-sketch based interfaces, the position, perceived width, as well as reverberation and equalization can be actually “sketched” in 3D space and time. The speed of movement over time, for example, can be controlled by sketching with a speed-overriding brush. The sketch interface is connected to a 192-channel loudspeaker array(Fig. 16(d)) through the open sound control(OSC) protocol(72) to render virtual sound sources and ambient sounds in 3D space. The array system consist of 6 line arrays each of which has 32 loudspeaker units. Each line array produces a virtual sound source, and 6 virtual sources produced by the line arrays are controlled to create an auditory illusion of a single sound source at the listener location(Fig. 16(e)).

These intuitive interfaces can also contribute to the interaction between the spatial sound rendering system and the general user. An array system combined with the user tracking device can produce a virtual sound source following the user, irrespective of his/her position(Fig. 15(b)). Implementing the essence of a sketching interface on consumer electronic devices, such as tablets or mobile phones, will enable people to interact and play with 3D sound objects.

(a) (b)

Figure 15 – Demonstration of spatial sound control using gesture-based interface with user tracking
(a) positioning with hand gesture (b) virtual sound source following the user’s head position

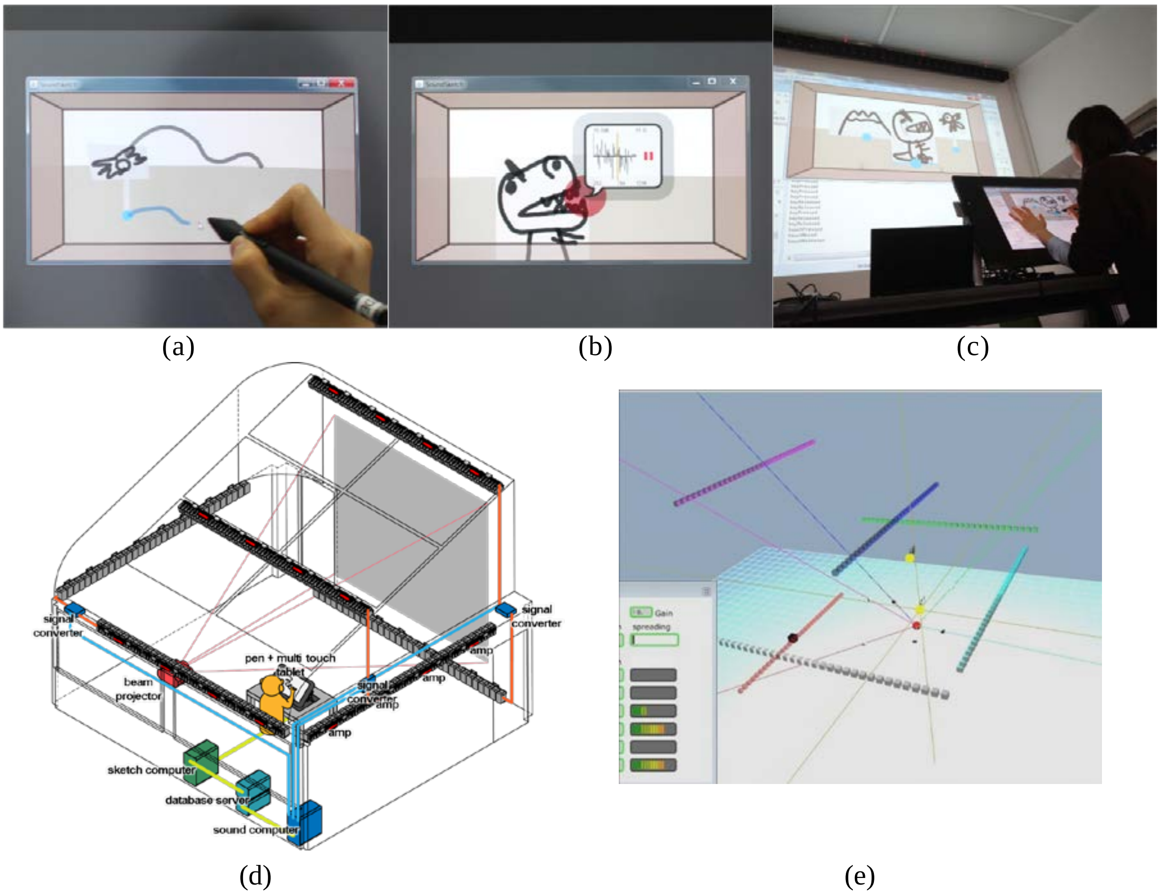


Figure 16 – SoundSketch interface (a) Path and velocity generation (b) Sound clip mapping (c) Interface overview (d) Configuration of the 192 channel loudspeaker array system (e) Sound renderer for multiple virtual sources

6. Summary

Theories and implementation examples of generating a sound field in a desired shape were introduced. Shaping sound fields by multiple loudspeakers is a classical problem, in that linear superposition of multiple sound fields is utilized, but it requires careful design of target field, object functions, and configuration of constraints to obtain the sound field we really want to have.

It was explained that the non-existence issue of the sound field reproduction can be problematic if the defined target sound field is not physically realizable. The sound field manipulation approach is less susceptible to this problem, because it aims at the creation of sound field without any pre-defined target field. The principles and concepts of sound field manipulation techniques, such as acoustic contrast optimization, wavenumber domain focusing, planarity control with regularization, were explained in a unified way. All these techniques can be expressed as optimization problems with different types of object functions and constraints. However, the manipulation approaches only concern the control of sound quantities included in object functions and constraints, which can lead to abrupt changes or degradation of other sound quantities that are not considered in the problem. To control various aspects of the sound field, the sound manipulation approach is evolving to new joint optimization techniques incorporating multiple object functions and constraints.

Practical issues on the construction of array systems and its possible future direction were presented. A paradigm(SoT) to incorporate networked sound objects as elements of an array system is introduced, and technologies for realizing the concept were addressed. As a pilot study, the concept was implemented on mobile phones with built-in loudspeakers and microphones. Individual mobile phones were tightly synchronized in time, and their relative locations, as well as acoustic transfer functions, were identified using acoustic signal transmissions. From the identified information, the mobile phones were controlled to serve as a unified 5.1 channel audio system.

Finally, the user-interface for the sound field control system was discussed. For the intuitive interaction with sound field control systems, the SoundSketch interface that can actually 'sketch' the desired sound quantity in space and time was demonstrated.

All the sound manipulation theories, array control technologies, and user-interfaces are essential parts required to precisely sketch the sound field we want to hear. The ultimate goal of sound sketch, however, would only be realized when we can precisely and separately control specific sound qualities or subjective measures of interest. Although current technologies based on objective measures are far from ideal, continuous efforts to discover the relation between the subjective and objective measures will make it possible to draw our own sound pictures.

Acknowledgement

This work was supported by supported by the Samsung Research Funding Center of Samsung Electronics under Project Number SRFC-IT1301-04, the Ministry of Trade, Industry and Energy (MOTIE) grant funded by the Korea government (No. 10037244), and the BK21 project initiated by the Ministry of Education.

The author acknowledges Prof. Yang-Hann Kim(Dept. of Mech. Eng., KAIST), Mr. Hyosu Kim, Prof. Insik Shin and Prof. Junhwa Song(Dept. of Comp. Sci., KAIST), and Prof. S. H. Bae(Dept. of Industrial Design, KAIST) for their invaluable contributions to *Sound Ball*, *Mobile Maestro*, and *SoundSketch* projects.

References

1. Camras M. Approach to Recreating a Sound Field. J Acoust Soc Am. 1967;41(6):1590.
2. Spors S, Wierstorf H, Raake A, Melchior F, Frank M, Zotter F. Spatial Sound with Loudspeakers and Its Perception : A Review of the Current State. Proc IEEE. 2013;101(9):1920–38.
3. Hansen PC. Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM monographs on mathematical modeling and computation; 1998.
4. Tikhonov AN, Arsenin VY. Solutions of Ill-Posed Problems. Math Comput. 1978;32:1320–2.
5. Morozov VA. On the solution of functional equations by the method of regularization. Sov Math Dokl. 1966;7(1):414–7.
6. Tikhonov AN, V. GA, Stepanov V V., Yagola A. Numerical Methods for the Solution of Ill-Posed Problems. Springer Science & Business Media; 1995.
7. Lawson CL, Hanson RJ. Solving Least Squares Problems. Society for Industrial and Applied Mathematics; 1995.
8. Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. Taylor & Francis; 1979;21(2):215–23.
9. Poletti M. An Investigation of 2-D Multizone Surround Sound Systems. Audio Engineering Society Convention 125. San Francisco, CA: Audio Engineering Society; 2008. Preprint 7551.
10. Poletti MA. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. J Audio Eng Soc. Audio Engineering Society; 2005 Nov 15;53(11):1004–25.
11. Ward DB, Abhayapala TD. Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Trans Speech Audio Process. 2001;9(6):697–707.
12. Choi JW, Kim YH. Integral approach for reproduction of virtual sound source surrounded by loudspeaker array. IEEE Trans Audio, Speech Lang Process. 2012;20(7):1976–89.
13. Daniel J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction. Copenhagen, Denmark; 2003.
14. Boone MM, Verheijen ENG, Jansen G. Virtual reality by sound reproduction based on Wave Field Synthesis. Audio Engineering Society Convention 100. 1996. Preprint 4145 (B–4).
15. Verheijen E. Sound Reproduction by Wave Field Synthesis. Delft University of Technology; 1998.
16. Bojarski NN. Inverse Scattering. Naval Air Systems Command Rep., contract N00019-73-C-0312, (Naval Air Systems Command, Washington, D.C., 1973); 1974. p. 3–6.
17. Bojarski N. A survey of the near-field far-field inverse scattering inverse source integral equation. IEEE Trans Antennas Propag. 1982 Sep;30(5):975–9.
18. Devaney AJ, Porter RP. Holography and the inverse source problem. Part II: Inhomogeneous media. J Opt Soc Am A. OSA; 1985 Nov 1;2(11):2006.
19. Kim Y-H, Choi J-W. Sound Visualization and Manipulation. John Wiley & Sons; 2013.
20. Choi J-W, Kim Y-H. Generation of an acoustically bright zone with an illuminated region using

multiple sources. J Acoust Soc Am. 2002 Apr;111(4):1695–700.
21. Choi J-W, Kim Y-H. Generation of acoustically bright and dark zone using multiple sources. Proc Inter-Noise 2002. Dearbone, MI, USA; p. N322.
22. Druyvesteyn WF, Garas J. Personal Sound. J Audio Eng Soc. Audio Engineering Society; 1997 Sep 1;45(9):685–701.
23. Elliott SJ, Cheer J, Choi J-W, Kim Y. Robustness and regularization of personal audio systems. IEEE Trans Audio, Speech Lang Process. 2012 Sep;20(7):2123–33.
24. Lee C-H, Chang J-H, Park JY, Kim Y-H. Personal sound system design for mobile phone, monitor, and television set; Feasibility study. J Acoust Soc Am. Acoustical Society of America; 2007 Nov 1;122(5):3053.
25. Cheer J, Elliott SJ, Kim Y, Choi J-W. Practical Implementation of Personal Audio in a Mobile Device. J Audio Eng Soc. Audio Engineering Society; 2013 Jun 7;61(5):290–300.
26. Park J-Y, Chang J-H, Kim Y-H. Generation of Independent Bright Zones for a Two-Channel Private Audio System. J Audio Eng Soc. Audio Engineering Society; 2010 Jun 8;58(5):382–93.
27. Cheer J, Elliott S. Design and implementation of a personal audio system in a car cabin. Proceedings of Meetings on Acoustics. Acoustical Society of America; 2013. p.055009.
28. Choi J-W. Spatial manipulation and implementation of sound. Korea Advanced Institute of Science and Technology; 2005.
29. Dawoud M, Anderson A. Design of superdirective arrays with high radiation efficiency. IEEE Trans Antennas Propag. 1978 Nov;26(6):819–23.
30. Cox H, Zeskind R, Kooij T. Practical supergain. IEEE Trans Acoust. 1986 Jun;34(3):393–8.
31. Choi J-W, Kim Y, Ko S, Kim J. Super-directive loudspeaker array for the generation of personal sound zone. 125th Audio Engineering Society Convention. 2008.
32. Golub GH, Loan CF Van. Matrix Computations. 3rd editio. JHU Press; 2012.
33. Shin M, Lee SQ, Fazi FM, Nelson PA, Kim D, Wang S, et al. Maximization of acoustic energy difference between two spaces. J Acoust Soc Am. Acoustical Society of America; 2010 Jul 16;128(1):121–31.
34. Coleman P, Jackson P, Olik M, Pedersen JA. Optimizing the Planarity of Sound Zones. Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception. Guildford, UK: Audio Engineering Society; 2013. paper 5–1.
35. Coleman P, Jackson PJB, Olik M, Møller M, Olsen M, Abildgaard Pedersen J. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array. J Acoust Soc Am. Acoustical Society of America; 2014 Apr 1;135(4):1929–40.
36. Møller MB, Olsen M, Jacobsen F. A Hybrid Method Combining Synthesis of a Sound Field and Control of Acoustic Contrast. Audio Engineering Society Convention 132. Audio Engineering Society; 2012. Preprint 8627.
37. Betlehem T, Teal PD. A constrained optimization approach for multi-zone surround sound. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2011. p. 437–40.
38. Wu YJ, Abhayapala TD. Spatial Multizone Soundfield Reproduction: Theory and Design. IEEE Trans Audio Speech Lang Processing. 2011 Aug;19(6):1711–20.
39. Abhayapala T, Wu YJ. Spatial Soundfield Reproduction with Zones of Quiet. Audio Engineering Society Convention 127. Audio Engineering Society; 2009. Preprint 7887.
40. Jacobsen F, Olsen M, Møller M, Agerkvist FT. A comparison of two strategies for generating sound zones in a room. 18th International Congress on Sound and Vibration. 2011.
41. Choi J-W, Kim Y-H. Manipulation of sound intensity within a selected region using multiple sources. J Acoust Soc Am. Acoustical Society of America; 2004 Aug 2;116(2):843.
42. Choi J-W, Jang J-H, Kim Y-H. Wavefront control by wave number domain focusing. Proc of the 150th Meeting Acoustical Society of America. Minneapolis, Minnesota; 2005.
43. Chang J-H, Choi J-W, Kim Y-H. A plane wave generation method by wave number domain point focusing. J Acoust Soc Am. Acoustical Society of America; 2010 Nov 24;128(5):2758–67.
44. Jackson PJ, Jacobsen F, Coleman P, Abildgaard Pedersen J. Sound field planarity characterized by superdirective beamforming. Proceedings of Meetings on Acoustics. Acoustical Society of America; 2013. p. 055056.
45. Nelson PA, Orduña-Bustamante F, Hamada H. Multi-Channel Signal Processing Techniques in the Reproduction of Sound. Audio Engineering Society Conference: UK 7th Conference: Digital Signal Processing (DSP). Audio Engineering Society; 1992.

46. Nelson PA. Active Control Of Acoustic Fields And The Reproduction Of Sound. J Sound Vib. 1994 Nov;177(4):447–77.
47. Nelson PA, Orduna-Bustamante F, Hamada H. Inverse filter design and equalization zones in multichannel sound reproduction. IEEE Trans Speech Audio Process. 1995 May;3(3):185–92.
48. Choi J-W, Kim Y-H. Active Control for the Enhancement of Sound Field. Proc Active 04. Williamsburg, Virginia;
49. Elliott SJ, Cheer J. Regularization and robustness of personal audio systems [Internet]. ISVR Technical Memorandum 995. 2011 [cited 2014 Aug 15]. Available from: http://eprints.soton.ac.uk/207989/1/Pub12685.pdf
50. Cai Y, Wu M, Yang J. Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint. J Acoust Soc Am. Acoustical Society of America; 2014 Feb 1;135(2):734–41.
51. Coleman P, Jackson PJ, Olik M, Olsen M, Møller M, Pedersen JA. The influence of regularization on anechoic performance and robustness of sound zone methods. Proceedings of Meetings on Acoustics. Acoustical Society of America; 2013. p. 055055.
52. Hansen PC, O'Leary DP. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. SIAM J Sci Comput. 1993;14(6):1487–503.
53. Park J-Y, Choi J-W, Kim Y-H. Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system. J Acoust Soc Am. 2013;134(1):EL112–8.
54. Giusto D, Atzori L, Morabito G, Iera A, Internet T. The Internet of Things: 20th Tyrrhenian Workshop on Digital Communications. New York, NY: Springer Science+ Business Media, LLC; 2010.
55. Ashton K. That "Internet of Things" Thing. RFID J. 2009;22:97–114.
56. Knapp C, Carter G. The generalized correlation method for estimation of time delay. IEEE Trans Acoust. 1976 Aug;24(4):320–7.
57. Benesty J, Elko GW, Mersereati RM. Real-time passive source localization: a practical linear-correction least-squares approach. IEEE Trans Speech Audio Process. 2001;9(8):943–56.
58. Zhang M, Lan H, Ser W. On comparison of online secondary path modeling methods with auxiliary noise. IEEE Trans Speech Audio Process. 2005 Jul;13(4):618–28.
59. Ming Zhang, Hui Lan, Wee Ser. Cross-updated active noise control system with online secondary path modeling. IEEE Trans Speech Audio Process. 2001 Jul;9(5):598–602.
60. Mills DL. Internet time synchronization: the network time protocol. IEEE Trans Commun. 1991;39(10):1482–93.
61. Atkinson B, Blank T, Isard M, Johnston JD (JJ), Olynyk K. An Internet Protocol (IP) Sound System. Audio Engineering Society Convention 117. Audio Engineering Society; 2005. Preprint 6211.
62. Blank T, Atkinson R. Synchronization Strategies for IP Networked Home Audio Equipment. Audio Engineering Society Conference: UK 20th Conference: Convergence. Audio Engineering Society; 2005.
63. Peng C, Shen G, Zhang Y, Li Y, Tan K. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. Proceedings of the 5th international conference on Embedded networked sensor systems. ACM; 2007.
64. Kim H, Lee S, Choi J-W, Bae H, Lee J, Song J, et al. Mobile Maestro: enabling immersive multi-speaker audio applications on commodity mobile devices. Ubicomp '14. Seattle, WA, USA;
65. Kuriyama J, Furukawa Y. Adaptive Loudspeaker System. J Audio Eng Soc. Audio Engineering Society; 1989 Nov 1;37(11):919–26.
66. Kim S, Kong D, Jang S. Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position. J Audio Eng Soc. Audio Engineering Society; 2008 Apr 15;56(4):243–54.
67. Pomberger H, Zotter F, Sontacchi A. An ambisonics format for flexible playback layouts. Proc 1st Ambisonics Symposium. 2009.
68. Brix S, Melchior F, Roder T, Wabnik S, Riegel C. Authoring Systems for Wave Field Synthesis Content Production. Audio Engineering Society Convention 115. Audio Engineering Society; 2003. Preprint 5972.
69. Ahrens J, Geier M, Spors S. The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods. Audio Engineering Society Convention 124. Audio Engineering Society; 2008. Preprint 7330.
70. Melchior F, Pike C, Brooks M, Grace S. On the Use of a Haptic Feedback Device for Sound Source Control in Spatial Audio Systems. Audio Engineering Society Convention 134. Audio Engineering Society; 2013.

71. Bae S-H, Kijima R, Kim W-S. Digital styling for designers: 3D plane-symmetric freeform curve creation using sketch interface. Proceedings of the 2003 International Conference on Computational Science and Its Applications: Part III. Montreal, Canada: Springer-Verlag; 2003. p. 701–10.
72. Wright M, Freed A. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. 1997;