Vol.2015-GN-94 No.23
2015/3/13

高臨場感を与える視聴覚ディスプレイのための音像制御システム

中谷 彰皓[1]　片桐 滋[1]　大崎 美穂[1]

概要：遠隔地間の協調的作業の支援を目指して開発を進めている視聴覚ディスプレイのための，多チャンネル・多音源対応の音像制御システムを紹介する．本システムは，大型ディスプレイ上に正確に音像を生成することを目指して提案されている音像生成方式，音響反射板方式のための音出力制御用ソフトウェアである．低遅延ドライバインタフェースである CoreAudio を用い，音響反射板方式で用いられる 6 台の平面スピーカからの出力音間の同期をほぼ完全な形で実現している．また，音像間には 21 ミリ秒程度のずれはあるものの，複数の音源に対応する複数音像を同時にディスプレイ上に生成することも可能となっている．

キーワード：音像定位，音響反射板方式，多チャンネル音出力制御

1. はじめに

通信技術やマルチメディア技術の発達に伴い，テレビ会議システムや，遠隔地間の協調的作業を支援する遠隔コラボレーション支援システムの研究開発が盛んに行われている（例えば t-Room[1]）．そうしたシステムにおいては，ユーザ間における（知覚的な）臨場感あるいは同室感の達成レベルがコミュニケーション・作業の質に大きく影響する．そのため，特に視覚的臨場感の向上を目指した，大型の映像ディスプレイ（以下，ディスプレイと呼ぶ）の利用が盛んに行われるようになった．しかし，ディスプレイの大型化は，それが生成する映像に付随する音声データを生成するスピーカの配置を制約し，かえって聴覚的臨場感，特に音源の方向・位置の同定，即ち音像定位の正確さを低下させ易い．例えば，大型ディスプレイを用いて，遠隔地間で全身動作などの視覚情報を共有することができても，口唇位置とな異なる方向から音声が聞こえてくるなどの，正確かつ自然な情報共有を妨げる状況が起こってしまう．

上記のような状況を受け，筆者らは，大型ディスプレイの利用を前提とした上で，そのディスプレイ上の映像と音像との位置を一致させることを目指す視聴覚ディスプレイの開発を進めてきた [2]．本ディスプレイは，例えば縦型 60 インチの液晶ディスプレイなどのような，利用者の等身大映像の表示が可能な大型ディスプレイの左右端に，剛体の音響反射板によって囲まれた複数の平面スピーカからなるスピーカユニットを配置し，ディスプレイ面上に音像を生成することを目指すものである．本稿では，この視聴覚ディスプレイ方式を紹介し，特に，その音像生成手続きを制御するための多チャンネル音声出力制御システム（ソフトウェア）の開発の現状を報告する．

[1] 同志社大学大学院 理工学研究科
Graduate School of Science and Engineering, Doshisha University

2. 音の方向知覚

初めに，人が持つ音（音源）の方向知覚のメカニズムを概説する．

まず，基本的用語を定義する．音信号を発するものを音源と呼び，受聴者が音源として知覚する（位置や大きさを伴う）像を音像と呼ぶ．

人は両耳の鼓膜で受け取った情報から，音の空間を知覚する．この音空間の知覚は，方向知覚と距離知覚，広がり知覚などから成り立ち，その全てが音像の知覚の手がかりとなる．

人が音の方向を知覚する仕組みは，左右方向からの音を知覚する「水平面」と，前後上下方向の音を知覚する「正中面」とに分けられる．水平面の方向知覚は，頭部伝達関数（HRTF: Head-Related Transfer Function）によって生じる両耳間時間差（ITD: Interaural Time Difference），両耳間レベル差（ILD: Interaural Level Difference）を用いることで行われる [8]．一方，正中面における方向知覚は ITD や ILD を用いることによって行うことができず，水平面のそれと比較して著しく精度が低下し，日常生活においてもしばしば誤った判断を行う．正中面における方向知覚の手

がかりは，HRTF による振幅スペクトルに起因するという説が有力だが，未だ必ずしも正確には解明されていない．

なお，大型ディスプレイに相当する壁面上における音像（の位置あるいは方向の）定位に関しては，音源と音像定位位置の水平面方向と正中面方向のずれは正中面方向で多く，比率は 6.4 倍になったと実験報告もなされている [6]．また，筆者らが研究を進めている視聴覚ディスプレイと同形状の板面上の音像の定位精度については，筆者らもかかわった実測例もある [7] 実測の結果，筆者らが提案する（60 インチ縦型）視聴覚ディスプレイの場合，水平面上の方向定位精度は 70%程度であり，正中面方向のそれは 40%程度であることがわかっている．

また，人が実際に音の方向知覚を行う場合には，HRTF や ITD，ILD などの基本的な機能を利用するだけでなく，頭の向き（両耳の向き）を変えるなどの動作を用いることで知覚精度を向上させる．慣れてくると片耳だけでも方向知覚ができるのは，動作に伴って変化する音の到来方向による音質の違いを記憶しているためとされている [9]．

次に，上述した人の方向定位能力を前提として，仮想的に音像を生成する最も一般的なステレオ再生方式について概説する．この方式は，水平面上の左右位置に 1 つずつのスピーカを配置する．左右のスピーカがコヒーレントな信号を出力し，左右の耳に到達した音信号の音圧レベルが同等であり，到達した時間差が $1ms$ 未満である場合，音像は左右のスピーカの中央付近に定位される．しかし，音圧レベルが左右の耳で異なるとき（ILD= 0 でないとき），または到達時間が一定量以上異なるとき（ITD が一定の値以上のとき），音像はより大きな音圧レベルの信号を到達させたスピーカの方向，より早くに音声を到達させたスピーカの方向に支配的な影響を受ける．そのため，通常，この標準的な 2 チャンネルステレオ音像生成システムにおいては，全てのスピーカからの距離が等しい位置で受聴することが前提となっている．

図?は，大型ディスプレイの左右端に，上記の標準的なステレオスピーカユニットを配置している様子を図解している．上段の解説より，ディスプレイ中央の正面に位置する受聴者は，原理的には，ディスプレイ面上に正確な音像定位を行い得るように考えられる．一方，図 2 は，同じディスプレイとスピーカユニットの配置において，受聴者が左スピーカに近い位置で受聴している様子を示している．この場合，左スピーカからの出力音が支配的に影響し，音像は，左スピーカに近いディスプレイ面上の位置あるいはそのスピーカ自体の位置に定位されることになる．

3. 音響反射板方式

筆者らの視聴覚ディスプレイが用いる音像生成方式は，大型ディスプレイの利用を前提として，その正面のみならず，もっと広い受聴エリアにおいても，受聴者による正しい音像方向定位を実現することを目指すものである [3]．図 3 は，その実装の様子を図解している．本方式は，例えば協調的作業に従事する利用者の全身映像を等身大で表示することが可能な，60 インチ縦型液晶ディスプレイを用い，その左右端のそれぞれに，L 字型の剛体である音響反射板によって囲まれた 3 つの平板スピーカからなるスピーカユニットを配置する．図中左側が，本方式（以降，音響反射板方式と呼ぶ．）による機器の配置図であり，図中右側は，スピーカユニットで用いられている平面スピーカの寸法図である．各スピーカユニットでは，中央位置とその上下に，合計 3 台の平面スピーカが取り付けられている．また，図からもわかるように，それらはディスプレイ面と直交するように置かれている．結果的に，同じ高さにある左右のスピーカは互いに対峙するように配置される．

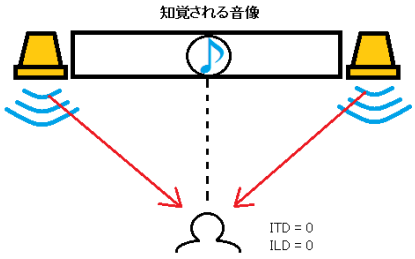

図 1 ステレオ:中央軸上の場合.

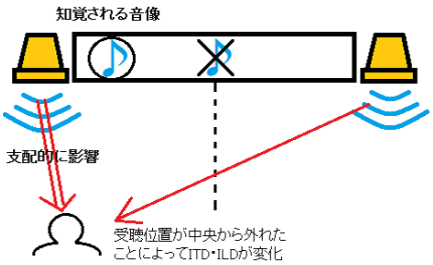

図 2 ステレオ:左右に外れた場合.

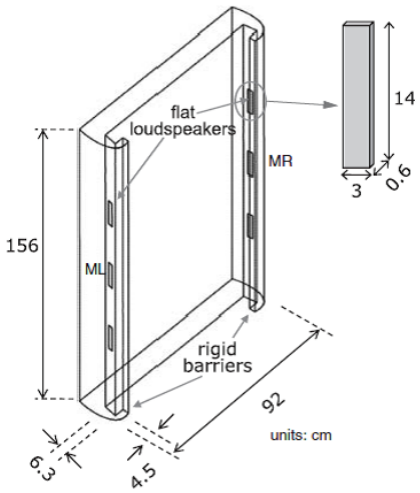

図 3 音響反射板方式の機器配置図 (文献 [3] より引用).

一般的な 2 チャンネルステレオ方式（図 1 等を参照.）と比較すると，スピーカ付近から直接的にディスプレイ前面に向かう音の広がりが音響反射板によって抑制され，同時に，音響反射板とディスプレイ面からの反射によってディスプレイ表面に音像が生成され易い構造になっていることがわかる．

図 4 は，音響反射板方式において，ディスプレイ中央の正面に位置する受聴者にスピーカ出力音が到達する様子を図解している．このとき，標準的なステレオスピーカ方式とほぼ同様に，各スピーカからは直接音が到達する．一方，図 5 は，同方式において，ディスプレイ左側に位置する受聴者にスピーカ出力音が到達する様子を図解している．この受聴者に対しては，音響反射板が作用して直接音が遮られ，近傍のスピーカからは回折音が，遠方のスピーカからは直接音が届く．回折音は，直接音に比べ到達時間が大きく，かつ音圧レベルは低下する．その結果，両耳間到達時間差や両耳間音圧レベル差が生じにくく，受聴位置条件による方向定位の偏りが抑制されることになる．

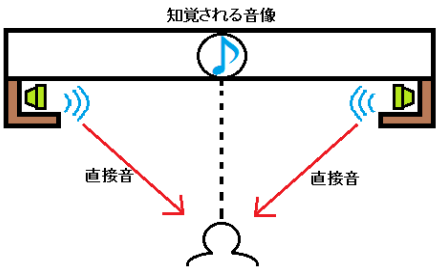

図 4 音響反射板方式:中央軸上の場合.

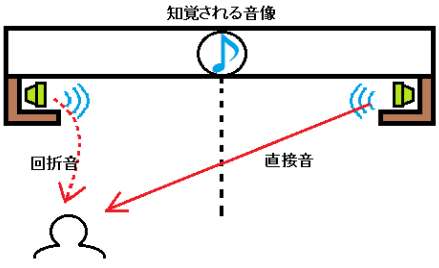

図 5 音響反射板方式:左右に外れた場合.

これらの効果により，音響反射板方式は標準的なステレオ方式よりも（ディスプレイ前の）広い受聴エリアで，より正確な音像定位をもたらすことが期待される．

4. 音像制御システム

音響反射板方式を用いてディスプレイ上の指定の場所に音像を生成するためには，6 台の平面スピーカそれぞれの出力の音圧レベルや位相を制御する必要がある．本稿で紹介する音像制御システムは，その制御を実現するためのソフトウェアである．

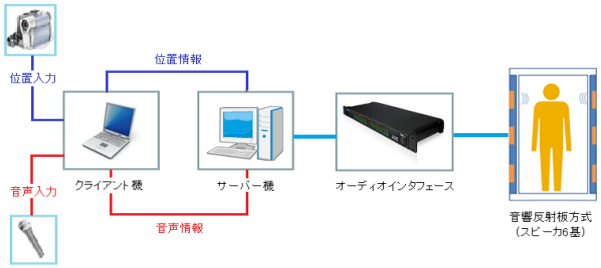

図 6 音響反射板方式のシステム構成.

本システムは，ネットワークによって接続される遠隔コラボレーション支援システム等への適用を考慮しているため，サーバークライアントの構成をとる．スピーカによる出力を制御するサーバと，マイクによる音声データ等の入力を制御するクライアントとから成る．図 6 は，本システムの概要を図解している．システムは以下のように動作する．まず，クライアントからサーバに接続を行い，データ伝送のための伝送路の確保を行い，続いて，音声出力ストリームの生成を行う．次に，クライアントは，マイクから音声データを入力として得，また，音像を生成すべき位置データ（ディスプレイ上の座標データ）を得て，それらをサーバへ送る．一方，サーバは，受け取った音声データ

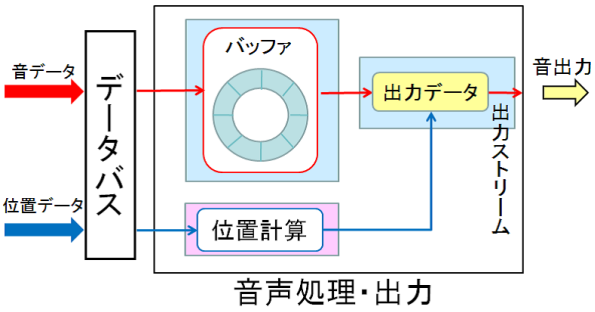

図 7 サーバ内の処理.

と位置データを用いて，スピーカからの音出力データへの変換を行い，そうして得られる音信号を実際にスピーカに出力する．なお，サーバは，データバスを通してクライアントからの音声データと位置データを受け取る．音声データはリングバッファに格納され，位置データは指定位置に音像を生成するための音圧の重み算出に用いられる (図 7 参照).

4.1 音圧制御

先述したように，音像生成位置は，各スピーカの音圧レベルと位相を制御することによって行うことができる．開発した本システムには，特に，音圧レベルの制御機能を実装した．

位置情報はディスプレイのピクセル位置に準じて設定されている．大型ディスプレイに対して水平面方向を x 軸と

し，正中面方向を y 軸とした xy 座標を用いて計算を行う．座標は次の範囲で定義されている．

$$0 \leq x \leq 1080 \tag{1}$$

$$0 \leq y \leq 1920 \tag{2}$$

左右スピーカの音圧を制御するパンニング法には様々な版が存在する．本実装では特に，2 チャンネルステレオ方式のパンニング法の代表とされるタンゼント則を採用する．このパンニング法は，図 8 に示されるように，受聴者の正面と生成する仮想音像との角度を ϕ，各スピーカとの角度を ϕ_0 として，式 (3) に基づいて重みの係数 w_L と w_R を導き出す．ただし，このときスピーカの配置及び人間の頭部形状に左右対称性を仮定している．

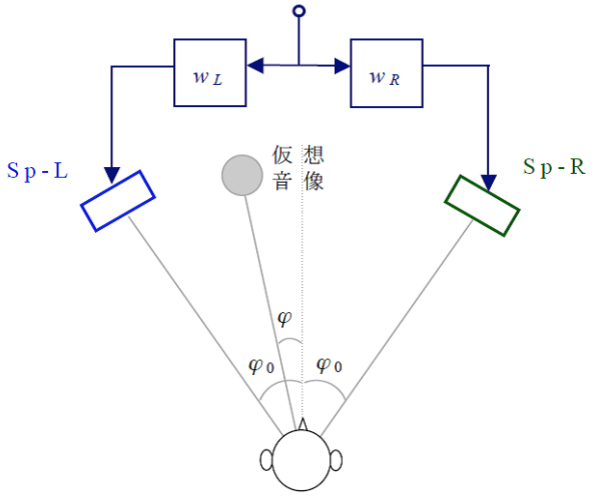

図 8 ステレオ再生における仮想音像の生成 (文献 [5] より引用).

$$\frac{\tan\phi}{\tan\phi_0} = \frac{w_L - w_R}{w_L + w_R} \tag{3}$$

ここで，これらの重みの比により，左右の音圧レベル比である X_L と X_R は式 (5) のように求められる．

$$X_L = \frac{w_L}{w_L + w_R} \tag{4}$$

$$X_R = \frac{w_R}{w_L + w_R} \tag{5}$$

次に，縦に配置されたスピーカの上下間の音圧レベル比を，音圧レベルの距離減衰の知見に基づいて行う．薄型スピーカを用いる場合，音源は線音源と見なし，音源からの距離が 2 倍になるごとに音圧レベルが 3dB ずつ減少する [10]．この知見に基づいてシミュレーションによって近似的に値を設定したものが，式 (6) である．また，式 (6) に用いられる $y^\star$ は式 (7) によって左右一対となったスピーカの段ごとに求める．y_{height} はスピーカの高さを，y_{max} はディスプレイの y 座標の最大値を示す．

$$Y_{Level} = a * exp(b * y^\star) + c * exp(d * y^\star) \tag{6}$$

$$a = -4.554 * 10^{-7} \qquad b = 18.1$$
$$c = 1.002 \qquad d = -1.376$$

$$y^\star = |\frac{y - y_{height}}{y_{max}/2}| \tag{7}$$

以上の手順で求められた X_L または X_R と，Y_{Level} を乗算することにより，各スピーカチャンネルの音圧の重みが求められる．

4.2 多チャンネル出力制御

音響反射板方式の説明から，そこで用いられる 6 台のスピーカの出力信号間の位相差が適切に制御される必要があることがわかる．こうした要請に基づき，本制御システムも，6 台のスピーカからの出力の同期を目指して，多チャンネル出力制御を行う．そのため，特に，こうした多チャンネル制御に適していると考えられているドライバインタフェース，CoreAudio[11] を，サウンドカード等とのオーディオインタフェースを確立するために採用する．

制御システムは，多チャンネル出力を行うために，クライアントから受信した音声データをリングバッファに格納する際に，多チャンネル出力に対応したデータ形式へと変換する．音声データ変換の概図を図 9 に示す．まず，受信した音声データを設定されたビットレート分のビット数で分割し，音声波形の標本点の振幅値を示すサンプルのデータブロックに分割する．次に，分割されたサンプルごとに，使用するチャンネル数分のデータブロックの複製を行い，それらを羅列することで新たなデータとし，リングバッファに格納する．このようにして複製して記述されたそれぞれのブロックが各チャンネルの出力信号として対応する．出力の際にこれらのブロックに対応するチャンネルの音圧を乗算することで，多チャンネル出力による音像制御を行っている．

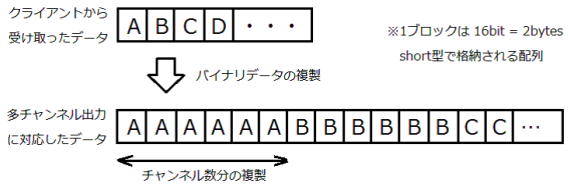

図 9 多チャンネル音声データの生成.

4.3 複数クライアントの接続

1 面のディスプレイ上に複数の音源が登場する（対応する音像を生成する）状況は珍しいことではない．本制御シ

ステムも，複数のクライアント（それぞれが一つの音源あるいは音像に対応する）を接続し，複数の音像を同時に生成する機能を実装する．

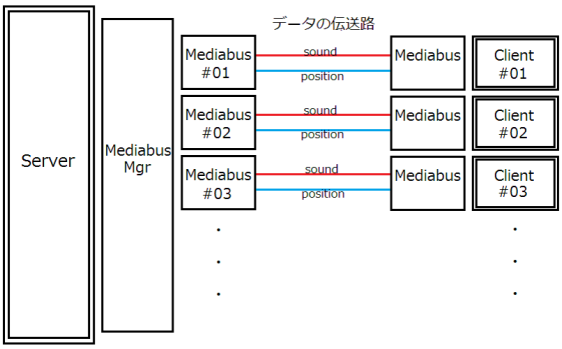

図 10 Mediabus によるデータ伝送．

そこで，クライアントとの接続には仮想のデータバスとして用いる Mediabus を実装した．図 10 はクライアントと接続を行っている Mediabus の概図である．Mediabus はクライアントとの接続が確立するたびに，Mediabus と音声制御・出力部が 1 対 1 で対応するように生成される．Mediabus 内にはデータ伝送チャンネルが設けられ，扱うデータの種類に応じてチャンネル数を増減させることができる．現行の実装ソフトウェアでは，音声データと位置データを扱うため，Mediabus 内のチャンネル数は 2 とし，また同時に生成される Mediabus の数は 10 と設定している．

5. システム評価

開発したシステムの動作確認を兼ねて，各チャンネルからの出力信号を解析し，チャンネル間の同期性や複数音像出力時の音像間の時間差の検証を行った．なお，出力音は全て，16bit のビットレートと 44.1kHz の標本化周波数で表現されている．

5.1 チャンネル間同期

単一の音像を表現する出力系において，各出力の同期性能を測った．使用した音源は 1000Hz のサイン波と 1sec 間隔のパルス波，ホワイトノイズ，マイクによる破裂音の入力である．システムの音圧制御の都合上，同時に使用するスピーカは最大 4 チャンネルのため，4 チャンネルずつの出力を収録し，解析を行った．結果を 1 に示す．なお，表内の数値の単位は標本数[*1]であり，チャンネルごとに得られた波形間の

破裂音において稀に 1 標本のずれが見られたが，これは 0.023ms 程度の誤差であり，音の方向定位における先行音効果発生の閾値と比較しても極めて小さな値であるため，

[*1] 44.1kHz の標本化周波数で標本化することに対応する標本の意味．

表 1 チャンネル間の出力時間差

音源	max	min	average
サイン波	0	0	0
パルス波	0	0	0
白色雑音	0	0	0
破裂音	1	0	0.033

問題とはならないと考えられる．結果として，開発した制御システムは，チャンネル間でほぼ完全に出力の同期を確立しているものと判断される．

5.2 複数音像間同期

複数の音像を同時に生成し，音像間の同期の精度を測った．マイクによる破裂音の入力を 3 つのクライアントに同時入力することで 3 つの音像を同時に制御し，出力の収録と解析を行った．結果を表 2 に示す．表内の数値の単位は表 1 と同様に標本数である．

表 2 音像間の出力時間差

max	min	average
3983	33	941.3

ここでは，生成した音像間で平均して 21.3ms の出力時間のずれを観測した．音源入力が同時であるため，この時間差はクライアント間またはサーバ内の Mediabus，出力制御インスタンスの処理の差によるものであると考えられる．また観測された時間差の振れ幅は大きく，本実験で観測された最大値のような時間差（90.3ms）を生じた場合，映像と音声のずれを引き起こし，違和感の原因となることが予想される．平均値についても，相関のある音源を用いた複数の音像提示をする場合には十分に知覚できる程度の誤差（21.3ms）であり，同期性改善の必要があるだろう．

6. おわりに

提案・実装した音像制御システムは，音響反射板方式のための使用を前提としているものの，他の多チャンネル音出力用システム等にも転用可能である（音以外の多チャンネルセンサデータを扱う場合等へも原理的には適用可能と考えられる．）．そのため，このシステムを利用することで，遠隔地間の協調的作業の支援などに限らず，映像と音声を使用する様々なシステム（例えば，ディジタルサイネージなどとして）で一層の臨場感の向上が見込まれる．また，今回は，収音系に関してはあまり触れていないが，音声入力を行う収音系を発展させることで，さらに豊かな視聴覚ディスプレイが可能となると思われる．

参考文献

[1] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai, Naomi Yamashita, Katsuhiko Kaji, Junji Yamato, and Kenji Nakazawa,

”t-Room: Next Generation Video Communication System,” Proceedings of World Telecommunications Congress at IEEE Globecom (WTC’08), 2008.
[2] 澤崎博章，山下春香，中谷彰皓，片桐滋，大崎美穂，”映像の有無を伴う音響反射板方式の音像定位制度の評価，” 電子情報通信学会技術研究報告. EA, 応用音響 113(242), 9-16, 2013.
[3] Gabriel Pablo Nava, Keiji Hirata, Masato Miyoshi, ”A loudspeaker design for sound image localization on large flat screens,” Acoustical Science and Technology Vol.31,No.4,pp.278-287,2010.
[4] Y.Makita, ”On the Directional Localization of Sound in the Stereophonic Sound Field,” EBU Review No.73-A, pp102-108, 1962.
[5] 安藤彰男，”高臨場感音響技術とその理論 Theory of Three-Dimensional Sound Field Reproduction ”, IEICE Fundamentals Review Vol3 No.4, pp33-46, 2009.
[6] 倉本敏行，佐多正至，高橋誠，”前方平面に配置したスピーカによる音像の定位，” 電子情報通信学会技術研究報告. MBE, ME とバイオサイバネティックス 101(733), 103-107, 2002.
[7] 松村友輔，”大型ディスプレイ上の音源位置に関する同定能力の分析，” 同志社大学理工学部情報システムデザイン学科 2014 年度卒業論文，2015.
[8] 飯田一博，森本正之，“ 空間音響学 ”，コロナ社，2010.
[9] 安田仁彦，“ 機械音響学 ”，コロナ社，2004.
[10] 古井貞熙，”音響・音声工学 ”，近代科学社，pp11-12,1992.
[11] Apple 社，” CoreAudioOverview ”