

音響制御プログラムの開発

Audio field control program

矢沢 晃
Akira YAZAWA

〒178-0061 東京都練馬区大泉学園町七丁目 1 0 番 1 0 号 E-mail: yazawa_amgs@
yahoo.co.jp)

ABSTRACT. It is well-known that Headphone system makes sound field localization inside of the head, not in front of head. The fundamental cue to localize sound source is time and strength difference. In this paper, I focused on the frequency characteristics which depend on the direction of the sound source and Digital signal processing technology can emulate Head front sound frequency characteristics to realize Head front sound and cancel reflection and distortion of headphone inside the head. Speaker system does not have enough front sound fields because of stereo effect. This processing technology also makes sound source field localization far from the head in Speaker system.

1. 背景

人間は2つの方法で音の定位認識を行っていると考えられる。左右の耳に届く音の強弱/時間差と、音源と耳の位置関係に対する周波数特性の違いの認識である。このため、通常 Speaker 再生の場合には Speaker のある位置に音が定位される。また、人間は Stereo 再生の場合左右の音の相関を認識することによりさらに音の定位を行っている。例えば、Vocal 音は通常正面に置かれるので左右の音に同じように録音されている。このため左右の相関がかなり強いと認識され、左右の Speaker の位置を結んだ中心に定位される。これが Stereo 効果である。しかし、Speaker の位置を結んだ線よりも外に出ることはなく広がり感が出ないという問題が発生する。

Head Phone の場合も同様に左右の耳の位置に Speaker が存在するためこの位置に音が定位するが、Vocal は Stereo 効果で Speaker の中心である頭の中心に音が定位してしまうことになる。また、本来存在しない耳の中や耳の外に Head Phone があるため、それによる反射/歪により周波数特性などが異なった音として聞こえてしまうことも問題となる。

2. 目的

Speaker, Head Phone のシステム及び音のソースもそのままに、これらの問題を Digital 信号処理技術で改善するのが今回の目的である。

音源と耳の位置関係によりその周波数特性が異なることに着目し、この周波数特性が頭、耳の形状によるものであるという仮定のもとに各種の解析を行い Digital 信号処理で再現することを行う。

現在の Speaker、Head Phone でも斜め/横からの成分は存在するので、この正面からの成分を加えることにより

正面から来る音の前方定位をさせ、広がり感を持たせる。

今回以下の5項目に関して検討を行う

外耳道の Emulate

耳介の Emulate

音の回り込み補正

左右の音を Incoherent に

Head Phone の歪/反射補正

それぞれの形状から Filter の形を決定しプロトタイププログラムを作成する。有限会社イクスクラに以下の機能がある Software を作成してもらい最も良い係数/遅延値の組み合わせを検討する。

簡易な入力方法で係数/遅延値の設定/変更

Real Time での係数/遅延値変更

係数/遅延値の保存/読み出し

これにより Head Phone での再生のみならず、Speaker でも音を前方に定位させると同時に広がり間のある音の再生を実現する。

また、マサチューセッツ工科大学(MIT)から公開されている Dummy Head による Impulse Response の解析も参考に行ってみる。従来この解析には測定時の Speaker 特性のキャンセルが必要となるが、Frequency Domain でこのキャンセルを行った場合には高周波が強調されてしまいうまくいかない問題がある。今回 Time Domain で行う。

3. 従来の前方定位方式

音の前方定位に関しての従来の方法を紹介する。

3-1. Dummy Head Recording(従来の前方定位例 1)

図 3-1 にこの方式の録音方法を示す。Dummy Head の鼓膜の近くの音を録音して実際の音に近づけようとするものである。ところが以下のように3つの問題点がある。まず、Head Phone を使用して聴く場合には、録音時にはな

かったHead Phoneが耳の中または頭の外に存在してしまうため反射/歪などが発生してしまい出力される音が本来の特性と異なってしまいます。

第二にこのDummy Head Recordingの音をそのままSpeakerで再生しようとするとう鼓膜までの経路が重なってしまうことになりSpeaker再生できないという問題が発生する。

第三にこの方法は特殊な状態で録音しなければならないので通常のCDのような既に録音されてしまったものには対応できないという大きな問題がある。このようにDummy Head Recordingは良い方法ではあるが、特殊な録音方法であるのでSpeakerでもHead Phoneでも問題が発生してしまい、一般には受け入れられていない。

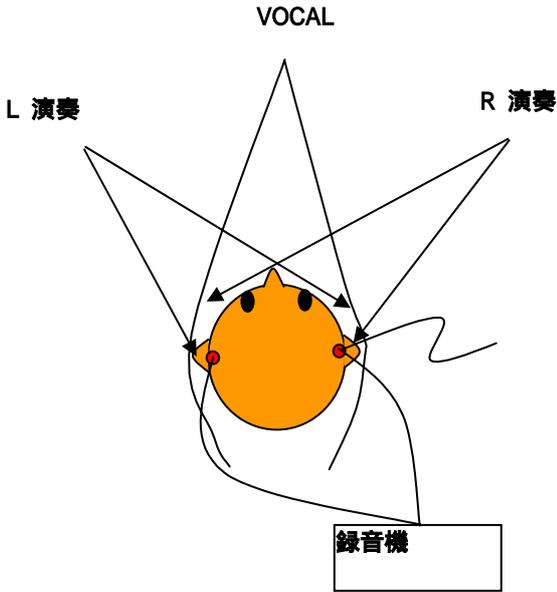


図 3-1 Dummy Head Recording

3-2 HRTF を Emulate(従来の前方定位例 2)

この方法はSpeakerの置かれた場所から耳までの経路のHRTF(Head Related Transfer Function)を測定してSpeakerでの再生時の環境を作り出すことを行っている。(図 3-2)

人間はこの左右の耳までの時間差/強弱、耳と音源の位置関係による周波数特性からSpeakerの位置に音を定位させる。音源がひとつでSpeakerの位置と同じ場合は問題ないが、通常複数の音源があるため音源の数だけ

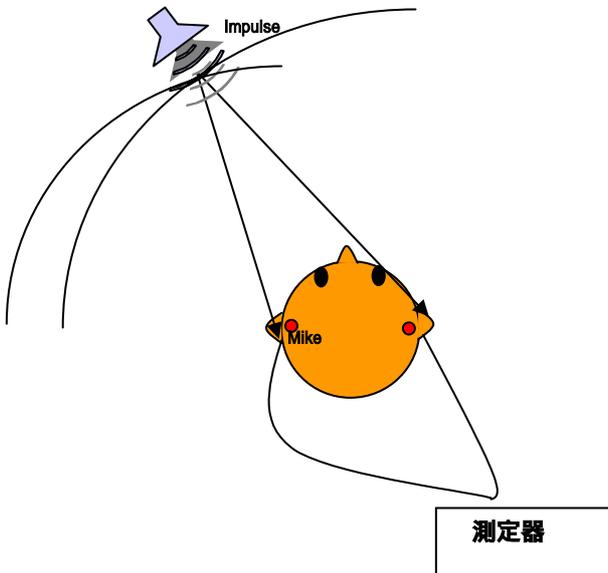


図 3-2 HRTF を Emulate

Speakerが必要となってしまいます。一方、2ch Stereoは全ての音を2chに凝縮しているので本来の音源の位置でなくてSpeakerの位置に定位してしまうということになる。

3-3 音源定位の原理

人間の音源定位認識は左右の耳に届く時間差/強弱と、音源と耳の位置関係による周波数特性から行われている。図 3-3 は音源と耳の位置関係を示したものであり、頭の正面にある場合に角度0度、頭の側面/耳の上にある場合に90度になっている。図 3-4 はそのときの特性であり、それぞれの角度によりその周波数特性が異なっているのが分かる。これは頭/耳の形によるもので、この違いを人間が認識し音の定位を行っている。図 3-6 は等ラウドネス特性と呼ばれ、人間が認識できる限界の音圧レベルを示したものである。これによると人間の歌声あたりに相当する4KHzあたりが小さい音でも認識できることが分かる。図 3-5 はMITのデータの中で0度の特性の逆数をとったものであるが、この両者はほぼ一致している。等ラウドネス曲線は人間の声を認識しやすくコミュニケーションを図りやすくする特性と従来から認識されているが、この特性がMITのデータの中で正面からの特性と一致しているということはそれだけ正面からの信号が重要であるという証にもなる。

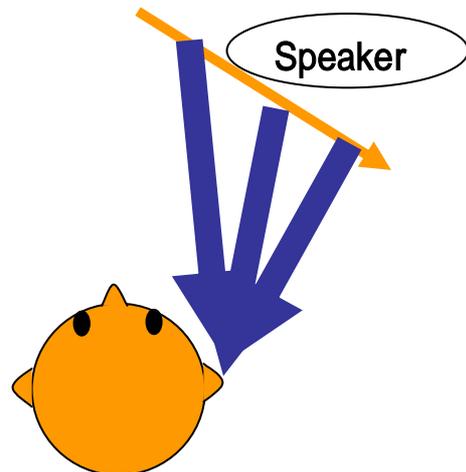


図 3-3 音源と耳の位置関係

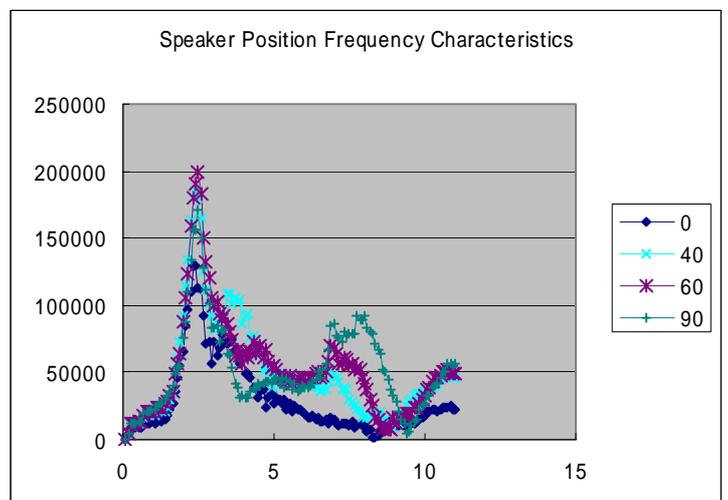


図 3-4 Speaker 位置 VS 鼓膜の周波数特性 (MIT Data より)

を出すものも開発されている。

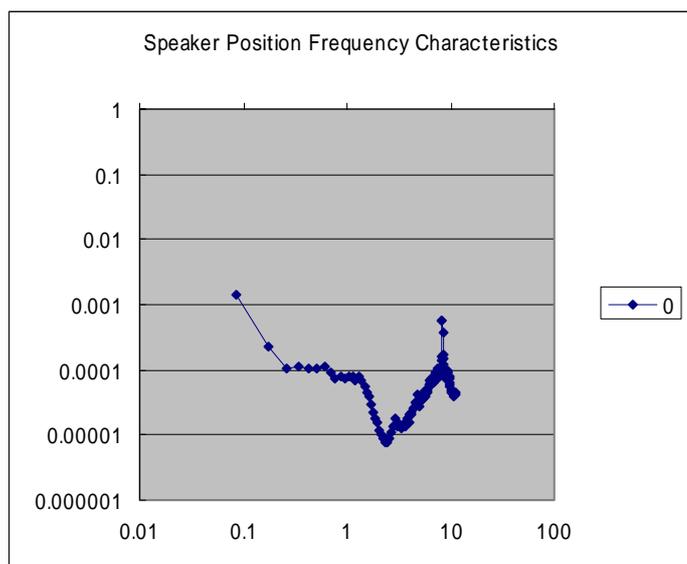


図 3-5 0度の特性の逆数

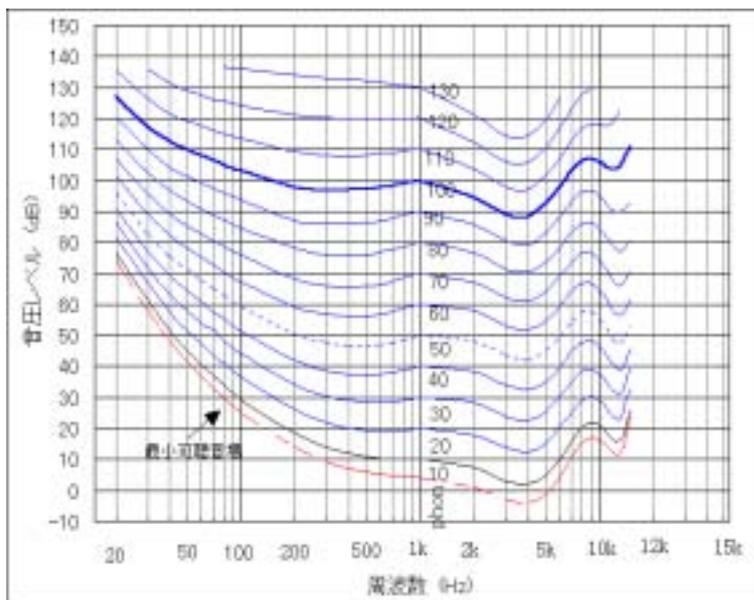


図 3-6 等ラウドネス曲線

<http://acoust.arch.kumamoto-u.ac.jp/labnoiseonatu.html>

4. Speaker Systemの問題点

図 4-1 は Speaker System を表している。本来 Vocal は頭の正面にあり、その正面からの周波数特性を認識することにより正面に音定位される。ところが Speaker 再生の場合、Vocal は正面からの周波数特性が存在しないのに Stereo 効果により左右に同じ音が入っていることだけで正面に定位させられていることから人間にとっては不自然な音になっており違和感がある。また、Speaker が前方に置かれているときは良いが、車の Speaker のように人間の横に置かれたりするレイアウトの場合には音の定位が人間の頭のあたり、後ろあたりでなってしまう。

これを避けるために、最近 Note Book タイプの PC では LCD 表示部に Speaker を配置することにより正面から音

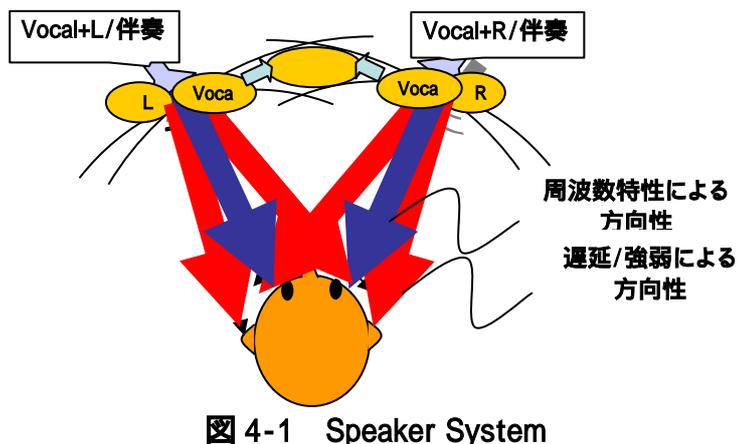


図 4-1 Speaker System

5. HeadPhone System での問題点

図 5-1 は Head Phone System を示している。Head Phone では Speaker が耳の位置に置かれていることから耳の位置に音が定位される。伴奏などのように左右の音の相関が弱い音はそのまま耳の位置に定位しているが、Vocal のように左右の音の相関が強い音は Stereo 効果でその中心である頭の中心に定位されるという問題点がある。また、Head Phone が耳の中または頭の横に存在するため本来存在しない反射 / 歪が発生してしまい人間の音の定位に必要な特性を変えてしまうという問題もある。

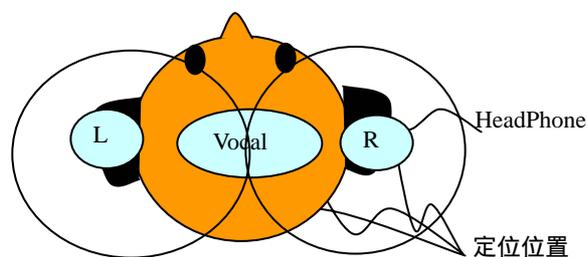


図 5-1 Head Phone System

6 検討内容

今回、検討された技術は以下の 5 点である。

外耳道の emulate

耳には鼓膜につながる外耳道が存在する。この外耳道は片側がオープン、もう一方が鼓膜でできている筒のようになっている為、定在波が存在する。この外耳道特性を信号処理により再現する。また、この外耳道特性こそがラウドネス曲線(中域が良く聞こえる)を実現するものになっている。このラウドネス曲線は人間が人の声を効率良く認識するように出来た機能であり、これを実現することにより、本来の音が再生できる。

耳介の Emulate

正面からの音は頭にぶつくと頭の周り回りこむように進んでいく。つまり、耳からみた場合の音源位置から

考えると正面の音は最も耳介の反射の影響を受けやすいと考えられる。一方、Speaker の位置が左右に広がるに従いこの反射も弱くなり耳の真横にきたときにはこの反射の成分がなくなる。この特性を実現する。

音の回り込み補正

人間は頭の周りを回りこんで伝わる音の位相差を認識することでも、前方定位を認識している。また、音の回り込みには回折が起こる。これは LPF として音に影響を与えている。この回り込み音の位相 / 時間差により人間は前方定位を実現している。最初に届く耳介 / 外耳道特性の音だけでは前方定位は実現出来ず、後から届く回り込みの音との両方で定位を行っている。

左右で Incoherent な音にする。

人間は、左右の音の相関を検出して定位を行っている。つまり、左右の伴奏は相関がないため左右の Speaker 位置に定位し、正面から来る Vocal は左右の音に同じように録音されているため左右の Speaker 位置の中心に定位される。つまり Vocal も左右の相関がないように工夫することにより中心には定位しなくなる。

サイン波の場合左右で位相の半分の遅延(逆位相)を作ると、相関がない(Incoherent)と認識される。つまり、Vocal あたりの周波数の半周期の遅延時間を作ることによって前方定位を実現することができる。しかし、現実に逆相となる時間差を加えることは本来の音自体に影響を与えるので人間が気がつかない程度の遅延を加えてみる。これは Head Phone のみのできることであり、Speaker の場合はその聴く位置が左右対称でなく左右の音にはほとんどの場合時間差が存在し、そのまま Incoherent になることから加える必要はない。

Head Phone の歪 / 反射補正

本来耳の中や頭の外に存在しない Head Phone が存在してしまうことから発生する音の歪 / 反射がある。これを打ち消すような信号を Digital 信号処理により作り出す。これは Head Phone のタイプ、大きさによって変わってくるので、係数の補正が必要となる。

また、この反射は Head Phone の外への 5KHz から 6KHz あたりのシャリシャリと聞こえる音の発生源にもなっており、この補正を行うことでこのシャリシャリ音も無くすることも可能ではないかと考えた。

7 MIT Data 解析(裏付け)

MIT から Dummy Head による Impulse Response data が公開されている。

これは音源と Speaker の全ての位置に対するデータが測定されており、このデータをもとに今回のデータの裏付けを行ってみる。

そもそも Dummy Head で耳の特性を測定するのは耳の理想的な Impulse Response を求め、それを通常の音に適用して完全な音の定位を行おうとするものである。この方法として Impulse Response をそのまま測定する方法と、全ての周波数を含む信号を与えて周波数特性を測定し、それから Impulse Response を求める方法がある。ところが、周波数特性でも音の定位に重要な部分とそうでない部分があるため、周波数特性が理想的でも Impulse Response に変換するときに押さえるところとそうでないところの切り分けが出来ていないと、結局意図しない特性になってしまう。今回の MIT のデ

ータは直接 Impulse Response を測定しているが、そのまま使用すると高周波成分が強くてうまく定位しない。これは、使用する Speaker が理想的な Impulse を作り出すことが出来ず Speaker 特性のキャンセルが必要になるためである。これを Frequency Domain で行っても上記のような問題が発生してしまうので、今回はこれを Time Domain で直接変換を行って上記検討時の参考とする。

8. 解析結果

外耳道の emulate

図 8-1 のように外耳道は片側が開かれた筒のような構成になっており、これにより定在波がある。MIT のデータを FFT した図 8-2 の矢印 1 の部分がこの外耳道による特性を示している。この特性に合う Filter を決定し、係数 / 遅延値は人間の耳の大きさにより異なるので、イコスタラの Program で Real Time で切り替え確認して決定した。尚、Head Phone 再生のときには耳のところに Speaker があるためこの外耳道特性が全くないためこの処理は必須であるが、Speaker 再生のときは実際の耳でこの特性が現れるので、効果を小さくする必要がある。

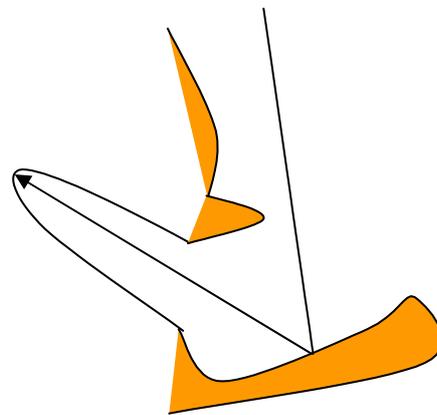


図 8-1 外耳道

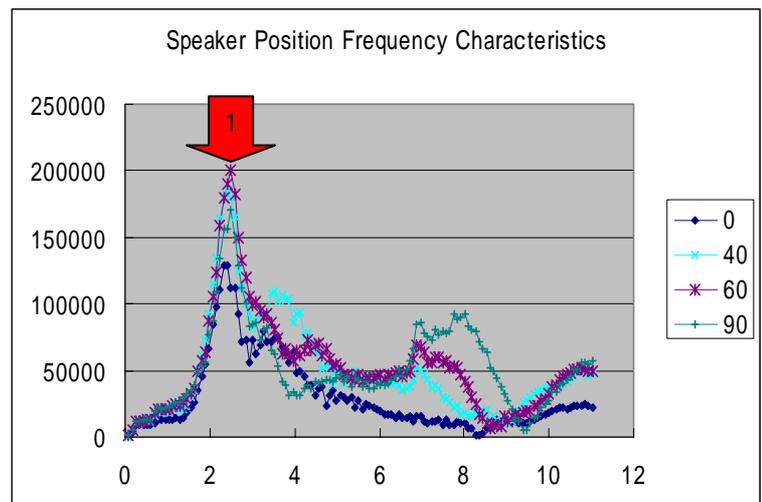


図 8-2 外耳道特性

耳介の Emulate

図 8-3 のように最初に鼓膜に届く音と反射されて届く音でこの特性は決まる。MIT データ図 8-4 の矢印 2 の部分がこれに相当。

MIT の Dummy Head で得られた Data で分かるように Speaker の位置により 8KHz あたりの特性が異なり、Speaker の位置が前になるに従い谷が現れてくる。この 8KHz あたりの谷は謎とされていた。今回の解析により、これが耳介の反射で再現させることが確認できた。

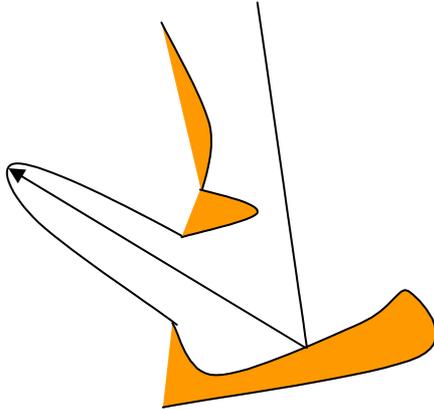


図 8-3 耳介

Speaker Position Frequency Characteristics

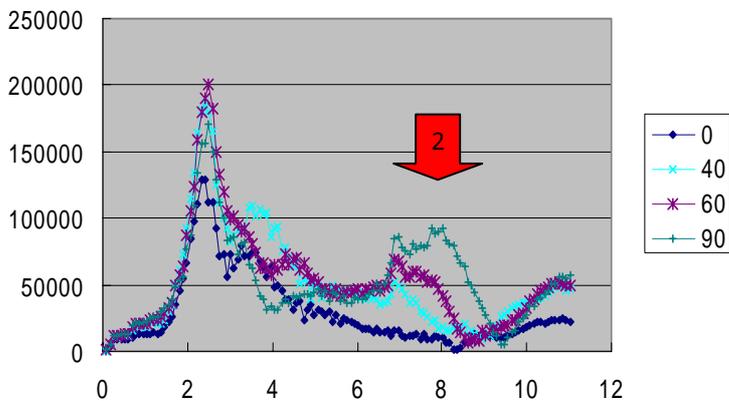


図 8-4 耳介特性

音の回り込み補正

図 8-5 は音の回りこみを示す概略図であり、MIT データ図 8-6 の矢印 3 がこの特性に相当する部分である。従来からの手法では Speaker から逆の耳に聞こえてしまう音 (Cross Talk) をキャンセルするために加えられるもので、左右の音を遅延させ Filter をかけ、逆側の音に加えることにより実現される。ところが、この逆側の耳に加えることはうまくいかないことが判明した。特に Vocal を含む場合には効果が得られない。この理由は 2 つ考えられる。まず第一に Vocal は左右に均等に録音されているため、もともと Vocal には左右がなくそれを不用意に逆側に加えても本来自然界では有り得ない遅延となってしまう、人間の脳が混乱するためと考えられる。第 2 にはこの手法が従来の Speaker 再生を Emulate しようとい

う考えから行われているものであるが、前述したように Speaker の位置に音を定位させても結局は本来の自然界での音の認識とは異なっており、あえて逆側に加える必要もないと考えられる。オーケストラのように Vocal がなく左右で共通な音がない場合にはこれは有効であると考えられる。また、Cross Talk をキャンセルする手法は聴き手の位置/頭の方角で特性が変わってしまうので、実際に実現するとしても様々な制限が出来てしまうという問題もある。

今回は、予定していたように逆側の耳に加えるのではなく、正側の耳に加えることを行った。これは、耳介で回り込んだ音がそのまま遅延されてまた耳に入力されることになる。この特性は今回の解析では重要な役割を演じていることが分かった。

回りこみの信号は外耳道特性と逆の周波数特性でかつ回りこみによる遅延があることにより自然な音となる。この外耳道と耳の回りこみ補正の組み合わせが非常に重要である。

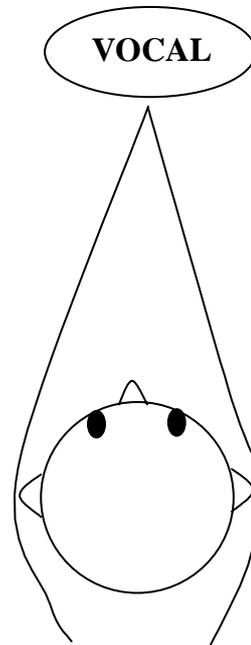


図 8-5 音の回り込み

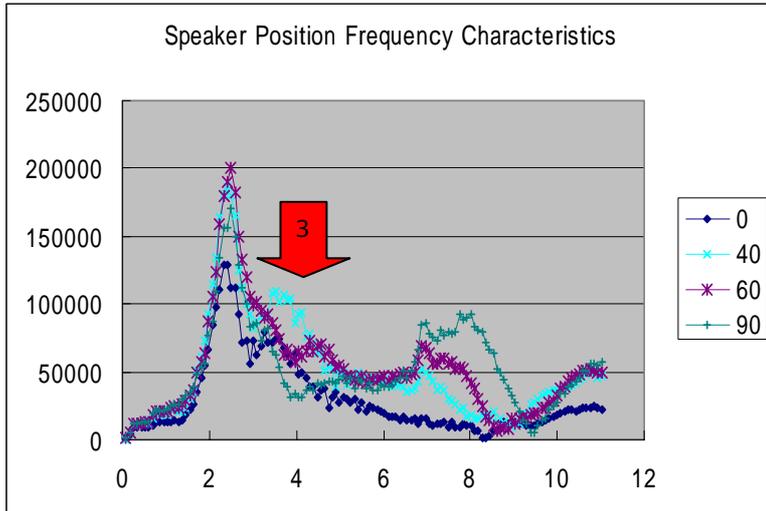


図 8-6 回り込み特性

左右で Incoherent な音にする。
 今回は、耳介で反射して耳に到着する音に対して左右で時間差を与え、これにより少しでも Incoherent になるようにした。最初に耳に届く音に時間差を与えなかったのは、この差により定位自体が移動してしまうためである。
 耳介での反射に左右で 44.1kHz サンプルングで 2 データ程度の遅延差程度の差をつけることで前方定位が強調されることが確認出来た。これは、時間にして 45us、距離にして 1.5cm 程度に相当するもので、人間は非常に小さな時間差でも認識できることになる。

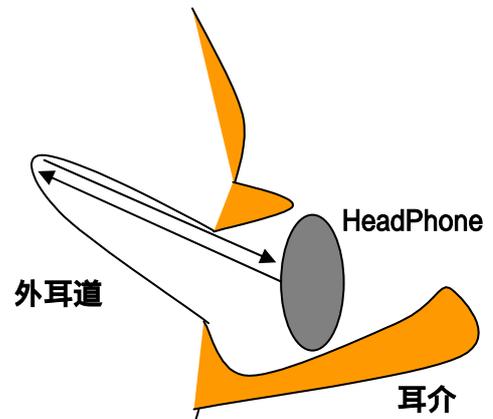


図 8-7 音源と耳の位置関係特性

尚、この補正は当然ながら Head Phone の場合のみ有効であり、Speaker のときには適用しない。

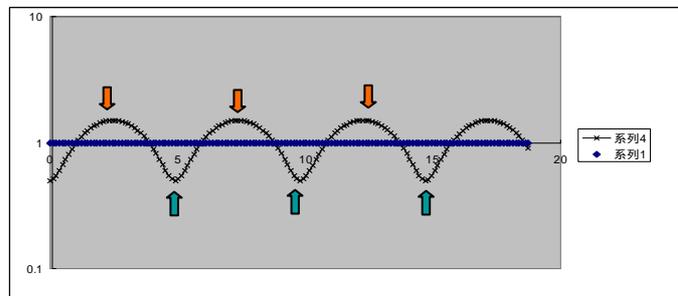


図 8-8 Head Phone の反射/歪補償

Head Phone の歪 / 反射補正
 (Head Phone が耳の中にあるために生じる反射特性の逆特性 Filter の作成)

図 8-7 で示されるように耳の中や外に Head Phone があるために歪/反射が発生してしまう。耳の中の反射は電子通信情報学会 Technical Report of IEICE より報告されており、図 8-9 はその周波数特性である。これは、外耳道が Head Phone でふさがれるため両側をふさがれた筒のようになっているために発生してしまう特性であり、これを打ち消す図 8-8 のような特性をもつ Filter を作成しキャンセルする。

矢印で示される山谷が逆になるような特性となっている。Head Phone にはその形状からいくつかの種類があるが代表的なものは

- Inner Type 耳の中に入れて使用するタイプ
- Outer Type 耳の外に装着されるタイプ

に分けられる。

Inner Type の場合は Head Phone が耳の中にあるため遅延時間が短く、Outer の場合には耳の外にあるため遅延が大きくなることを予想していたが実際には Inner Type とほぼ同じことが判明した。これは、Inner Type と Outer Type では耳の中での Head Phone の位置が異なることから反射する場所も異なると考えられる。

- Inner Type の場合には外耳道の奥での反射
- Outer Type の場合には耳の中央での反射

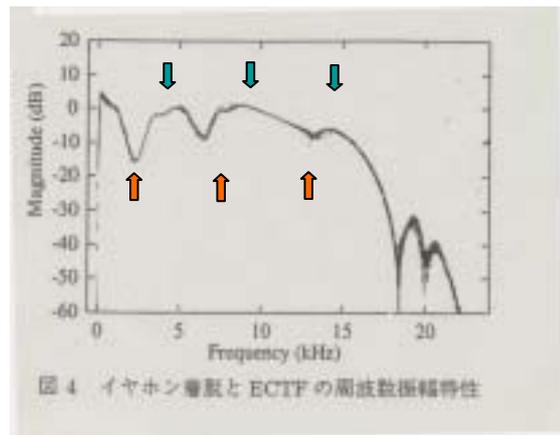


図 8-9 Head Phone の反射/歪

電子通信情報学会
 Technical Report of IEICE
 EA2000-31(2000-10)

最終的な特性は図 8-10 のようになり、Wiener and Ross が測定した音が正面からの時のデータ(図 8-11)とほぼ

同じ特性となっている。
 そのときのシグナルフローは図 8-12 のようになる。

図 8-12 シグナルフロー

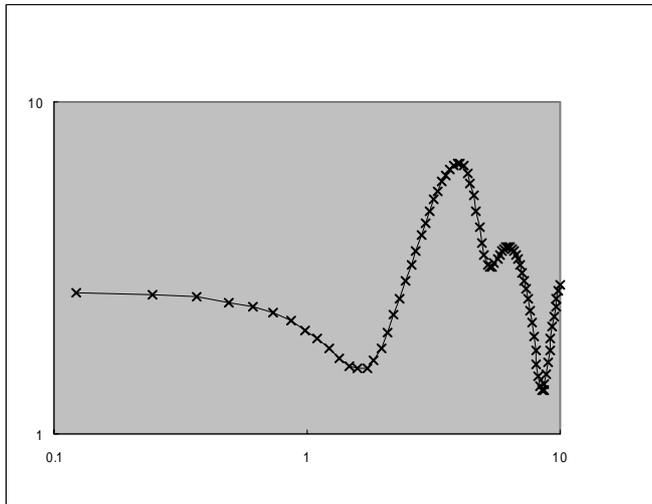


図 8-10 最終特性

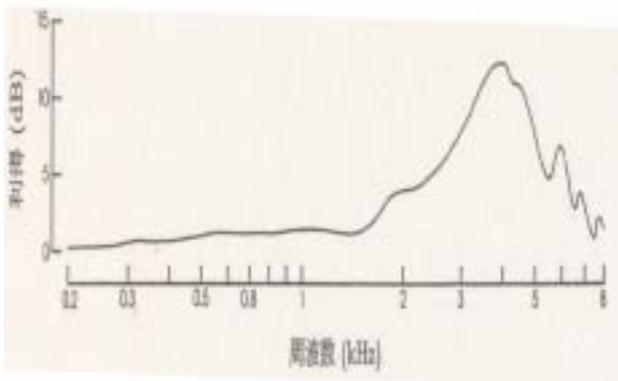


図 8-11 Wiener and Ross 周波数特性

9. MIT のデータを使用しての解析

MIT の Dummy Head 特性は以下のものがある

Dummy Head での Impulse Response (図 9-1)

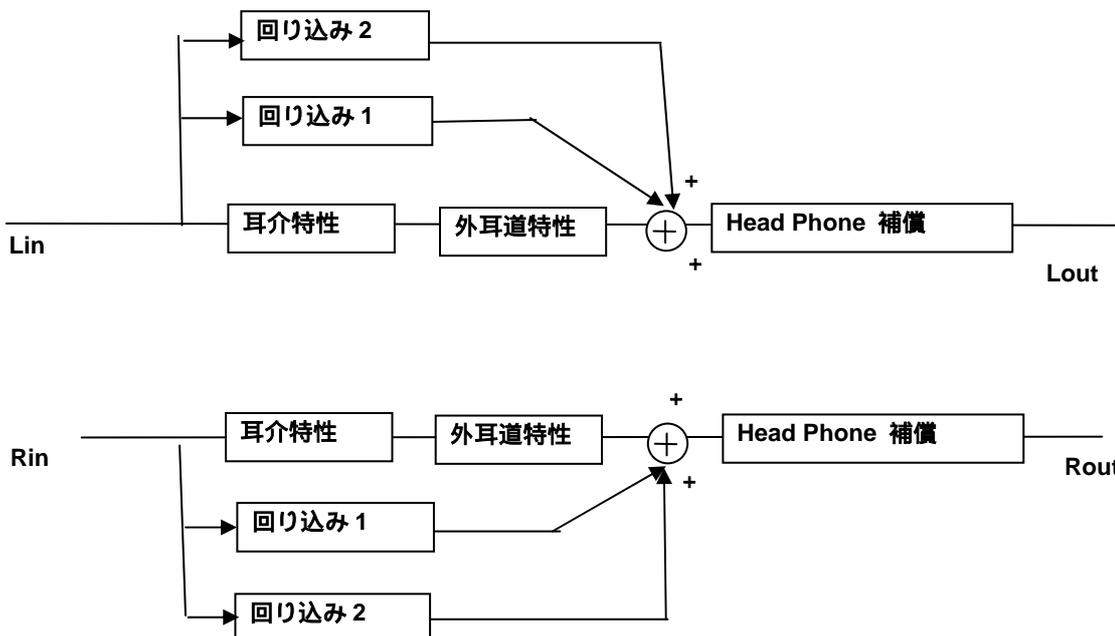
ソースとなった Speaker の Impulse Response

(図 9-2)

今回はこの Speaker の Impulse Response の逆特性を Dummy Head の Impulse Response に掛け合わせることで Time Domain でのキャンセルを試みた。

ところが、途中で発振してしまうという問題が発生した。これは MIT のデータに問題があると考えられる。それは、通常測定には同じマイクを用いなければならないが、MIT のデータでは Speaker Impulse 測定に Dummy Head Impulse を測定したのと同じマイクを使用していない。MIT によると Dummy Head を分解している時間がなかったとのことであるが、このためどうしても特性のずれが発生してしまったのではないと思われる。ただ、発振するまでの特性からは上述したように、今回の解析で得られた特性と同様な傾向が認められた。

人間にとって音の定位認識は、外耳道での定在波/耳介による反射/耳介の回り込みなどの時間的な要素が非常に大きく、これが再現されないとうまく定位されない。いくら周波数特性が似ていてもこれらのキーとなる特性がきちんと再現されるのとされないのでは音の定位は大きく変わってきてしまう。



た方向に音源があると認識してしまうことになる。
 また時間差 10ms、距離にして 3m を超えると、さらにひとつの音源と左右両耳の位置関係ではあり得ない時間となり、ここでやっと反射であると認識するのであろう。

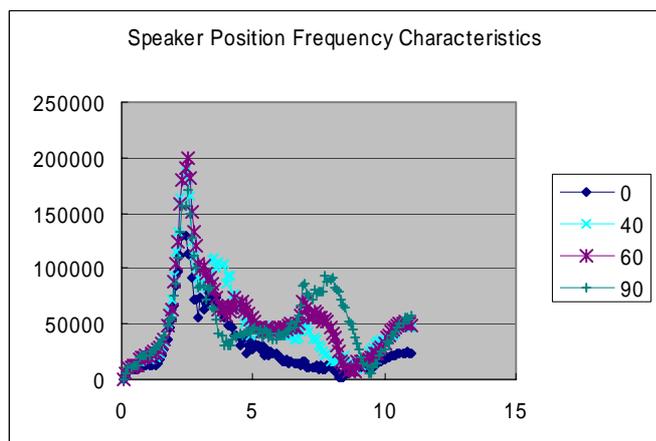


図 9-1 音源に対する耳の位置特性

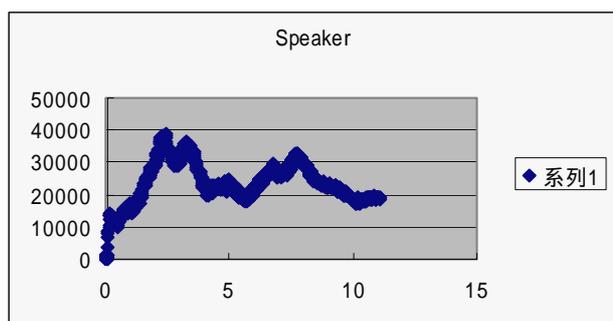


図 9-2 Speaker 特性

10. 遅延の認識

人間は複数の音源から同じ音が出ている場合には聞き手の耳に早く到達する音源方向に音像定位するという Hass 効果、別名先行音効果がある。このため遅延時間が

0ms から 1ms の場合は遅延差で音を定位

1ms から 10ms の場合は先行音の方向に定位

10ms 以上の場合は遅延をエコーと認識

のように定位される。

通常同じ音はひとつの音源から発生し、それを左右両耳で認識するため、この遅延差が人間にどう認識されることが問題になる。

0ms から 1ms の場合は遅延の距離にすると 0m から 0.3m となり、ひとつの音源から左右両耳の関係を考えると、音源の位置を変えることでこの遅延が存在する位置関係が成り立つのでこの遅延値から人間は音の定位を認識できる。

ところが 1ms から 10ms つまり距離にして 0.3m から 3m の場合はひとつの音源と左右両耳との関係を考えると反射を考えない限り通常ありえない位置関係になる。すると人間は音源がどこにあるか認識できずに最初に聞こえ

従来からのサラウンドサウンドは想定するホールや劇場に存在する反射やそれにもなう遅延時間を測定して、それを Digital 信号処理で行なわれる。この遅延値は 10ms 以上であり、人間は反射によるエコーと認識するのである。

一方、今回の検討のように人間の音の定位認識を再現するためにはこの反射を想定した遅延値ではなく、人間が本来音の定位を認識できる遅延時間、つまり 1ms 以内にする必要がある。今回の検討では最大 0.7ms 程度の遅延時間を使用しており、サラウンドサウンドとは異なることがわかる。

11. 本技術の展開

ターゲット市場としては、Head Phone を使用する携帯機器全般になる。つまり、従来からの CD, MD, MP3 のような携帯型の音楽再生装置のみでなく、最近では PDA、携帯電話でも MP3 の音が再生出来、このようなタイプの装置が全て対象となる。ということからビジネス規模は現在どんどん広がっている。

一方、Speaker System である映画やテレビなどの場合には正面に前方の音だけを出力する Speaker を設けるなどの方式もあるが、通常の方式で録音された CD などの音の再生には前方だけのチャンネルはなく、今回の技術を使用することで、音の前方定位が可能となる。

また、車載のような Speaker が人間の横にあるような劣悪な Speaker 位置のオーディオシステムでも正面に音像定位することが可能となる。

12. 参加企業及び機関

有限会社 イクスクラ 新井清嗣

13. 参考文献

[1] 堀内 俊治, 穂刈 治英, 島田 正治, 適応逆フィルタを用いた頭外音像定位に関する検討, TECHNICAL REPORT OF IEICE EA2000-51(2000-10)

[2] 矢野 隆, 等ラウドネス曲線
<http://acoust.arch.kumamoto-u.ac.jp/labnois>

[3] Bill Gardner, Keith Martin, Dummy Head data
<http://sound.media.mit.edu/KEMAR.html>

[4] イェンス ブラウエント, 森本 政之, 後藤 敏幸
 空間音響, 鹿島出版社