両耳聴モデルを用いた 3 次元音源方向定位システムの開発*

堀 畑 聡^{*1}, 片 山 岳 彦^{*2}
三 宅 哲 夫^{*3}, 章 忠^{*4}

3-D Sound Localization Using the Binaural Model

Satoshi HORIHATA^{*5}, Takehiko KATAYAMA,
Tetsuo MIYAKE and Zhong ZHANG^{*5} School of Dentistry at Matsudo, Laboratory of Physics, Nihon University,
2-870-1 Sakaecho-nishi, Matsudo-shi, Chiba, 271-8587 Japan

The robot has evolved to meet various needs in the industrial world and the home flexibly today. However, an intellectual robot has not been developed yet. In the future, the intellectual robot that will think for myself, judges, and acts is needed. One of the important senses added to the robot is aural. It is said, "Sound Localization" to presume the direction of coming among these. It aims at the achievement of three dimension sound localization with two microphones that assume the application of the humanoid robot the auditory system in this research. Then, binaural model using dummy head was made as man's auditory model. In this research, the sound source direction is decided by two angles, one is the horizontal angle α , and the other is vertical angle β . The former is presumed from the difference of the time of arrival of the sound from the source to a right and left ear, and the latter is presumed from spectrum correlation analysis using HRTF (Head Related Transfer Function). HRTF is a transfer function of spectrum modulation caused because the sound is diffracted and reflects by the ruggedness of the head and the earlobe. The directional information of the sound is taken out by using information that system has as knowledge. The direction location experiment was done in the anechoic chamber for some sources. The correct presumption rate became about 90% on the average, and was able to show the possibility of the sound localization by this technique.

Key Words: Sound Localization, Binaural Model, Transfer Function, Head Related Transfer Function

1. 結 論

今日、産業界や家庭における様々なニーズに柔軟に対応するロボットが必要とされている。だが、それらはまだ人間からの一方的な命令で動くものが多く、ロボット自身で考え、判断し、行動するものはない。これからは、状況に応じて様々な業務を行う自律型ロボットが求められている。その自律型ロボットに重要な認識機能である聴覚は視覚に比べて周囲環境の認識に寄与する割合は低いと言われているが、視覚が使えない暗闇や視野外でも音は聞くことができる。つまり視覚の及ばない範囲を聴覚で補うことができる点で、聴覚は周囲環境認識の重要な機能の一つであると言える。

その聴覚機能に人間はある音を聴くとその到来方向や距離を大まかに推定することが出来る機能を持っている。このうち到来方向を推定することを音源方向定位という。この音源方向定位に関する研究は以前より多くの研究者により進められてきた。しかしそれは通常 3 個以上のマイクロホンアレイを用いるもの^{1,2}や、2 個のマイクロホンを用いてもその音源定位が可能な範囲が限定されているもの³⁻⁵であった。ヒューマノイドロボットの聴覚システムへの応用を考えたとき、2 個の固定マイクロホンシステムを使う利点は構造的に単純で小型化に有利であること、2 チャンネル分の信号処理で済むことなどが挙げられる。また、人間は頭部や耳介などのため、同じ音を聞いても音の到来方向によって微妙に違った音に聞こえている。この到来方向によって変化する音の伝達関数のことを頭部伝達関数 (HRTF: Head Related Transfer Function) という。人間の音源方向定位において、HRTF は主に垂直方向の方向推定に利用されており、水平方向の音源方向推

* 原稿受付 2006 年 3 月 27 日。

^{*1} 正員、日本大学松戸歯学部 (☎ 271-8587 松戸市栄町西 2-870-1)。^{*2} 豊橋技術科学大学工学部 (☎ 441-8580 豊橋市天伯町雲雀ヶ丘 1-1) [現: マツダ(株) (☎ 730-8670 広島県安芸郡府中町新地 3-1)]。^{*3} 豊橋技術科学大学工学部。^{*4} 正員、豊橋技術科学大学工学部。

E-mail: horihata.satoshi@nihon-u.ac.jp

定には左右の耳への音の到達時間差が主な情報源となる。人間の頭部形状を模したダミーヘッドはそういった人間の聴覚特性を反映しており、音源の各方向により異なる音響伝達特性を持っている。

そこで本研究では、聴覚モデルとしてダミーヘッドを用いた両耳聴モデルを構成した。また、ここでは水平角 α と垂直角 β の二つの角度を求めることで音源の方向を推定する。その定位原理は人間の音源方向定位と同様に水平方向（水平角 α ）の推定には両耳への音の到達時間差（BTD=Binaural Time Difference）を利用し、垂直方向（垂直角 β ）の推定にはダミーヘッドのHRTFを利用する。解析に際してその準備段階として、白色雑音を用いてあらかじめあらゆる方向に対してダミーヘッドのHRTFを測定し、データベースを作成する。それに加えて定位対象音は一度聞いたことがあるという仮定でその情報を知識として用意する。実際に音を観測したときには知識として持っている音の情報を元に観測音に畳み込まれた方向情報を取り出し、データベースから特定することで音源方向を推定する。音源方向定位実験には3種類の定位対象音（「あ」 $/a/$ 、「い」 $/i/$ 、「う」 $/u/$ ）を用いた。その結果から、本手法の有効性を示す。

2. 原 理

2.1 方向定位 音源方向定位のうち、特に水平方向の定位においては両耳に到達する音波の時間差（BTD）、振幅レベル差（ILD: Interaural Level Difference）が主要な手がかりとなる。

人は頭部が存在するために、周波数 1kHz 以下の音は回折により反対側の耳に到達できるが、周波数 1kHz 以上の音は頭部で反射して直接は反対側の耳に到達できないことが知られている。よって 1[kHz]以下の音はそのBTDによって音源の方向を知り、それ以上の周波数の音は左右の耳に到達する音圧の差(ILD)を検知して方向を知る。ただ、音源位置が正面からずれて側方へ行くほどILDが方向定位に寄与する割合はBTDのそれより低いといわれている。よって、定位対象とする音が低周波を含んだ広帯域信号である場合にはBTDによる手がかりがILDによる手がかりよりも優位に働くとされている⁶⁾。

方向定位のうち仰角方向の定位については主に頭部や耳介での反射によって生み出されるスペクトルの違いが利用されており、これによって音源が正中面上にある場合の上下・前後方向の定位ができる。耳介は音波を集約する役割も持っており、主に 4[kHz]以上の音の方向定位に役立っているとされている。実際に人

間が方向定位を行うときにはこれらの機能のみではなく、頭を動かすことで変化する音の情報を手がかりにしている。頭を動かせば左右の耳と音源との相対的な位置関係が変化し、頭部・耳介の影響で音が鼓膜まで到達するまでに伝播する経路も変わるので音の「聴こえ方」も変化する。この聞こえ方の変化も利用して人間は経験的に音の上下・前後方向を定位していると考えられる。よって、上下・前後の方向定位には頭部や耳介における音の反射や遮蔽が大きな影響を及ぼしていることになる。つまりこれらの影響を受けやすい波長の短い音、即ち高周波が上下・前後方向の定位の支配的要因となっている。これは高周波成分を持たない音源では前後・上下の判定が困難になることを示唆しており、実際に被験者の頭部を固定した状態で低域に帯域制限した信号を聴かせる実験を行うと上下前後方向の誤判定が多くなることが報告されている⁷⁾。このことから、人間は方位角方向については低周波、仰角方向については高周波を手がかりにして方向定位を行っていることがわかる。つまり信号が純音であるよりも様々な周波数が混じった複合音である方が方向定位のための情報が多く、定位しやすいといえる。また、人間は両耳間の時間差の弁別域（分解能）は 10[μ s]、強度差の弁別域は 0.5 [dB]、水平方向の角度分解能は正面方向で約 1° という報告もある⁸⁾。この音に関する角度や時間の弁別域は音源の種類や個人差による変動が非常に大きく、実際の日常環境での音源方向定位にはこれほどの精度は得られないと考えるべきである。

2.2 音源方向の定義 音源の方向は水平角 α と垂直角 β の二つのパラメタで定義する。方位角 ϕ は定位原点から音源を見た時の水平面上の角度であり、図1に示すように定位原点を頂点とした円錐母線上の点は全て同じ水平角である。垂直角 β は水平角 α によって表現される円錐の底面中心E、円錐底面と水平面の交点をHとして、 $\angle SEH$ で表現される角度である。

2.3 方向推定の原理 水平角 α は音源から左右のマイクロホンに音が到達する時間差 BTD を知ることによって推定できる。BTD は左右の耳に入る時系列信号の相互相関関数を計算し、得られた分布のピーク値の時刻から求められる。

垂直角 β についてはHRTFを利用して推定する。本研究ではあらかじめ白色雑音を用いてHRTFを音源の各方向ごとに測定し、データベース化する。データベースとして収録されている音源方向は α が 0°（正面）から 80° まで 10° 刻みで 9 方向、 β は α の角度によって

4～12方向に区切られており、全78方向である(表1, 図2)。これは定位原点から見た音源位置を仮想球面上で均等に配置させるためである。なお、今回はダミーヘッドが左右対称であることから正中面より右側についてのみ実験を行った。また本システムでは原理上、白色雑音を用いたHRTFデータベースとは別に実際に発生される音声である定位対象音をあらかじめ正面から観測し、知識としてデータベース化しておく必要がある。

ϕ : Azimuthal angle α : Horizontal angle
 θ : Elevation angle β : Vertical angle

Fig.1 Direction of sound source.

2.4 BTD と水平角の関係 音源が水平面に存在する場合($\phi = \alpha$)のBTDと水平角の関係について述べる。ここでは議論を簡単にするために図3のようにダミーヘッドは直径 d [cm]の球であると仮定する。音波の伝播経路として、右耳へは音は直接入り、左耳へは点 p までは直進してそこからは顔面に沿って左耳へ到達すると仮定する。定位原点から音源までの距離 r はダミーヘッド両耳間距離 d より十分大きい仮定($d/r \ll 1$)を導入する。このときの、水平角 α とBTDの関係は次式となる。

$$BTD = \frac{d}{2c} (\alpha + \sin \alpha) \quad (1)$$

ここで、 c は音速を表す。

2.5 相互相関関数によるBTDの推定 BTDの推定には右耳と左耳の信号の相互相関関数による時間差を用いる。右耳、左耳に入ってくる時系列信号をそれぞれ

$o_R(t)$, $o_L(t)$ としたとき、その相互相関関数は式(2)の畳込み演算で求めることができる。

$$\gamma_{RL}(\tau) = \int_{-\infty}^{\infty} o_R(t) \cdot o_L(t - \tau) dt \quad (2)$$

ここで τ は両信号の時間差に対応する変数である。実際の計算には、時間領域での畳込みが周波数領域での乗算となる原理を利用して式(2)の高速フーリエ変換を用いて周波数領域で処理し、短時間での計算を行っている。

Table 1 Direction of sound source of HRTF data base

		β [deg]
α [deg]	0~20	0,30,60,90,120,150,180,210,240,270,300,330
	30~40	0,36,72,108,144,180,216,252,288,324
	50	0,45,90,135,180,225,270,315
	60	0,60,120,180,240,300
	70~80	0,90,180,270

Fig.2 Measurement position

Fig.3 Relation between arrival time interval and horizontal angle of sound

2.6 頭部伝達関数 (HRTF) 音源から聴取者の鼓膜までの音の伝達特性を頭部伝達関数とよぶ。HRTFには聴取者の頭部や耳介の形状が影響しており、音が鼓膜に到達するまでの間に回折・反射を起こす。頭部形状は通常、上下・前後方向には非対称であるから音の到来方向によって回折・反射のしかたも異なる。よって HRTF は方向毎に違った特性を示す。本研究で用いたダミーヘッドは人の頭部・耳介の形を模したものであり、人と同様の HRTF を持っている。

音源より放出された音声でダミーヘッドに入る直前の信号を $s(t)$ 、そのフーリエ変換 (スペクトル) を $S(f)$ 、ダミーヘッドのマイクロホン出力信号を $o(t)$ 、そのフーリエ変換 (スペクトル) を $O(f)$ とするとこれらの関係は式(3)、(4)のようになる

$$o(t) = s(t) * h(t) \quad (3)$$

$$O(f) = S(f)H(f) \quad (4)$$

ここで $h(t)$ はダミーヘッドのインパルス応答、 $*$ は $s(t)$ と $h(t)$ の畳込み積分を表す演算子、 $H(f)$ は $h(t)$ のフーリエ変換、すなわち HRTF である。 $O(f)$ と $S(f)$ がわかれば $H(f)$ がわかり、前述の理論から音源方向がわかる。しかし、 $s(t)$ は観測することができないので $S(f)$ も計算できないことになる。

そこで「システムは方向定位の対象とする音源を既知の方向から一度聞いたことがある」、すなわち「対象音源に関する知識 (データ) を持っている」という前提を設ける。そしてその知識として持っている対象音源のスペクトルを $O_0(f)$ (これを基準スペクトルと呼ぶことにする) とし、そのときの音源方向が α_0, β_0 であったとする。またこの方向に対応する HRTF を $H_{\alpha_0\beta_0}(f)$ (これを基準 HRTF と呼ぶことにする) とする。そこで、あらかじめあらゆる音源方向に対して HRTF を測定し、データベースとして作成しておく。そしてある音声を聞いたときに、その音源の到来方向の HRTF をデータベース内から参照し、特定できればその音源の方向を定位できる。

HRTF データベースの作成方法について説明する (図4)。HRTF の測定に際して音源はホワイトノイズを使用する。まず、ダミーヘッドに入る直前の信号 $s(t)$ を測定するために、定位原点に基準マイクを置いて真正面 ($\alpha=0^\circ, \beta=0^\circ$) からのホワイトノイズを観測し、その信号を記録する。次に定位原点にダミーヘッドを置き、真正面からホワイトノイズを試し

てダミーヘッドのマイクロホン出力信号 $o(t)$ を観測する。これらの信号をフーリエ変換したものをそれぞれ $S(f)$ 、 $O(f)$ とすると伝達特性は次式で求められる。

$$H(f) = \frac{O(f)}{S(f)} \quad (5)$$

さらに以下の式により、パワースペクトルの比を求める。

$$H(f) = \frac{O(f) \cdot O(f)^*}{S(f) \cdot S(f)^*} = \frac{Po(f)}{Ps(f)} \quad (6)$$

ここで、 $*$ は複素共役を表し、 $Po(f)$ 、 $Ps(f)$ は $o(t)$ 、 $s(t)$ それぞれのパワースペクトル値を表す。

Fig.4 Calculation method of HRTF

以上の操作で HRTF は求められるが、このとき用いたホワイトノイズは疑似ホワイトノイズであり、その統計学的性質上ランダムな誤差が含まれている。このランダム性による影響を軽減するために $s(t)$ 、 $o(t)$ 両方について連続で約7秒間のホワイトノイズ信号を収録し、その信号から順次1024点ずつを切り出して300個のホワイトノイズ時系列データとした。そしてこれら300個のスペクトルを平均したものの比を式(7)で計算し最終的な HRTF とする。

$$H(f) = \frac{[Po(f)]_{ave}}{[Ps(f)]_{ave}} = \frac{\sum_{n=1}^{300} Po(f)_n}{\sum_{n=1}^{300} Ps(f)_n} \quad (7)$$

左右のマイクロホン信号について式(7)の操作をそれぞれ行い、真正面の HRTF データが作成される。そして次に音源の位置を変えてダミーヘッドでホワイトノイズを観測し、同様に300回の平均化操作を行って HRTF データベースを計算する。

2.7 垂直角の推定の原理 次に音源より到来したある音声を観測・解析してその BTM から仮水平角 α が求められ、そのときの α を α_1 とする。 $\alpha = \alpha_1$ より参照されるデータベースは $H_{\alpha_1, \beta_0}(f)$, $H_{\alpha_1, \beta_1}(f)$, $\dots$, $H_{\alpha_1, \beta_{N-1}}(f)$ である。ここで N は 360° を N 等分したときの分割数であり、垂直角 β の分解能は $360/N$ となる。

次に参照された HRTF データベース $H_{\alpha_1, \beta_n}(f)$ と基準 HRTF ($H_{\alpha_0, \beta_0}(f)$) の比を求め、 $R_{H_n}(f)$ とする。

$$R_{H_n}(f) = \frac{H_{\alpha_1, \beta_n}(f)}{H_{\alpha_0, \beta_0}(f)} \quad (n=0, 1, 2, \dots, N-1) \quad (8)$$

式(8)により、知識として我々が予め持っている方向のみに依存する伝達関数が得られる。本研究ではこの基準 HRTF $H_{\alpha_0, \beta_0}(f)$ を真正面のデータとした。

同様に観測された信号のスペクトル $O(f)$ と基準スペクトル $O_0(f)$ の比を求め、 $R_s(f)$ とする。

$$R_s(f) = \frac{O(f)}{O_0(f)} \quad (9)$$

式(9)により、観測された音が持っている方向のみに依存する伝達関数が得られる。本研究での基準スペクトル $O_0(f)$ は真正面のデータを用いた。

得られた $R_{H_n}(f)$ と $R_s(f)$ の相関係数を式(10)で計算することで、観測音の方向に依存する伝達特性と HRTF により予め持っている方向に依存する伝達特性の類似性を定量的に求めることができる。

$$\gamma_n = \frac{\int_{-\infty}^{\infty} (R_s(f) - \bar{R}_s)(R_{H_n}(f) - \bar{R}_{H_n}) df}{\sqrt{\int_{-\infty}^{\infty} (R_s(f) - \bar{R}_s)^2 df} \sqrt{\int_{-\infty}^{\infty} (R_{H_n}(f) - \bar{R}_{H_n})^2 df}} \quad (10)$$

$$(n=0, 1, 2, \dots, N-1)$$

ここで $\bar{R}$ は f の全帯域における R の平均値を表している。原理的には式(10)で求められた相関係数 γ_n が最大となる方向が真の到来方向に最も近い垂直角 β_n と推定することができる。実際の相関係数の計算は式(10)を離散表現した式から求める。

本研究ではサンプリング周波数 $F_s = 48$ [kHz]、データ点数 1024 点としたので、スペクトルデータ 1 点辺りの間隔 $\Delta f = 48000/1024 = 46.875$ [Hz] となる。サンプリング定理によると理論上は 24 [kHz] (=512 点) までの相関計算が可能であるが、ここでは実験装置の周波数特性がフラットな部分 (200 [Hz] ~ 17 [kHz]) のみで解析する。

右耳と左耳による垂直角の推定を行い、 $\gamma_R(\beta)$ と $\gamma_L(\beta)$ を求める。今回は求められた左右の耳の相関係数 $\gamma_R(\beta)$, $\gamma_L(\beta)$ の各 β 毎の平均相関係数 $\gamma_{AVG}(\beta)$ を式(11)で計算し、この値が最大となる β を最終的な β の推定角度とする。

$$\gamma_{AVG}(\beta) = \frac{\gamma_R(\beta) + \gamma_L(\beta)}{2} \quad (11)$$

3. 実験装置

本研究ではダミーヘッドを用いて音源定位を行う。ダミーヘッドは人間の聴覚特性を模した録音ができるため、音に関する研究やコンサートホールの等の音響設計にも利用されている。実験装置は図5に示すように半ドーム状のフレームを組み、スピーカー側をレールに沿って動かすことで方向位置を与える。音源が水平面より下、即ち $\beta=180^\circ$ 以上の角度についてはダミーヘッドを倒立にセットすることで、下から音が聞こえる状態を作りだしている。供試音はスピーカーより出力され、ダミーヘッドでの観測信号は実時間処理可能なオーディオデバイスを経由して PC に保存され、オフラインで解析を行う。HRTF データベース作成で用いた基準マイクは HRTF の原理上、全周波数域にわたってフラットな特性を持つことが望ましいが、これについては公称再生周波数 15 [Hz] ~ 20 [kHz] で図 6 に示すような特性を持っている。またスピーカーの周波数特性は公称データが無いのでホワイトノイズを再生し、基準マイクでその信号を観測して求められた特性を図 7 に示す。図よりスピーカー側の周波数特性が基準マイクほどフラットではない。つまり定位原点で聴取される音は厳密にはホワイトノイズではない。しかし、本解析方法では信号同士の比をとるのでその特性は相殺されるために影響はないと考えている。但し、200 [Hz] 以下の低域と 17 [kHz] 以上の高域はパワーが落ちているのでスペクトルデータ自体の信頼性が落ちる。よって解析に利用できる区間は 200 [Hz] ~ 17 [kHz] と考えるのが妥当である。

4. 解析方法

実験は周囲雑音の影響を避けるために無響室にて行う。方向定位の対象音 (供試音) は男性の「あ」(/a/)、「い」(/i/)、「う」(/u/) の3種類の音源を対象とした。音はピーカーから出力し、発生した音をダミーヘッドで観測する。その方向は HRTF データベースと同様に正中面より右側 78 方向について行う。観測した信号は USB オーディオデバイス (EDIROL: UA-

1000) を用いて収録, PC に記録した後, 音源方向推定を行う. 音源 (スピーカー) から定位原点 (ダミーヘッド中心) までの距離を 2 [m] とする. 供試音の観測に関しては, 先に入力があった側 (今回の場合は右耳) のチャンネルの振幅が設定した閾値を超えた時点でトリガをかけて計測する.

音の収録範囲はトリガ点から 512 点さかのぼったところから 2560 点を収録する. 図 8 に示すように最初の 1024 点を BTD の検出に利用し, 512 点から 1024 点を 2 区間切り出してそれぞれのパワースペクトルを求める. これは音の立ち上がり部分を収録することで, BTD 検出の精度を向上すること, 無音部分のパワースペクトルへの影響をさけること, 二つのパワースペクトルを平均化することによる波形の安定化などをねらったものである.

BTD の検出分解能はサンプリング周波数に依存する. $F_s=48$ [kHz] の場合の時間分解能は 0.0208 [ms] となり, 本研究で使用するダミーヘッド (両耳間距離約 14cm) においては約 2° の方位角分解能に相当するが, HRTF データベースの関係上 α の推定は四捨五入によって 10° 刻みで行う.

Fig. 5 Experimental apparatus

Fig. 6 Frequency response of ECM8000

Fig. 7 Frequency response of MPC-PS30

水平角 α の推定式は音源から両耳への音の伝播距離の差より理論的に求められるが, ダミーヘッドの表面形状を正確にモデル化する必要がある. 本研究では実験より α と BTD の関係を次式で近似し, α 推定に利用する.

$$\alpha = 1.031 \times 10^5 \cdot \text{BTD} \quad [\text{deg}] \quad (12)$$

実際に理論の α と実験により得られた α との比較を図 9 に示す. 図は横軸に BTD, 縦軸に水平角 α をとったものである. 図の実線が理論値, 黒丸が実験値, 点線が最小二乗法により引かれた近似直線である. この式では理論値と推定値が少し異なる角度もあるが概ね推定には問題ないと考えられる.

β の推定においては各方向のスペクトルの特徴をつかむことが重要である. そのために観測音スペクトル, HRTF スペクトルとそれぞれの基準スペクトルに対してケプストラム分析による平滑化・対数化を用いて波形の特徴を引き出すための処理を行う. ケプストラム分析を用いた理由は将来, 音声認識との組み合わせを考えたときの拡張性を考慮しているためである. また HRTF スペクトル比 R_{hm} と観測音スペクトル比 R_s の相関を計算するとき重要となるのがその周波数範囲である. $F_s=48$ [kHz] であることから観測される信号は 24 [kHz] までとなるが, 今回実験を行った供試音の主な成分は 5 [kHz] 以下に分布している. よって相関を計算する周波数範囲に関しては, 最初 5 [kHz] 以下の範囲とした. 実際の解析では 5 [kHz] までの範囲内でいくつかの異なる相関周波数範囲を設定し, より正解率が高くなる周波数範囲の検討も行った.

Fig.8 Data point in collection of test sound

 Fig.9 Relationship of horizontal angle α and BTD

5. 解析結果

5.1 垂直角 β の推定結果 図 10 に α の角度毎にまとめた供試音「あ」についての垂直角 β の定位正解率 (5 回の実験の平均) を示す。このときの相関周波数範囲は 234.375 [Hz] ~ 4125 [Hz] とした。結果から、比較的角度 α の小さい方が正解率が低く、角度 α の大きい方が正解率が高くなった。「あ」の平均正解率は約 87% となり、ここでは示さなかったが、「い」と「う」についてはそれぞれ 82%, 84% の正解率となった。表 2 に誤判定の多い方向と、その誤推定先を示す。表中の矢印の先が誤判定先方向であり、例えば α

$=0^\circ$, $\beta=90^\circ$ のときは $\beta=270^\circ$ と判定してしまうことを意味している。結果より全ての α において、ダミーヘッドの上下・前後に対称な方向へ誤判定を起こす可能性が高いことがわかった。このような上下・前後の誤判定は人間でもよく見られる傾向であり、この誤判定の解消には高周波成分が大きく寄与していると考えられる。ここでは示していないが特に高周波成分の少ない「う」は正面付近での正解率が低い結果となったことは、上記の理由によると考えられる。これは相関を求める際の周波数範囲が影響していると考えられる。そこで、相関周波数範囲を最適化することで正解率の向上を検討した。供試音「あ」についての周波数範囲の変更は上限を 4687.5 Hz とした。変更後の結果を図 11 に示す。図 10 と比較すると各水平角 α において正解率が向上していることがわかる。また、結果より $\alpha=30^\circ$ 以下と $\alpha=80^\circ$ のところで正解率が低下していることがわかる。 $\alpha=80^\circ$ において正解率が低下しているのは、 $\beta=180^\circ$ において水平角の誤推定が発生することに影響されている。これについてはダミーヘッドの形状の影響で BTD が小さくなっていることが原因と考えられる。しかしながら、原理的には α が大きければ BTD が大きくなるが、実際の計測においては α , β の変動が影響しているとも考えられる。現在、この問題の明確な原因は検討しているが、実用上は $\alpha=80^\circ$ はかなり極端な音源方向のため、両耳間の音圧差などを用いることで補足できると考えている。 $\alpha=30^\circ$ 以下に関してはダミーヘッド下側で、前後に対称な方向で誤推定する確率が非常に高い結果となった (例えば $\beta=300^\circ$ を $\beta=240^\circ$ と誤推定する)。これはダミーヘッドを下方向から見たときに前後での形状の変化が乏しく、HRTF スペクトルがよく似ているためと考えられる。「あ」の全体の正解率も約 93% となり最適化前の結果から比べ若干向上した。「い」と「う」についてもそれぞれ最適な周波数範囲の上限を設定した結果、それぞれ 85%, 87% の正解率となった。各音源とも相関周波数範囲の変更によって正解率は若干向上したが、その設定方法については各音源のスペクトル波形に注目し、振幅がある程度大きい範囲を任意に選び出して試行錯誤的に行ったものである。よって、相関周波数範囲の最適化手法の確立は今後の課題である。

5.2 断続音を想定した方向定位 これまでは音源が単発の場合を対象にして解析を行ってきた。しかし単発の音は人間でも方向定位が難しくなる。そこで非常ブザーのように断続的に発せられる音に対する本システムの有効性を調べた。この場合、同じ位置から

発生する音を複数回授聴することが可能になり、音源の情報が増加して定位の精度は向上すると考えられる。この状況と同等の条件として同じ音を5回聴いたことを想定し、「あ」(/a/)について5回の実験で得られたスペクトルをシステムに与え、そのスペクトル波形を平均化して解析を行った。

Fig. 10 Correct answer rate of location according to horizontal angle of voice /a/

Table2 Direction location of mis-presumption

		Misjudgment of Vertical angle β [deg]
α [deg]	0	90 $\rightarrow$ 270, 300 $\rightarrow$ 240
	10	300 $\rightarrow$ 240
	20	300 $\rightarrow$ 240
	30, 40	288 $\rightarrow$ 252
	60	300 $\rightarrow$ 240

「あ」を複数回授聴した場合の結果を図12に示す。図より、20°と30°を除いては100%の正解率となった。また、平均の定位正解率は97.4%となり、単発音の定位に比べて定位精度の改善が確認できた。これは信号情報量の増加によってスペクトル波形の変動の影響が低減されたためと考えられる。ただし、 $(\alpha, \beta) = (20, 300)$ については誤判定を起しており、その誤判定先は $(\alpha, \beta) = (20, 240)$ である。つまり表3に示した誤判定の傾向は残っており、前後方向の判定は比較的難しいことがわかる。

Fig. 11 Location correct answer rate of horizontal angle according to changed frequency range

Fig. 12 Correct answer rate to two or more repetition sounds

6. まとめ

本研究ではダミーヘッドを用いて音の両耳間到達時間差および頭部伝達関数(HRTF)スペクトルの相関解析による音源方向定位を試みた。3種類の音源について供試実験を行った結果、定位正解率が高いもので93%、低いもので82%となり、本手法による方向定位の有効性を示すことが出来た。但し、定位対象音によって相関をとる周波数範囲などの解析上のパラメータを変更する必要があること、特定の方向($\alpha=30^\circ$ 以下および $\alpha=80^\circ$ など)について誤推定が多くなることがわかった。前者の問題に関しては今後音声認識との組み合わせによって改善できる可能性がある。後者の問題に関しては単純に周波数スペクトルの相関に注目するだけでは困難なので、位相や音圧、両耳間の信号差などの他の情報を用いる必要がある。また、本研究は無響室にて実験をおこなったため、室内の反射や雑音の影響について検討していない。そのため実空間で

の適用には無響室の場合、100%近い正解率が理想である。今の段階での最低の正解率 82%はまだまだ実用的には不十分であると考えられる。今後はシステムの改善と正解率の向上、実空間での実験による反射や雑音の影響の検討が課題である。

音は高周波ほど遮蔽されやすく、方向による特徴も高周波の方に現れやすいが、今回用いた供試音については比較的低い周波数範囲に分布していることが推定正解率に影響を及ぼしていると考えられる。また、ここではただ1つの音源のみであったが、複数の音源が存在する場合は音声認識法や話者認識法、独立成分分析などを前段階で使用することで特定の音源を抽出し、方向推定を行うシステムの構築が必要である。

本手法は観測音をすでに聞いたことがあると言う前提が必要となる。これは、異常音への適用に対しては対象音のみの必要とすればよく、また、ロボティクスへの適用ではいろいろな音のデータベースを作成する必要がある。なお、実用化においてはデータベースをどのように構築するかが問題となる。この問題に関しては、音声認識技術を組み込むことで解決することを考えている。

よって、今後は他の解析手法の検討を含めた正解率の向上と一般的な環境への対応として任意の音源に対する方向定位を目指す。

文 献

- (1) Sasaki, K. and Hirata, K., 3D-Localization of A Stationary Random Acoustic Source in Near-Field by Using 3 Point-Detectors, *Transactions of the Society of Instrument and Control Engineers*, Vol.34, No.10(1998), P.1329-1337.
- (2) Kunita, M. et al., Pan-Tilt Camera Control using Sound Source Localization, *IPSI SIG-CVIM, Computer Vision and Image Media*, 133-6(2002), pp.41-48.
- (3) Nakashima, H. and Mukai, T., A Learning System for Estimating the 3-dimensional Sound Source Position using Binaural Hearing, SI 2004, *Proceedings of The 5th SICE System Integration Annual Conference*, (2004), pp. 1102-1103.
- (4) Nakashima, H. et al., Improvement of Sound Source Localization Ability Using Inter-aural Level Difference Information, *IEICE, D-II, Vol.87*, No.3(2004), pp.919-922.
- (5) Nakashima, H. et al., Self-organization Learning of Localization Ability Using Spectrum Information, SI 2003, *Proceedings of The 4th SICE System Integration Annual Conference*, (2003), pp.746-747.
- (6) Itoh, M. et al., Three-Dimensional Sound Image Localization by Interaural Differences and the Median Plane HRTF, *Technical Report of IEICE*, EA2001-39 (2001), P. 17-23.
- (7) Kimura, D. and Suzuki, Y., A consideration about the effect of head movement on the sound localization, *Technical Report of IEICE*, EA2001-44 (2001), P.57-64.
- (8) Oda, R., Evolutionary biology of binaural processing, *The Journal of the Acoustical Society of Japan*, Vol.58, No.3(2002), pp.193-198.