

## 音響テレプレゼンスロボットにおける頭部形状簡略化の音響的・知覚的影響

### Acoustical and Perceptual Influence from Simplifying Head Shape of an Acoustical Telepresence Robot: *TeleHead*

○戸嶋 巖樹 (NTT コミュニケーション科学基礎研究所)

青木 茂明 (金沢工業大学)

\*Iwaki TOSHIMA (NTT CS Lab.), Shigeaki AOKI (Kanazawa Ins. of Tech.)

[toshima@brl.ntt.co.jp](mailto:toshima@brl.ntt.co.jp), [aoki\\_s@neptune.kanazawa-it.ac.jp](mailto:aoki_s@neptune.kanazawa-it.ac.jp)

**Abstract**— We built an acoustical telepresence robot named *TeleHead*, which has a user-like dummy head and whose movement is synchronized with the user's head movement in real time. An accurate-shape user-like dummy head improves sound localization accuracy, but making an accurate-shape user-like dummy head for all users is not realistic. We are trying to simplify *TeleHead*'s head shape by taking the effect of head movement into consideration. In this work, we made two types of simplified dummy heads, a ball-like dummy head and a ball-like dummy head with a user-like pinna. At first, we compared HRTF between dummy heads. Dummy heads are acoustically different between each other. Effect of pinna is large and in case of using user-like pinna, HRTF is more similar than the other. Then, we used the dummy heads in sound localization experiments. The experimental results show that the pinna is very important for sound localization in the median plane. Head movement can improve sound localization and subjects can localize sound with another person's pinna. The results indicate the possibility of using a ball-like dummy head with a generic pinna for acoustical telepresence robots.

#### 1. はじめに

究極の通信技術は、使用者があたかも遠隔に居るかのように感じる技術であり、それをテレプレゼンス技術と呼ぶ[1]。テレプレゼンスロボットとは、そのテレプレゼンス技術を支えるために、遠隔において使用者の代わりに動くロボットである。使用者が完全に遠隔に居るかのような感覚を得るためには、使用者が遠隔に意図通りの作用を及ぼすことと、使用者に遠隔の感覚を正確に伝えることの、双方の技術を実現する必要がある。遠隔に身体を持つことは、遠隔において環境に物理的に作用できる可能性を意味する。他のあらゆる信号処理的通信手段は、原理的にこの可能性を持ち得ないか、または、目的とする環境への作用を達成するための新たな装置が必要と考えられる。したがって、テレプレゼンスの実現にはロボット技術が重要な役割を果たすと言える。

これまでに、力覚や触覚を中心として様々なテレプレゼンスが試みられてきた。筆者らは、環境理解という意味において重要な役割を果たし、また、コミュニケーションにおいても重要な位置を占める聴覚・音響のテレプレゼンスロボットの製作を試みている。聴覚は、空間的には全方位に対して機能し、周囲への警戒をはじめとする環境の把握に重要な役割を果たしている。このように、聴覚による音響環境の理解は、環境把握にとって重要であり、テレプレ

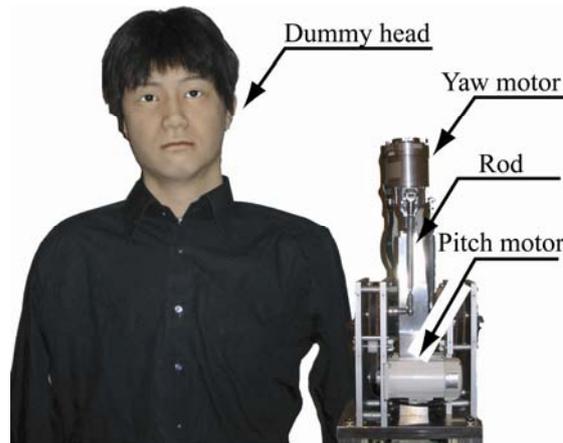


Fig. 1 Acoustical telepresence robot: *TeleHead*. It has a user-like dummy head and synchronizes with user's head movement in three degrees of freedom.

ゼンス技術によって環境を伝達しようとするとき、音響環境の伝達は欠かせない。

身体を持つことの利点は、遠隔において環境へ直接影響を与えることのみではない。人間は自らの身体を通して環境を理解する[2]。音響環境の理解においても、自己の音響的性質の果たす役割は大きい。特に自己の頭部形状に由来する頭部の音響的性質である、頭部伝達関数 (Head-Related Transfer Function: HRTF) は音源の方向同定、すなわち音像定位において大きな役割を果たすことが知られている[3]。さらに、有効な手がかりは受動的な身体情報だけではない。積極的に頭部を動かすことも音像定位を行う上での重要な手がかりとなることが知られている[4, 5]。頭部運動を考慮したバイノーラルシステムも考案されている[6-8]。このように遠隔に身体を持ち、能動的に動くことが出来るロボット技術は様々な側面でテレプレゼンス技術の実現に欠かせないものである。

筆者らはダミーヘッドを使用者の頭部運動に追従動作させることにより、頭部形状および頭部運動の双方の効果を得ようと考えている。これまでに、使用者と同形状の頭部を持ち、使用者の頭部運動に追従する、音響テレプレゼンスロボット、テレヘッドを製作した(図1)。また、頭部形状および頭部運動の効果を音像定位実験で確認した[9]。また、ロボットを用いる方法は、音響環境の伝達という側面だけを見ても、HRTFやダミーヘッド録音を利用した音響環境

の再現手法[10]と比較して、音源の性質や運動を事前を知る必要がなく、メリットがある。しかも、HRTFを測定することの困難さ[11]、およびその困難さを克服し、容易な測定を可能にするために必要な専用の測定系[12]、あるいはHRTFの個人性による定位感の低下問題[13, 14]も回避することができる。以上のように音響テレプレゼンスの実現にロボティクスを用いることは、現実的かつ多くのメリットがある。

一方で、ロボティクスを使うことのデメリットも考える必要がある。筆者らはこれまでに、騒音や追従軌道、遅延の問題について、それぞれ議論を行ってきた[15, 16]。本稿では頭部形状を使用者個人にカスタマイズして製作しなければならない点に注目し、どの程度の簡略化が現実的に可能であるか検討する。まず、簡略化ダミーヘッドのHRTFについて、実頭と比較する。次に、そのダミーヘッドを用いて、音像定位実験を行い、頭部運動の有無や簡略化の条件を変化させ、頭部形状簡略化の可能な範囲について考察する。

## 2. テレヘッド

### 2.1 テレヘッドの概要

テレヘッドの外観は図1に示したとおりである。使用者の頭部形状を象って製作したダミーヘッドが3個のモータによって、Yaw, Roll, Pitchの3方向に駆動するようになっている。テレヘッドの概要を図2に示す。赤い矢印で表されているのが、音の流れである。ダミーヘッドの外耳道入り口で集音した音信号をアンプとヘッドホンを通じてそのまま使用者に提示する。使用者のHRTFの効果が、使用者に酷似したダミーヘッドによって自動的に加味される仕組みである。青い矢印は頭部運動情報の流れである。ヘッドホンに3次元姿勢センサとしてFastrak (Polhemus) を装着し、120 Hzで使用者の頭部姿勢を検出する。テレヘッドはこの情報に基づいてダミーヘッドをPCによって制御した。制御周期は10 msである。騒音は人間の聴感度の高い1-4 kHzの範囲で最大24 dB SPLであり、静かな図書館程度の騒音である。また、頭部運動に対する追従性能は安定しているが、遅延時間が約80 msある。遅延時間と騒音が現状のテレヘッドではトレードオフの関係にあり、聴覚の心理物理実験を行う都合上、遅延時間を長くし、騒音を最小化した。

### 2.2 簡略化ダミーヘッド

本稿で論じる簡略化ダミーヘッドの形状について、図4, 5に示す。図4の写真のように、ラグビーボール状の本体の側面に耳介を貼り付けた形で製作した。本体の幅155mmは図1の写真で用いた精密に使用者の頭部形状を再現したダミーヘッドの両耳間の幅と等しい。正面は厚さ約1mmのシリコン、内壁は厚さ約4mm程度のFRP樹脂である。また、図5の中心部分はマイクロホンを接続するための音響インピーダンスの高い素材部分であり、ロボットの振動がマイクロホンに音として伝わることを防止する。耳介部分は取り外しや変更が可能な構造になっており、外した場合は図5の様な耳介無しのダミーヘッドとなる。図1の写真にあるような精密ダミーヘッドは頭部全体の型取り、MRIや光三次元計測に基づく修

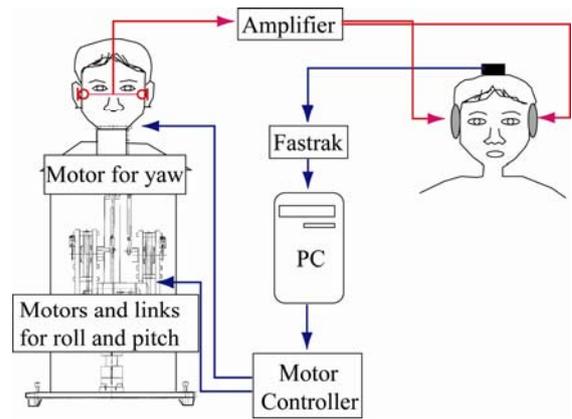


Fig. 2 Outline of *TeleHead*. *TeleHead* is synchronized with the user's head movement and the sound collected with microphones in the dummy head is transmitted to the user by headphones. Blue lines are the flows of head posture data. Red lines are the flows of acoustical signal.

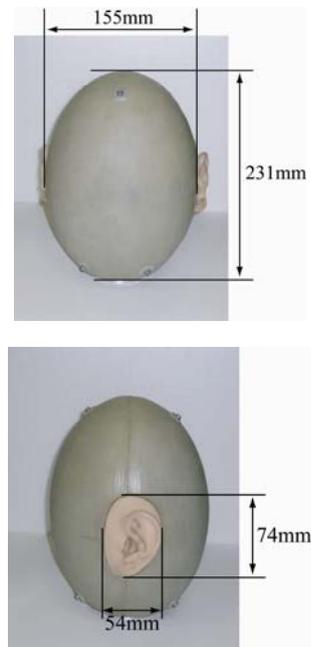


Fig. 4 Simplified dummy head with accurate pinna. Its front view (top panel) and side view (bottom panel).

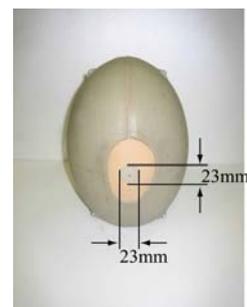


Fig. 5 Ball-like dummy head (no pinna).

正、が必要であり、計測のみでは制作不可能であるし、型取りのみでは十分な効果が得られなかった[17,18]. これらの手順による製作には、高いスキルをもった職人が必要となる。一方、今回制作したダミーヘッドは、耳介つきのもので、特に本人の耳介を模した耳介を使用する場合は耳介の型取りが必要となるが、それ以外には使用者の頭部の写真数枚で十分である。型取りによる耳介の制作は使用者にとっては数10分の負担であり、制作者にとっても1日程度の負担で制作可能である。従って、本稿で取り上げる簡略化ダミーヘッドは、精密なダミーヘッドと比較して、格段に制作過程が簡略化されたものである。以降、まず音響的性質の再現性について測定し、次に知覚的効果について実験した結果を述べる。

### 3. HRTF 測定

#### 3.1 測定方法

実頭とダミーヘッドの頭部伝達関数 (HRTF) は、頭部の中心から音源までの距離を1.2mとして、無響室内で測定した。HRTFは音源から自由音場における頭部中心位置までの伝達関数  $H_{sp-center}$  と、音源から左右の外耳道入り口までの伝達関数  $H_{sp-l} \cdot H_{sp-r}$  の比で表し、周波数  $\omega$  と頭部中心位置から見た音源の相対位置 (方位角, 仰角, 距離) =  $(\theta, \phi, r)$  を変数として、式(1)の様に表される。

$$HRTF(\omega, \theta, \phi, r) = \frac{H_{sp-l \text{ or } r}(\omega, \theta, \phi, r)}{H_{sp-center}(\omega, \theta, \phi, r)} \quad (1)$$

実頭の測定では、測定方位は全方位角と仰角-40~90°で合計143点を測定点とした。各測定点は、正中面と水平面は10°おきに、その他の点は隣り合う測定点との間の仰角と水平角が最大でも20°以内に収まるように設定した。これは、HRTFの測定時間を90分以内に納めて、被験者の負担を減らすためである。ダミーヘッドは長時間の測定に耐えられるので、全方位角と仰角-40~90°で5°おきに1873点、ないし10°おきに469点を測定点とした。実頭の測定点はダミーヘッドの測定点のサブセットとなっている。

HRTFの測定には、実頭やダミーヘッドの左右の外耳道入り口付近をシリコン印象材でそれぞれ型取りした耳栓に装着した小型コンデンサマイクロホン (Panasonic, WM62-AT102) を用いた。即ち、実頭もダミーヘッドも外耳道をマイクロホン付きの耳栓で塞ぐ状態で HRTF を測定した[10]。なお、いずれの場合も、測定を開始する前に、レーザーポインターを用いて両耳珠と鼻頭の位置が常に同じ位置になるように位置あわせを行った。ダミーヘッド測定時にはテレヘッドの駆動部等の電源は落とした。また、測定の最初と最後に正面からのHRTFを測定し、大きな誤差が途中に発生していないことを確認した。音源はサンプリング周波数 48 kHzの最適化引き延ばしパルス (TSP) 信号[19]を使用し、10回の平均をとった。HRTF は512点のFFTで算出した。

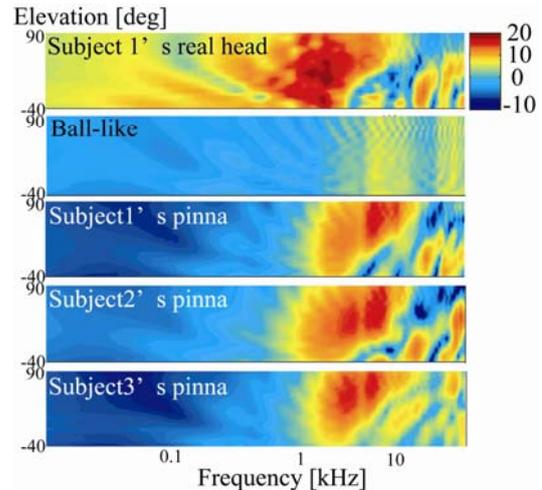


Fig. 6 Results of HRTF measurement for subject 1's real head, the ball-like dummy head, that with subject 1's pinna, that with subject 2's pinna, and that with subject 3's pinna.

#### 3.2 測定結果

HRTFの測定結果を図6に示す。正面のHRTFを表している。横軸は周波数であり、縦軸は仰角である。正面を0度とし、-40度から90度 (真上) までのゲインが色で示されている。暖色の領域でゲインが高い。音像定位の観点からは、6~13kHzあたりに斜めに伸びている低ゲイン領域の形状が重要であるとされている。一番上の実頭の測定結果と比較して、上から2番目の耳介無しの例は大きく異なっている。また、その他の3枚は耳介有りの簡略化ダミーヘッドのHRTFを表しており、3段目のものは、実頭の耳介を象った耳介を装着したものである。耳介形状が等しいダミーヘッド・実頭間のHRTFは比較的近いことが分かる。

次に、HRTFをスペクトル差で表したのが表1である。スペクトル差は式(2)で示される。

$$D_{HRTF}^{FFT} = \sum_d (\sqrt{\sum_{\omega} (|H_i - H_j|)^2 / N_{\omega}}) / N_d \quad (2)$$

ここで、 $d$  は方向、 $\omega$  は周波数、 $H_i, H_j$  はそれぞれ、周波数と方向についてそれぞれの測定点数、 $H_i$  および  $H_j$  は  $H_i(\omega, d)$ 、 $H_j(\omega, d)$  を省略したものである。

実頭を RH, 簡略化ダミーヘッド DH で表している。RH や DH の後ろの数字は耳介形状であり、RH1 の耳介形状を模した耳介を装着した簡略化ダミーヘッド

Table 1 Spectral differences between HRTFs [dB]

	RH2	RH3	DH1	DH2	DH3
RH1	7.38	7.90	7.50	7.87	7.66
RH2	-	8.67	8.85	8.48	8.30
RH3		-	8.23	8.30	6.42
DH1			-	6.79	6.63
DH2				-	6.59

が DH1 である。また、耳介無し の簡略化ダミーヘッドは DH0 と表記することにする。表 1 には載せていないが、DH0 は RH と約 10dB の差があり、図 6 から明らかであったが、DH0 は実頭と音響的に大きく異なる形状であると言える。逆に DH1, DH3 は音響的に近い HRTF となった。

#### 4. 音像定位実験

音像定位実験とは、被験者がある条件下において提示された音に関して、その音源の方向を正しく把握しているかどうかを調査する実験である。スピーカ等の実音源を用いる場合は、スピーカ位置と回答位置が近いほど音像定位精度が高いと結論できる。また、HRTF を用いた仮想音源を用いる場合は、実験者が意図した音像位置を音の提示位置と考えて音像定位精度を測定する。聴覚にとって、音像定位機能は最も基本的機能の 1 つと考えられるため、音像定位実験の結果は音刺激を提示するシステム全体としての質の定量的評価と言える。実音源や仮想音源といった条件に依存せず、人間が直接に音を聴いた場合と同程度の音像定位精度が得られる系があるとすれば、音像定位に関して十分に高品質な音響バーチャルリアリティを実現した系であると結論できる。また、音像定位精度の比較により、系の質を比較することも可能である。

##### 4.1 実験方法

実験の写真を図 7 に示す。図 7 から分かるように、本稿では、正中面にスピーカを配置した音像定位実験を行った。頭部形状の簡略化は HRTF が使用者とロボットで不一致となるといふ影響をもたらす。これは主として正中面の音像定位精度に影響を与えると考えられるため、頭部形状簡略化の影響を測るには正中面の音像定位実験を行うことが妥当と考えた。スピーカは 15 度間隔で -45 度から 75 度まで 9 個を設置した。ただし、-40 度から 80 度までの場合や、-50 度から 70 度までの場合もあり、被験者はこれを知らない。ダミーヘッドの中心からスピーカまでの距離は 1.2 m とした。この距離は頭部近傍における HRTF 変化の影響を受けない距離である [17]。刺激音は、持続時間 5 s、音圧 65 dB SPL 程度の白色雑音で、刺激間隔は 8 s とした。再生毎に生成し直し、また、音量も  $\pm 5$  dB 程度の範囲で変化させることで、音質、音量等による学習効果を回避した。

実験条件は頭部形状と頭部運動について変化させた。音像定位には頭部形状と共に頭部運動が深く関わっていることが分かっており、頭部形状の簡略化も頭部運動と切り離して考えるのは合理的でない。被験者は 3 名、いずれも成人男性 (30 代 2 名、40 代 1 名) で、聴力は事前にオーディオメータで測定し正常であることを確かめた。頭部形状の条件は 3 通り。ただし、ダミーヘッドは耳介のみ精密に再現したものを、各被験者に対して製作したため、合計 4 体である。

1. ラグビーボール状のダミーヘッド (DH0)
2. DH0 に被験者 1, 2, 3 の耳介を接続したもの (DH1/2/3)



Fig. 7. Photograph of the setup for sound localization experiments. TeleHead with simplified dummy head is set in the anechoic room. Loudspeakers are set in the median plane from -45 deg to 75 deg at intervals of 15 degrees 1.2 m in front of TeleHead.

##### 4.2 実験結果

実験結果の生データを図 8 に示す。頭部運動ありの条件である。最上段は各被験者の耳介形状を模した耳介を接続したダミーヘッドによる結果、最下段は耳介無しのダミーヘッドによる結果である。中段のブロックはそれ以外、つまり、Subject1 にとっては、DH2 と DH3、Subject2 にとっては、DH1 と DH3 など、他者の耳介を装着したダミーヘッドを用いた結果である。横軸は音刺激の提示方位 (角度) であり、縦軸は回答方位である。いずれも正面を 0 度とした。マイナスは下方向、プラスは上方向である。従って、提示方位と回答方位が一致する場合、即ち正解は常に対角線上となる。各被験者において、元々の定位精度が異なることもあり、生データにはばらつきが見られるが、傾向として、最下段の DH0 を使用した場合の定位精度が低いことが読み取れる。また、いずれの被験者においても最上段、つまり、被験者の耳介形状を模した耳介を装着した場合の定位精度が高いことも読み取れる。提示方位と回答方位の一致度を相関係数で表したものを図 9, 10 に示す。図 9 は被験者毎、図 10 は被験者間をまとめて検定した結果も示しているが、その結果も生データからの印象と一致している。

##### 4.3 精密ダミーヘッドとの比較

テレヘッドにおいては、被験者の頭部形状を精密に模したダミーヘッドも用意している。図 1 は精密ダミーヘッドを装着した写真であり、被験者 1 の精密ダミーヘッドを DH1a と表す。精密ダミーヘッド

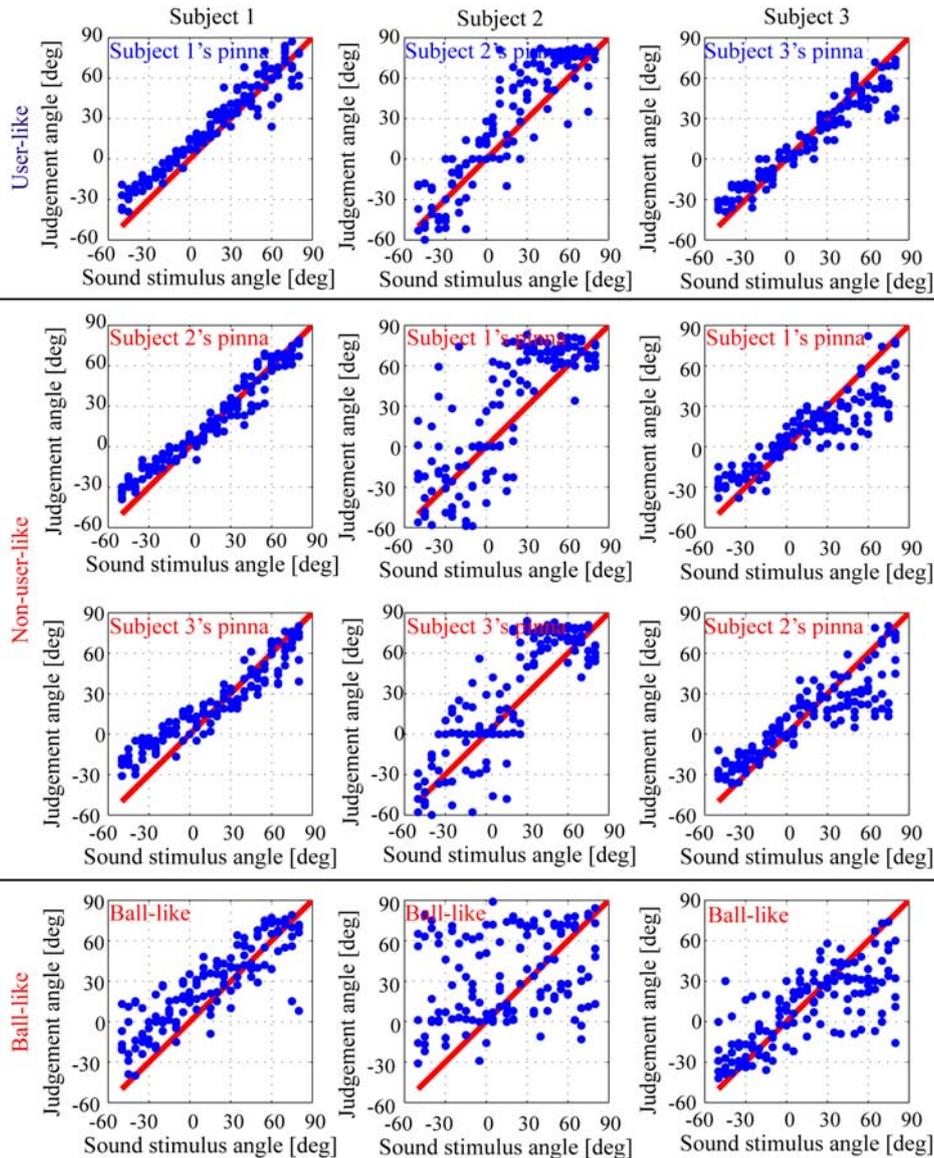


Fig. 8. Results of sound localization experiments in synchronized (head-movement) condition. Top panels show the results using dummy heads with a user-like pinna. Bottom panels show the results using the ball-like dummy head. The others show the results using the dummy heads with a nonuser-like pinna. Left column shows the results for subject 1, center column shows those for subject 2, and right column shows those for subject 3.

と簡略化ダミーヘッドの差を評価する。DH1aは、音響的(HRTF)にはRH1と約4.6dBの差異であった。また、音像定位実験の生データを図11に示す。この時の、正中面音像定位における相関係数は頭部運動無しで0.94、有りで0.98であった。この精度は、実頭で直接聞いた場合の定位精度と概ね同程度である。このように精密ダミーヘッドは確かに効果がある。しかし、簡略化ダミーヘッドは、特に使用者の耳介を装着した場合で、しかも頭部運動がある場合について、精密ダミーヘッドに近い性能を発揮する可能性があることが分かる。

## 5. おわりに

1. ラグビーボール状の簡略化ダミーヘッドを制作した。
2. 簡略化ダミーヘッドには耳介の装着が可能であり、使

用者の形状を模した耳介を作成し、HRTF測定と音像定位実験を行った。

3. HRTFに関して、使用者の耳介を使用した簡略化ダミーヘッドの場合は、実頭の場合と近い数値となった。
4. 音像定位実験の結果は、頭部運動有るかた被験者と耳介形状が一致している場合に、もっとも高い音像定位精度となった。
5. 耳介無し条件では、常に低い音像定位精度となり、特に頭部運動無しの条件では、実験結果がほとんど意味を成さなかった。
6. 精密ダミーヘッドと比較した場合、音響的には近い精度とはならなかったが、耳介形状一致で、頭部運動有り条件での定位実験の結果は良好であった。これは頭部運動によって、頭部形状を簡略化可能となる可能性を示唆している。

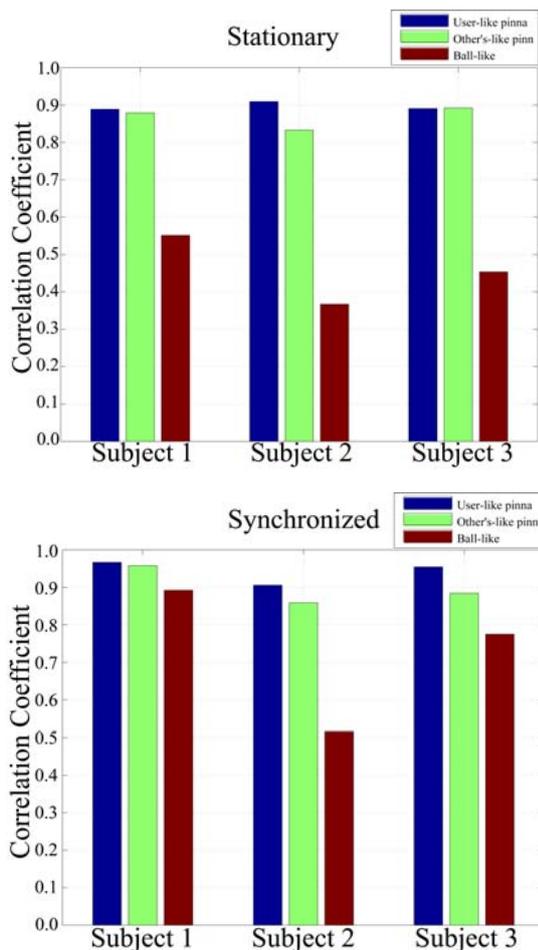


Fig. 9. Correlation coefficient of each result. Upper panel shows the results in the stationary condition and lower panel shows those in the synchronized condition. Blue bars show the results for the ball-like dummy head with user-like pinna. Green bars show the averaged results for two kinds of dummy head with the non-user-like pinna, and brown bars show those for the ball-like dummy head.

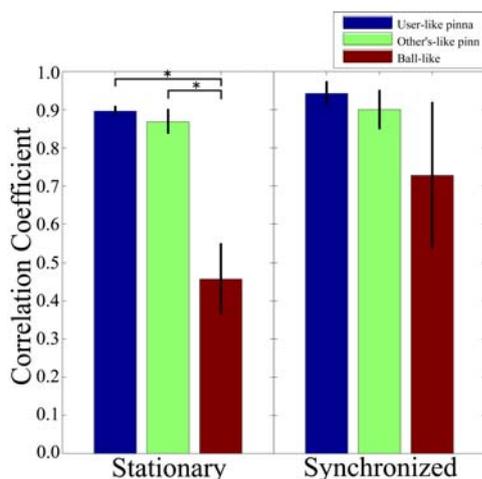


Fig. 10 Averaged correlation coefficient for all subjects. The left graph shows in the stationary condition, and the right graph shows in the synchronized condition. Error bars shows standard deviations of each result.

### 参考文献

[1] R. M. Held, and N. I. Durlach: "Telepresence", Presence: Te-

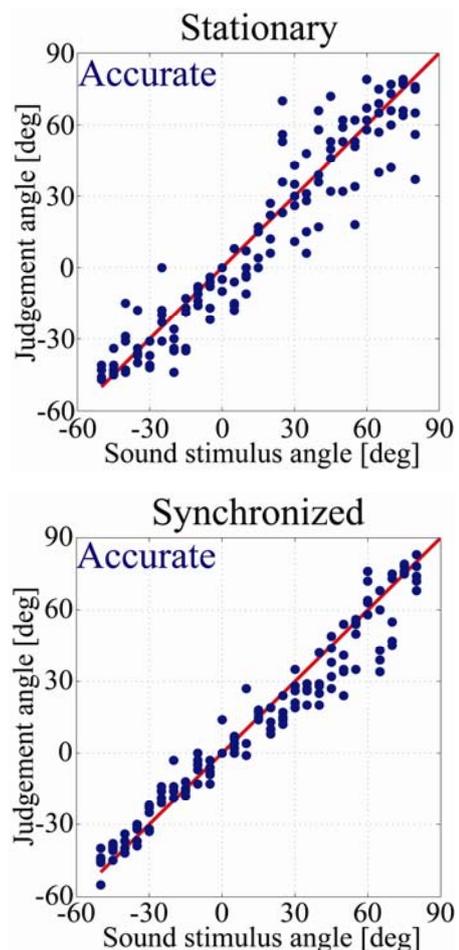


Fig. 11 Results of sound localization experiments using with the accurate user-like dummy head. Correlation coefficient is 0.9427 in stationary condition and is 0.9750 in synchronized condition.

leoperators and Virtual Environments vol. 1, pp. 109 - 112, 1992.

[2] R. Pfeifer, and C. Scheier, "知の創成-身体性認知科学への招待", 共立出版, 2001.

[3] J. Blauert: "Spacial hearing: The psychophysics of human sound localization", MIT Press, Cambridge, Mass., 1997.

[4] H. Wallach: "The role of head movements and vestibular and visual cues in sound localization", J. Experimental Psychology, vol. 27, no. 4, pp. 339-368, 1940.

[5] F. L. Wightman, and D. J. Kistler: "Resolution of front-back ambiguity in spatial hearing by listener and source movement", J. Acoust. Soc. Am., vol. 105, no. 5, pp. 2841-2853, 1999.

[6] W. E. Kock: "Binaural Localization and Masking", J. Acoust. Soc. Am., vol. 22, no. 6, pp. 801-804, 1950.

[7] I. Toshima, H. Uematsu, T. Hirahara: "A steerable dummy head that tracks three-dimensional head movement: TeleHead", Acoustical Science and Technology, vol. 24, no. 5, pp. 327-329, 2003.

[8] V. R. Algazi, R. O. Duda, and D. M. Thompson: "Motion-Tracked Binaural Sound", J. Aud. Eng. Soc., vol. 52, no. 11, pp. 1142-1156, 2004.

[9] I. Toshima, S. Aoki, T. Hirahara: "An acoustical tele-presence robot: TeleHead II", Proc. of International conference on intelligent robots and systems 2004(IROS2004), pp. 2105-2110, 2004.

[10] H. Møller: "Fundamentals of binaural technology", Applied Acoustics, vol. 36, pp. 171-218.

[11] K. A. J. Riederer: "Repeatability analysis of head-related transfer function measurements", 105th Audio Eng. Soc. Conv., no. 4846, 1998.

- [12] D. N. Zotokin, R. Duraiswami, E. Grassi, and N. A. Gumerov: "Fast head-related transfer function measurement via reciprocity", J. Acoust. Soc. Am., vol. 120, pp. 2202-2215, 2006.
- [13] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman: "Localization using nonindividualized head-related transfer functions", J. Acoust. Soc. Am. Vol. 94, pp. 111-123, 1993.
- [14] J. C. Middlebrooks: "Individual difference in external-ear transfer functions reduced by scaling in frequency", J. Acoust. Soc. Am., vol. 106, no. 3, pp. 1480-1492, 1999.
- [15] 戸嶋巖樹, 青木茂明, "音響テレプレゼンスロボットの頭部運動制御", ロボティクス・メカトロニクス講演会, 1A1-N-057, 2005.
- [16] 戸嶋巖樹, 青木茂明, "音響テレプレゼンスロボットの頭部運動再現における聴覚的時間的余裕の定量的評価", 日本ロボット学会誌, Vol.25, No. 6, pp. 990-996, 2007.
- [17] 平原達也, 植松尚, 戸嶋巖樹, "頭部の3次元運動に追従するダミーヘッドシステム - テレヘッド -", AI チャレンジ研究会 2002
- [18] I. Toshima, S. Aoki, and T. Hirahara, "Sound Localization Using an Acoustical Telepresence Robot: *TeleHead II*", Presence, Vol. 17, No. 4, pp. 392-404, 2008.
- [19] Y. Suzuki, F. Asano, H. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses", J. Acoust. Soc. Am. , vol. 97, pp.1119-1123, 1995.