

聴覚実験における音響テレプレゼンスロボットの有用性 Acoustical telepresence robot for auditory psychophysics

○戸嶋 巖樹*
近藤 洋史*
Daniel Pressnitzer^{†‡}
柏野 牧夫*[§]

* NTT コミュニケーション科学基礎研究所,
[†]UMR 8158, CNRS and Université Paris Descartes
[‡]Ecole normale supérieure
[§]東京工業大学大学院 総合理工学研究科

Iwaki TOSHIMA*, Hirohito M. KONDO*, Daniel Pressnitzer (Univ Paris)^{†‡}, and Makio KASHINO*[§]

* NTT Communication Science Laboratories,
[†]UMR 8158, CNRS and Université Paris Descartes
[‡]Ecole normale supérieure

[§]Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

{toshima.iwaki, kondo.hirohito, kashino.makio}@lab.ntt.co.jp, Daniel.Pressnitzer@ens.fr

Abstract—“Embodiment” is one of the most important key words of studies of robots’ learning, human-robot interaction, and understanding of environment for robots. In the area of robot-audition and human auditory perception, this is also important. We use an acoustical telepresence robot: *TeleHead* that has human-like dummy head and is synchronized with user’s head movement in realtime. We can control “embodiment” of the *TeleHead* such as head shape and head movement. Then, we tried several psychophysical experiments; sound localization, delay discrimination, and streaming segregation. Then, we conclude that *TeleHead* can be used for psycho-physical experiments.

1. はじめに

急速に進歩するロボット技術を用いて、人間の性質を解明しようとする研究が多くなされている。そもそも人間の発達や知能を議論するにおいて、外界という未知なるものを如何にロボット（即ちシミュレートされた人間）に理解させるかということは大問題である。そこで、身体を通じた外界の理解に代表される、いわゆる「身体性」が提唱されて久しい[1]。これは、1990年代のBrooksらが強く主張し、ロボット研究で議論されてきた、知能は身体を必要とする、という議論の発展であると捉えられる[2]。このように、古くから知能そのもの、あるいはその獲得プロセスをシミュレーションするためには、物理的存在のある身体、すなわちロボットの身体が必要であると考えられてきた。近年、そのロボットの身体性能が人間に近付いてきたことにより、ロボット技術を用いた人間の性質解明への挑戦が活発になっ

てきたと言える。これは、言語が、その入れ物である脳と共に進化してきた [3]こととのアナロジーを持って受け止める事ができると考える。環境を理解し、知覚し、環境に働きかけるためには、自らの持つ物理特性をはじめとする諸性質を踏まえ、その理解と行動の可能な範囲において、環境を理解し、環

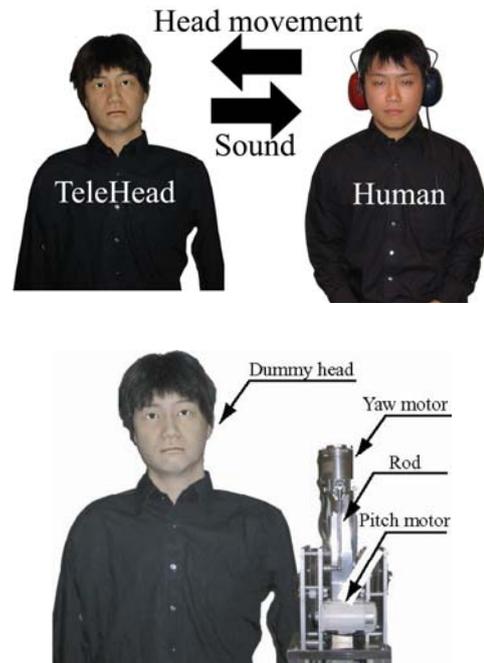


Fig. 1 Outline of *TeleHead* system and inside of the *TeleHead*.

境への働きかけを行うのが効率的である。

ところで、本論文で研究対象とする聴覚においても、身体性の役割は重要である。人間は個々に異なる形状の頭部を持つ。従って、同一の環境、同一の音源からの音であったとしても、鼓膜に伝わる音響信号は異なる。また、頭部運動は音響信号の測定点、即ち耳介の位置を変化させるため、音環境理解において、重要な要素となる。この場合、聞き手としての人間は、自らの頭部形状の影響、頭部運動の影響を事前に学習し、音環境を理解していると考えられる[4]。

前述のロボットを用いた人間のシミュレーションという意味においては、人間と頭部形状の寸分違わない頭部を持つロボットを製作し、人間の動きに追従動作し、さらにその影響を確かめるために、それらの性能を自由に变化させることが可能なロボットがあれば、それらの影響を自由に調べることが出来る。そこで、現状で可能な限り、上記の性能を満たすような、音響テレプレゼンスロボット、テレヘッドを作成した[5]。テレヘッドを用いることで、頭部形状や頭部運動の効果を任意に操作し、人間の性質解明に役立てたいと考えている。本稿では、テレヘッドによって知覚実験を行うにあたり、必要と考えられる基礎的性質、音響定位感や遅延の影響を評価し、それらを踏まえた上で、様々な要素が関係するストリーミング課題での活用の是非を検討した。

2. 音響テレプレゼンスロボット「テレヘッド」

我々は使用者と同形状の頭部を持ち、使用者の頭部運動に追従動作するロボット、テレヘッドを作成した[5]。概要を図1に示す。使用者は、テレヘッドの外耳道入り口付近に設置されたマイクロホンにより集音した音を、ヘッドホンを通じて聴く。頭部形状が使用者の頭部と一致していることにより、使用者の頭部が音環境に与える影響を計算することなく、頭部形状の影響を加味することが出来る。また、ヘッドホンに付けられた頭部運動センサが使用者の頭部運動を測定し、この情報に基づいて、ロボットの頭部が制御される。これにより、使用者の頭部運動の効果が遠隔の音環境で実現され、使用者に提示される。以上により、使用者は、遠隔において、自らの頭部形状と頭部運動の影響が加味された音を聴くことが出来る。

この時、実機を使うデメリットとして、騒音と遅延が考えられるが、騒音は24dB SPL程度であり、静かな部屋でも、何らかの刺激音を聞いている場合はほとんど知覚されない[6]。また、遅延は80ms程度（delay time, 動き出しのタイミング dead time としては10ms程度）は、後述の検討により、臨場感のある音を聴くという使用目的に対して、知覚上は問題ないことを確認している。

3. 音響知覚への影響

3.1 実験概要

音が聞こえる方向を正しく知覚すること、即ち音響定位感は聴覚の実験をする上で重要である。テレ

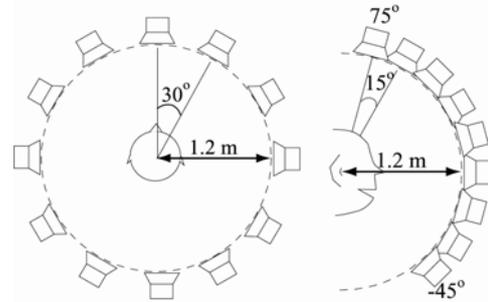


Fig. 2 Setup of sound localization experiments.

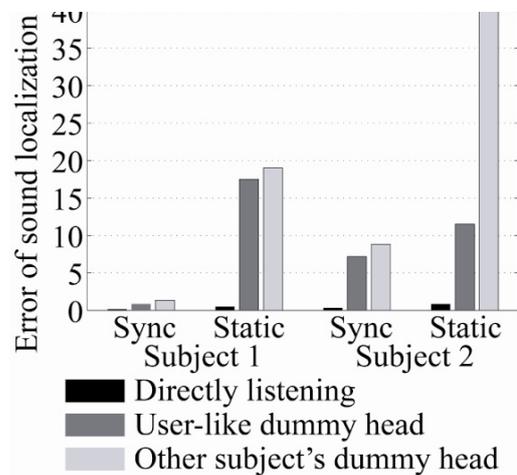


Fig. 3 Sound-localization error in azimuth plane.

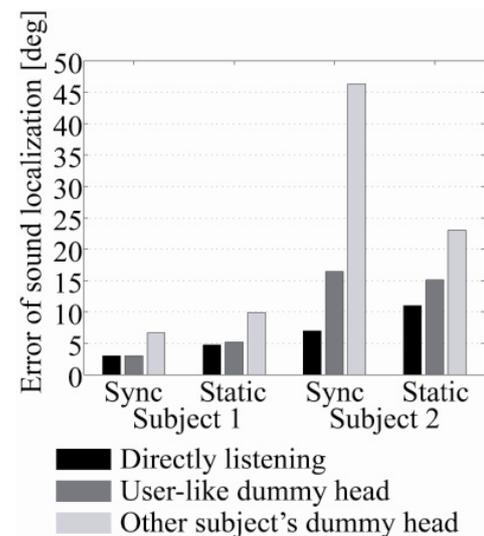


Fig. 4 Sound-localization error in vertical plane.

ヘッドを用いることで、どの程度音像定位感が得られるか実験的に明らかにした[6].

テレヘッドもしくは使用者自身の周りに図 2 に示すようにスピーカを配置した。左は水平面条件、右は正中面条件でのスピーカ配置を示している。テレヘッドを使用する場合、使用者は別室で受聴し、音の聞こえる方向を回答した。テレヘッドを使用しない場合は、スピーカ配置の中心に使用者が座り、直接受聴することとなる。この場合の回答は HMD を用いて提示された画面に対して、手に持ったマウスを用いて回答した。いずれの場合もスピーカは見えない。また、テレヘッドの特長である、頭部形状を一致させたことの効果、頭部運動を遠隔で実現したことの効果をそれぞれ検証する意味で、頭部形状を一致させた場合とさせない場合、また、頭部運動を行った場合と行わない場合についても実験した。

3.2 実験結果

結果を図 3, 4 に示す。図 3 は水平面、図 4 は正中面の定位誤差を表している。いずれも左のバーが直接受聴条件であり、中央が頭部形状一致条件、右が頭部形状不一致条件である。また、各被験者について、左のブロックが頭部運動有り条件、右のブロックが頭部運動無し条件である。頭部形状を制作することの困難さが問題となり、被験者数が 2 名と少ないが、この範囲においては、いずれの条件でも頭部形状が一致し、頭部運動が再現されている場合において、高い音像定位結果が得られている。これは頭部形状の再現、頭部運動の再現共に、定位感の向上に寄与していることを示している。また、テレヘッド使用で、頭部形状一致かつ頭部運動有りの条件で、少なくとも被験者 1 は直接受聴と同等の音像定位精度となっている。これは、テレヘッドシステムが、十分に高精度で機能している事を示しており、音像定位に関して、知覚実験が可能な装置であることを示唆する結果である。被験者 2 については、頭部形状一致、頭部運動有り条件でも、直接受聴と同等までは定位精度が向上しなかったが、元々被験者自身の定位精度が高くない事と、頭部形状の再現精度が被験者 1 のロボットに比べて若干劣ったことが分かっている。個人性の問題もあるため、安易な結論は述べられないが、少なくとも直接受聴並の音像定位精度が得られる結果があることは、テレヘッドの様なシステムが音像定位精度に使用できる可能性を示唆する結果と言える。

4. 遅延知覚

ロボットを用いる際に考慮すべき重大な問題に遅延がある。実機を用いる場合は、計測の時間（多くの場合無駄時間：dead time になる）に加え、機械のイナーシャによる動作遅延が不可避である。テレヘッドでは、系全体としての遅延が、60deg 程度のステップ状の頭部運動に対して 80ms 程度ある[6]. さて、

この 80ms が許容範囲かどうかという事について、以下の通りに実験した[7].

4.1 実験方法

テレヘッドを利用し、正面から音を出して、意識的に頭部運動をする。この際、トータル遅延時間の異なる 2 回の試行を行い、被験者はどちらの試行の方が、遅延時間が長かったかを回答した。被験者は 4 名、少なくとも耳介形状は一致するように頭部を制作した。テレヘッド自体の遅延が 80ms 程度あることから、遅延時間の短い条件は 80,100,120ms の 3 通りとし、これらと、これらよりもさらに長い遅延について、弁別できるかどうかを調査した。

4.2 実験結果

結果を図 5 に示す。Weber 則によれば、比較の基準となる短い遅延刺激の遅延時間が長くなれば、追加遅延時間の閾値が高くなることが予想された。しかし、結果は被験者によって閾値自体はことなるものの、そのような傾向は見られず、むしろ 40ms, 20ms などで、被験者毎に一定の閾値となった。これは、タスクへの回答が遅延時間自信ではなく、何か他の手がかりを利用しており、その結果、一定になったことを示唆しているのではないかと考えた。その手

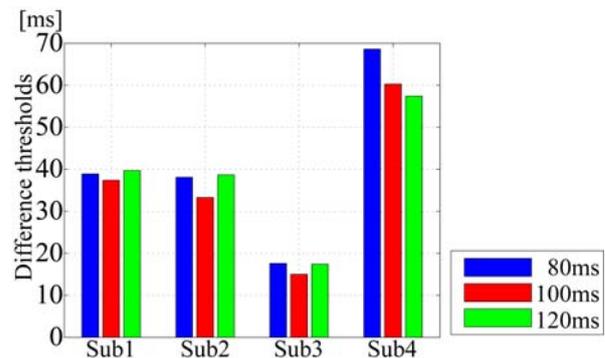


Fig. 5 Delay thresholds

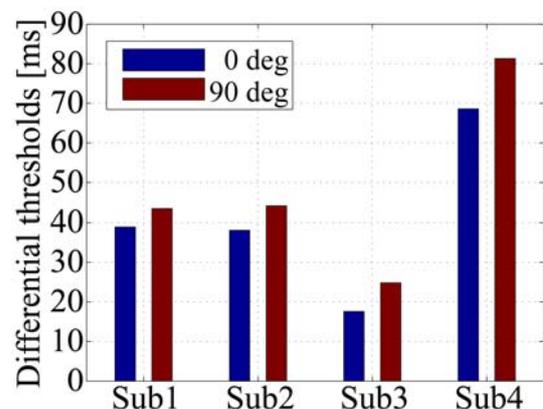


Fig. 6 Delay thresholds depend on sound direction

がかりを空間的な手がかり、つまり音像定位感の差と仮定して、いくつかの追加実験を行った。仮説は、頭部運動の遅れを定位感のずれとして知覚しているということである。仮説が正しければ、頭部形状を変化させ、定位感を低下させた場合、閾値は高くなる。また、音源の方向を定位精度の高い正面に配置した場合と、定位精度の低い横方向に配置した場合では、正面の方が、閾値が低くなる。さらに、広帯域雑音に比べて、音声では定位精度が低下すると考えられるため、閾値が高くなるはずである。

実際に正面と横方向に音源を配置した場合の結果について、図6に示す。いずれの被験者においても、正面方向に音源を設置した場合に閾値が下がり、遅延に対して敏感になった。他の2つの条件、頭部形状の変化と刺激音の変化についても、定位感によって遅延を弁別していることを支持する結果を得ており、空間知覚が遅延の知覚に重要であることを示唆する結果を得た。なお、その際の閾値が頭部運動の最高速度と遅延時間の積、即ち遅延に影響を受けた音像のずれ角度は約10度~17度となるあたりに遅延の閾値があることが実験的に分かっている。

ところで、人間の首の動きは瞬間的でも高々300deg/sec程度であり、テレヘッドの無駄時間10msの間に移動する距離は3deg程度である。また、遅延時間80msの間、300deg/secの運動を継続できたと仮定しても、80ms後に発生する音像の空間的誤差は12deg ($300\text{deg/sec} \times 80\text{ms} \times 1/2$)であり、これは広帯域雑音を正面から提示したとしても、知覚限界付近である。音声を用いた場合はさらに知覚されにくくなることも確認済みであり、テレヘッドは単体で通信を伴わずに使用する場合には、許容される遅延レベルであることが分かる。

さらに実用的な側面では、頭部形状の一般化が考えられるが、その場合、この運動遅延の影響はさらに小さくなることも分かっている[8]。また、同様のシステムを2台遠隔でつないでテレプレゼンスの質を評価するような検証も行われている。それによれば、TCP/IPで東北大学と富山県立大学を結んだ場合、テレヘッドシステムの頭部運動は数630ms、UDPで約140ms遅れ、遅延時間の揺らぎによって、両耳間時間差が定位方向で0.5degに相当する程度、誤差を持つことが分かっている。このようなより実用に即した場面での遅延の影響の評価等は今後の課題として残っていると云える[9]。

5. 複数の要因が関係する知覚実験へのテレヘッドの使用:ストリーミング課題

5.1 実験の背景・方法

ここまで頭部形状や頭部運動にまつわる、音像定位や、遅延の感覚について議論してきたが、様々な要因が関係する知覚現象も多くある。本稿では、その一例として、音脈分凝について、議論する。ストリーミング課題と呼ばれる知覚課題である。図7に

示すような音の時系列刺激が提示された場合、聞き手には多くの解釈が存在する。例えば、ABAやAAB等でセットとなるリズム的な音列が提示されているという解釈(1ストリーム)。あるいは、音の繰り返しの速さと高さの異なるAAAとBBBという2つの音列が同時に提示されているという解釈(2ストリーム)である。一般的には、聞こえはじめは1ストリームであるが、連続して聞いているとやがて2ストリームになると言われている。しかし、この現象は、音の空間配置や頭部運動、あるいは注意(聞き手が1ストリームと2ストリームのどちらとして知覚したいと考えているか)などが複雑に関係するため、その機序の解明は容易ではない。

ここで、テレヘッドを用いれば、空間的知覚、頭部運動、頭部運動の意識の影響を切り分けることが出来る。頭部運動にあわせてテレヘッドを動かす通常の条件に加え、頭部運動にかかわらずテレヘッドを動かさないことで、頭部運動の音響的效果を打ち消す条件、さらに、頭部運動していないのに、以前の使用者の頭部運動をテレヘッドが再生することにより、頭部運動の音響的效果のみを与え、自己運動

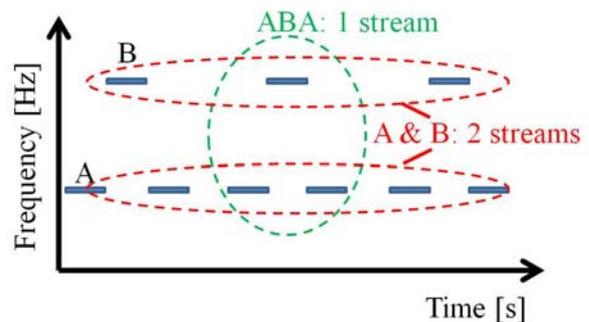


Fig. 7 Streaming segregation.

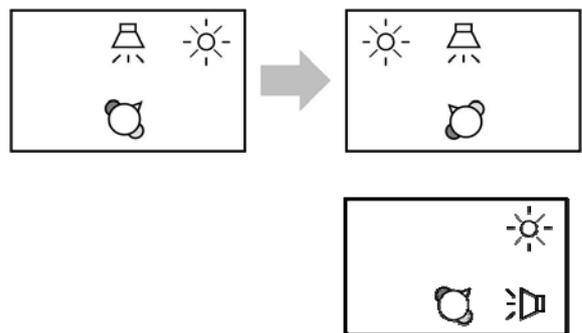


Fig. 8 Head movement of the experiments. Upper panel shows the normal situation. Under panel shows the condition that subject does not move the head and TeleHead moves his head. This makes sound direction change.

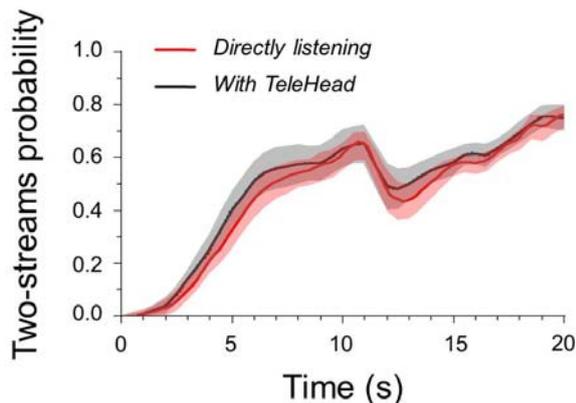


Fig. 9 Results of streaming task.

の効果を与えない条件、などがすぐに実施可能である。具体的な実験の例を図 8 に示す。始め被験者は右の LED を注視し、正面のスピーカからストリーミングの音響刺激を聴いている。あるタイミングで左の LED が点灯し、被験者は素早く頭部運動を行う。この時に、頭部運動の影響を受けて、ストリーミングがどのように変わったかを回答す。一方で、図 8 下に示す条件では、同じタイミングで被験者へは頭部運動の指示を出さず、TeleHead のみが直前の被験者の頭部運動と同じ運動を行う。この場合、被験者もしくはテレヘッドの運動後の、被験者と音源の相対的方向は一致する。従って、被験者への音響信号は一致しているが、自らの意識を伴って頭部運動したかどうか異なる。この 2 つの実験の結果を比較することで、自発的頭部運動の効果について論じることが出来る。

既存のバーチャルリアリティシステムでも、たまたみ込み演算による加減運動で音の運動や頭部運動を再現できる。しかしこれは、パラパラ漫画が動画に見えるのと同じ事であるが、このような複雑な知覚現象においては、連続運動と知覚されることが、本当に連続的に変化する音響信号の情報と等価かどうか疑問が残る。

もちろん、一方で、ロボットを使用する弊害も多く存在する。遅延や、頭部形状の再現精度の問題も大きな物である。そこで、常に、人間が直接受聴する場合とどの程度異なるのか、議論する必要がある。そこで、使用者にストリーミング課題を課し、時間と共に聞こえがどう変化するか、ロボットを介す場合と介さない場合を比較した。

結果を図 9 に示す。図 8 の上段の実験条件におけるストリーミング課題の回答結果を示している。横軸が時間であり、10 秒後に頭部運動を行うようにした。縦軸が 2 ストリームと知覚する確率である。はじめは 1 ストリームであるが、数秒後には 2 ストリームと聞こえる確率が上昇し、10 秒目に発生する頭部運動イベントによって、2 ストリームと聞こえる確率が少し低下し、運動後もう一度上昇していること

が確認できる。この実験をテレヘッドを介さず、被験者が直接音を聴いた場合が重ねて描かれているが、両者にはほぼ差がないことが読み取れる。これは、遅延や騒音など、様々な影響はあるものの、複雑な知覚現象を扱う実験においても、テレヘッドが有効に機能していることを示唆する結果と言える。

6. おわりに

頭部形状と頭部運動を再現する音響テレプレゼンスロボット：テレヘッドを用いて、音像定位やストリーミング課題によって、ロボットを用いる実験の可能性について検討した。

- ・音像定位課題においては、頭部形状を作り込むことにより、定位精度は向上した。特に頭部運動を再現することで、定位精度は直接受聴とほぼ変わらない程度まで向上することが確認された。

- ・遅延弁別課題においては、遅延の追加に対する閾値が定位感のずれに依存していることを示唆する結果を得、80ms の遅延ではほとんど、定位感のずれに起因する遅延を知覚することは無いことを示唆する結果を得た。

- ・ストリーミング課題は、頭部形状、頭部運動、注意、空間知覚など、多彩な聴覚の知覚が関連する複雑な課題であるが、この様な課題においても、テレヘッドを用いた結果は、直接受聴と変わらなかった。

以上から、テレヘッドは音像定位の様な単純な聴覚実験や、ストリーミング課題の様な複雑なものまで、広範に活用可能であることを示唆する結果を得た。

ところで、既に本稿におけるストリーミング課題でも、被験者自らが頭部運動する意識的運動と、ロボットが勝手に動く意識的では無いが物理的には等しい効果をもたらす運動の比較などを行えることを確認している[10]。また、頭部運動を手元のコントローラで操作することにより、頭部運動の意識は保ったまま、頸部の体性感覚の効果のみを検証し、体性感覚が音像定位に寄与することを示唆する結果も報告されている[11]。

このように、人間の外界との関わり、特に環境理解について考えるとき、「身体性」の重要さは、その物理的存在に留まらないと筆者は考える。今回新たに検討した複雑な聴覚現象の一つであるストリーミング課題においても、意識・情動といった問題が大きく関わってくる。従来、ロボティクス、特に創発ロボティクスの分野では、Pfeifer らの議論に基づき、身体性について、外界を理解するために用いる物理的フィルタの様に捉えてきた。しかし、知覚現象が物理世界の脳内表現の解釈の一つであるならば、そのフィルタには、意識・情動といったものも含まれるのではないか。そうだとすれば、意識的に行った運動の効果をバーチャルリアリティシステムによって、精確に再現し、意識や体性感覚の影響をそれ

ぞれ個別に取り除く手法は、人間理解の発展に多大な寄与をもたらすのではないかと考えている。

参考文献

- 1) R. Pfeifer and C. Scheler, “知の創成”, 共立出版, 2001.
- 2) R. A. Brooks, “Elephants don’t play chess”, *Robotics and Autonomous Systems*, vol.6 pp3-15, 1990.
- 3) T. Deacon, “Symbolic species: The co-evolution of language and the brain”, W. W. Norton and Company Inc, 1997.
- 4) J. Blauert, “Spatial hearing”, MIT Press, 1982.
- 5) I. Toshima, H. Uematsu, and 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.
- 6) I. Toshima, S. Aoki, and T. Hirahara, “Sound localization using an auditory telepresence robot: TeleHead II”, *Presence*, MIT Press, vol. 17, no. 4, pp.392-404, 2008.
- 7) 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”, 第24回ロボット学会学術講演会,
- 8) I. Toshima, S. Aoki, “Possibility of head-shape simplification for an acoustical telepresence robot: TeleHead”, *Journal of Robotics and Mechatronics*, Japan Society of Mechanical Engineers, vol. 21, no. 2, pp. 223-228, 2009.
- 9) 平原達也, 森川大輔, 岩谷幸雄, “インターネット接続したテレヘッドによる聴覚テレプレゼンス”, 音響学会秋期研究発表会, pp. 651-652, 2011
- 10) 近藤洋史, Daniel Presnitzer, 戸嶋巖樹, 柏野牧夫, “音脈分凝のリセットに対する頭部運動の影響”, *聴覚研究会資料*, vol. 41, pp. 365-370, 2011.
- 11) 吉崎大輔, 平原達也, “頭頸部の体性感覚情報が水平面音源定位に及ぼす影響”, 日本音響学会 2011 年秋期研究発表会 講演論文集, pp.479-480