

頭部運動が 3 次元空間の音像定位した音声了解度に及ぼす影響

○齋藤直，小坂哲夫，近藤和弘（山形大学大学院 理工学研究科）

1 はじめに

ヘッドホンなどを用いた音声の受聴は、通常モノラルで提示され、仮想空間上の 1 点（多くは頭部内部）に音像が定位されることに相当する。これは競合音が想定される場合でも 1 点に定位されるため、場合によっては非常に聞き取りにくい。そういった場合の音声了解度を向上させる技術として音像分離がある。複数の音声を 3 次元の仮想空間上の別々の位置に定位させることにより、実空間と同じような感覚で受聴ができ、競合音がある場合でも音声了解度の向上が見込まれる。しかし、ヘッドホンにより提示されている仮想空間上の音像は、人間が自然な状態で音声を聴く場合に行う無意識な頭部運動をしたとしても、ヘッドホンと共に動くため、音像と耳の位置関係は変わらない。頭部運動は音像位置の知覚に有効であり、了解度の向上の一助になることも容易に類推できるが、このような状況では頭部運動が反映されず、不自然な受聴状態であるといえる。これを解決するにはヘッドトラッキングを用いる。頭部運動を検知し、その動きを打ち消す向きに仮想空間上の音像を移動させ、実空間の定位置に音像が位置し続けるようにでき、より実空間に近い受聴環境が再現でき、了解度の向上も期待できる。

3 次元音響の先行研究の中でもヘッドトラッキングに注目したものもあるが、1 種類の音声を用いた実験が多く、複数人同時会話を想定するような、競合音を設定した場合の了解度の検討が不十分であった。また、3 次元音響にヘッドトラッキングを用いた場合の検討も十分になされていない。

本研究では競合音を設定し、3 次元音響を用いた仮想空間、および実空間での頭部運動が音声了解度向上に有効であるか検討する。

2 実験環境

2.1 音声了解度試験

音声了解度の評価には二者択一式音声了解度試験（DRT : Diagnostic Rhyme Test）を行う。DRT とは語頭 1 音素のみ異なる単語対からなる専用リストを用いて行う評価である。これらの単語の語頭子音は 6 つの属性から成っている（表 1）。この被験者は単語対の内の 1 単語を聴取し、単語対のどちらの音声も聴こえたか二者択一式で選択する。DRT は訓練されていない被験者でも評価が可能である。また、被験者に示す候補には必ず正答単語が含まれる。

この試験の正答率は二者択一であるため、まったくでたために回答しても 50% の正答率が期待される。そうした

偶発性を避けるため、以下の式により正答率を算出する。

$$\text{正答率} = \frac{\text{正答数} - \text{誤答数}}{\text{総提示回数}} \times 100 \quad (1)$$

表 1: 日本語 DRT 用単語対の一部

語頭子音属性	単語対の例
有声性 (Voicing)	財- 才 (zai - sai)
鼻音性 (Nasality)	万- 番 (man - ban)
継続性 (Sustention)	箸- 菓子 (hashi - kashi)
不規則性 (Sibilant)	ジャム- ガム (neru - deru)
抑音性 (Graveness)	梓- 楽 (waku - raku)
集中性 (Compactness)	焼く- 沸く (yaku - waku)

2.2 実験環境

本研究では図 1 で示す実験系を用いて、目標音と競合音をランダムな位置から同時に提示する実空間および仮想空間で評価を行う。実空間では該当する位置にスピーカーを設置、仮想空間では該当する位置に音像を定位させて評価する。方位は同図のように被験者正面を 0 deg とし、左右 ± 45 deg, ± 90 deg, 後方 180 deg とする。

目標音として女性話者が発声する DRT の単語、競合音としてバブルノイズを用いている。目標音が再生されている間、競合音も同時に再生される。

受聴位置での目標音の音圧を 60 dB とし、さらに SNR (Signal-to-Noise Ratio) が 0dB, -6dB, -9.5dB の 3 種類になるよう競合音の音圧を調整し、それぞれ検証する。

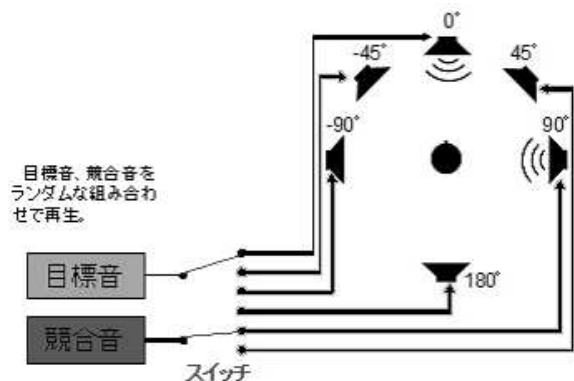


図 1: 実験系概要図

3 実空間での実験

3.1 頭部固定装置

実空間での頭部固定時は、モジュールに電池ボックスを繋いで作成した小型のレーザーポインタを装着する。こ

のレーザーポインタは、金属製で細身のカチューシャに取り付けている。被験者には、頭部固定条件での評価中はレーザーポインタが正面に設置した的の中心を指すように常に気を付けてもらうよう事前に説明した。

装着する装置を出来るだけ小型化することで、音声聴き取り時に、被験者の受聴時に違和感を与えない、つまり HRTF (Head Related Transefer Function: 頭部伝達関数, 4.1 参照) への影響を最小にしている。また、的を注視することで、目線の移動によって得られる視覚的情報も制限している。

3.2 実空間での実験結果と考察

図 2 に実空間での頭部固定時 (Restricted) と頭部非固定時 (Unrestricted) の DRT 結果を示す。全体的に音声了解度は、SNR の低下とともに正答率の低下がみられた。

評価データを詳しく見てみると、0 dB と -9.5 dB においては頭部固定時の方が音声了解度は若干高いという結果になった。傾向も似ている部分があり、目標音が右側から提示された時の方が左側から提示された時よりも正答率が高く、また、後方 180 deg に目標音があるときの正答率が他の角度に比べて低い。これは、真後ろからの音声の聞こえ方に慣れていないためと思われる。

各 SNR の結果に対し分散分析を行ったところ、0 dB では有意差が見られたものの、-6 dB、-9.5 dB では有意差は得られなかった。低 SNR において有意差が無くなるようである。

図 3 に試験音に含まれる目標音に対する競合音の相対角度別に正答率をまとめたものを示す (例: 目標音 -45 deg, 競合音 90 deg の場合、相対角度は 135 deg)。目につくのは、SNR: 0 dB で頭部固定時がどの方向に対しても同程度に高い正答率を示していることと、逆に SNR: -9.5 dB で方向によって正答率が大きく落ち込んでいる部分があることである。SNR: -9.5 dB では目標音に対して競合音が右側に位置するときに正答率が下がっていることがわかる。これはわずかながら SNR: -6 dB においても見られた傾向であった。これは被験者が受聴してから認識するまでの過程の中で、何らかの影響で正答率が低下するものと思われる。

以上より、実空間では SNR: 0 dB で頭部運動をしない方が音声了解度が高いといえる。

3.3 頭部運動の統制

先の実験における図 2 を見ると、SNR: -6 dB での傾向が他の SNR と大きく異なっていた。全ての実験において同じ環境、同じ条件で実験を行ったものの、頭部非固定の頭部運動は被験者の裁量に任せたであった。

この点のために結果に差が出てきてしまったのか、ここで検証した。

試験音の構成を変更する。競合音を全長 2 秒にし、目標音の再生開始を競合音の再生開始 1 秒後からにする。目標音を判別する語頭音素が試験音の中ほどのタイミングに来る構成である。使用した目標音は、先の実験で全体の平均値に最も近い傾向を示した音素の Nasality のみを使用する。

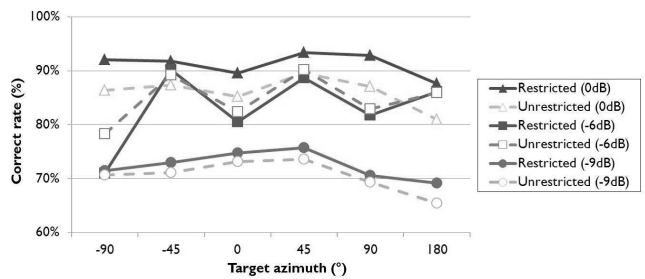


図 2: 実空間の目標音提示角度別正答率

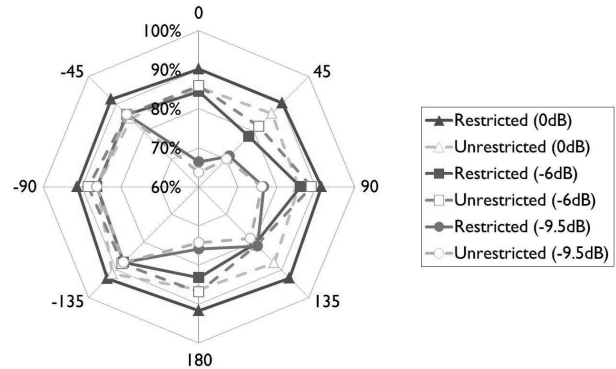


図 3: 実空間の目標音-競合音相対角度別正答率

被験者は、試験音再生中の 2 秒間で頭部を -45 deg から 45 deg に滑らかに水平移動させながら受聴する。この運動を統制しての頭部非固定時の測定と合わせて、固定時も再測定した。

頭部運動統制の実験結果 (-6dB:co) を、先の実験の -6 dB の Nasality のみでプロットしたもの (-6dB) と一緒に図 5 に示す。

全体的な正答率が上昇し、90 deg, 180 deg においては 5 ポイント以上向上した。しかしやはり頭部固定、非固定での有意差は得られない結果となった。また、グラフの概形も多くの提示方向で先の実験同様の結果になったため、先の実験において行っていた頭部非固定時のデータは、頭部運動の影響を加味出来たものであったといえる。

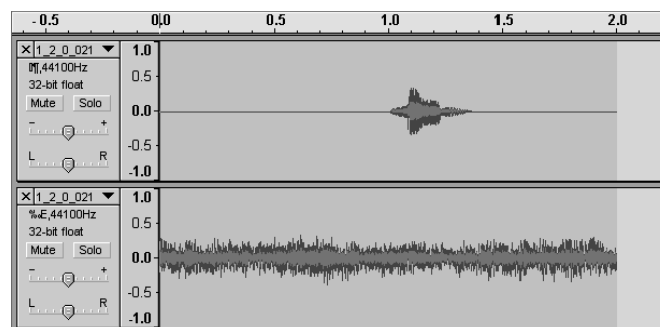


図 4: 頭部統制用試験音の例 (上が目標音, 下が競合音)

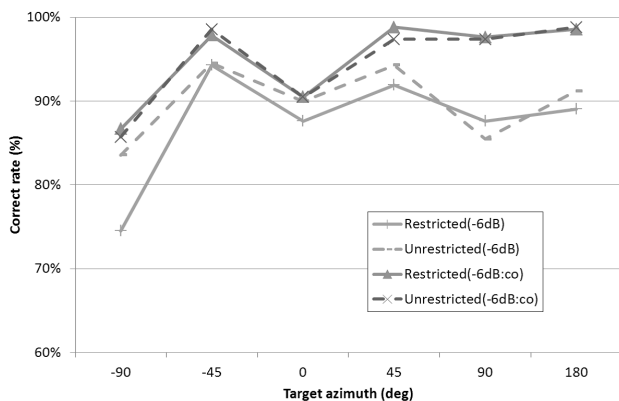


図 5: 頭部運動統制の有無の比較

4 仮想空間での音像移動

4.1 音像定位

まず初めに、音像移動が仮想空間内での音声了解度には有効であるか、実験を行う。仮想空間内でも、図 1 に示したスピーカーの角度に相当するように音像を配する。

立体的に音像を提示するために、HRTF を用いる。これは実空間において、音源から外耳道入口までで音声を受ける周波数特性のことで、音源の角度、耳周辺の形により変化する。一般的なヘッドホンは音源が外耳道に直結しているために頭部伝達関数が無視されており、立体的には聞こえない。提示音に任意方向の HRTF を畳み込むことによって、3 次元音響が提示できる。

ヘッドホンによる仮想空間上では、HRTF を畳み込んだだけでは音像と耳の位置関係に実際の頭部運動は影響しない。仮想空間において音像と耳の位置関係を動かすには、音像の方を動かす必要がある。HRTF を想定した頭部運動のタイミングに合わせて次々切り替えていけば、音像が移動して聞こえる。今回想定した頭部運動は 3.3 で行った統制の動きとする。

本研究の仮想空間内の設定として、前者の HRTF を畳み込んだだけの提示を「音像固定」とし、後者の想定した頭部運動に従って音像を移動させる提示を「音像移動」とする。

4.2 音像移動の実験結果と考察

図 6, 7 に仮想空間内での音像移動時 (Unrestricted) と音像固定時 (Restricted) の DRT 結果を示す。

0, -6 dB では総じて正答率が 90% を超えている。-9.5 dB においては音像移動時の正答率が低下している。これは低 SNR 時の目標音判別が困難な環境で、さらに目標音の音像が移動しながら提示されたために語頭音素がわかりにくくなったためと思われるが、それでも平均値としては実空間での実験よりも高い。この -9.5 dB 結果を見ると、実空間では高かった ± 45 deg の正答率が他より低い。3 次元音響とはいえ実際に聴いているのはヘッドホンからの音声になるため、被験者は聴き取りにくい -9.5 dB の低 SNR という状況では、通常のヘッドホン受聴時の反応と同じように両耳に対して同じように聞こえる 0 deg か、あるいは片耳の音声に集中しやすい ± 90 deg での目標音提

示が聴き取りやすく、音声了解度向上につながったと思われる。なお有意差は -9.5 dB でのみ得られた。

目標音と競合音の相対的な位置関係による正答率の変化は、競合音の位置の左右に関して対称な結果となった。目標音と競合音が同一方向にある場合の正答率が最も低い、目標音 0 deg、競合音 0 deg 提示の場合は 100% に近いスコアになっていたため、同一方向からの提示での音声了解度低下は、提示方向しだいであるといえる。

仮想空間では低 SNR 時、音像を固定しての受聴の方が了解度向上には有効である。

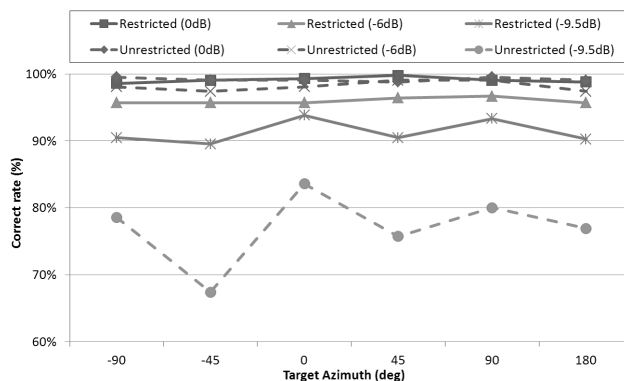


図 6: 仮想空間内の DRT 結果 (音像移動の有無)

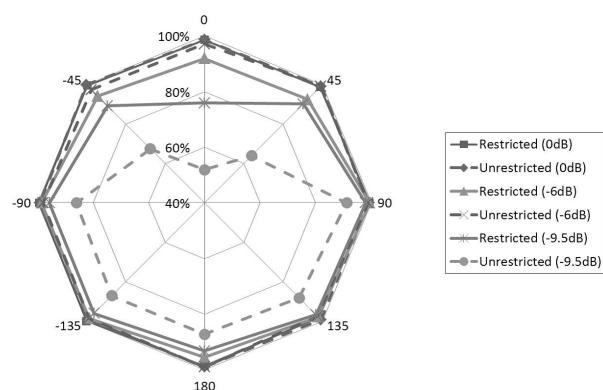


図 7: 仮想空間内の目標音-競合音相対角度別正答率 (音像移動の有無)

5 仮想空間での頭部運動

5.1 擬似的ヘッドトラッキング

前章での実験においては、試験音受聴時に実際に頭部を動かさずに実験を行った。前章では動かさなかった被験者の頭部を、本章においては受聴時に実際に動かし、擬似的にヘッドトラッキング状態を作り出し、DRT を行う。

前章の音像移動時に用いたものと同じ試験音を使用する。本章では頭部運動統制時と同じ頭部運動を実際にしながら試験音を受聴する。すると音像が実空間上の特定位置に固定されて聞こえる。これによって擬似的にはあるがヘッドトラッキングを模擬した状態になる。

5.2 仮想空間での頭部運動の実験結果と考察

実際に頭部を動かし、擬似的にヘッドトラッキングを行った場合の結果 (Head_Tracking) と、同じ試験音を使い、前章で行った頭部を動かさなかった場合の結果 (Unrestricted) を合わせて図 8, 9 に示す。

0 dB, -6 dB ではほとんど差がなく、正答率は全体的に高い。-9.5 dB において Head Tracking は Unrestricted と同様低下し、目標音提示角ごとの傾向も近い形であるが、むしろ Unrestricted よりも正答率は低くなった。なお、全ての SNR においてヘッドトラッキングの有意差は得られなかった。

図 9 を見ると-9.5 dB において、正面右側での正答率がわずかに下がっているのが見て取れる。これは前章の仮想空間内の実験では得られなかった傾向である。しかしこの傾向は実空間で得られた傾向であった。このことより、ヘッドトラッキングは実空間内での受聴に近い状態にすることが推測できる。

以上より、ヘッドトラッキングを行うのは音声了解度向上に対しては有効でないが、実空間での受聴のように聴こえることがわかった。

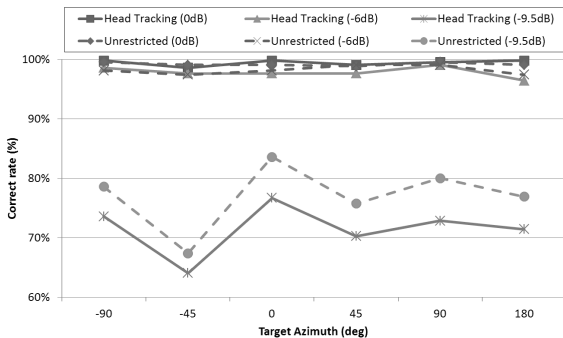


図 8: 擬似的ヘッドトラッキングの DRT 結果

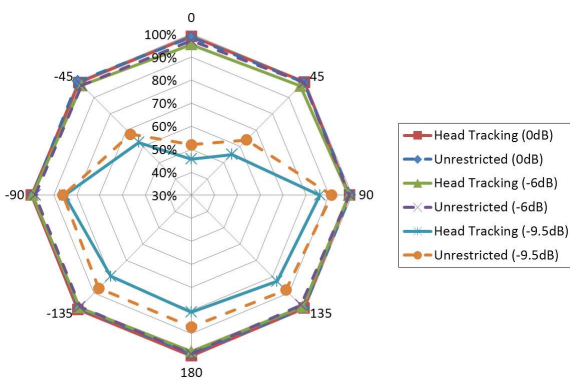


図 9: 擬似的ヘッドトラッキングの目標音-競合音相対角度別平均正答率

6 仮想空間での目標音固定

6.1 目標音の固定

ここまでの実験で、仮想空間内の受聴では音像を動かさず、また頭部も動かさない方が了解度が高いという結果が得られてきた。この章では目標音の音像を仮想空間上正面 0 deg に固定し、競合音は移動させながらという状況での実験を行う。

6.2 目標音固定の実験結果と考察

図 10 に目標音 0 deg 提示での各結果をまとめて示す。Restricted, Unrestricted は図 6 から目標音 0 deg 提示のものを抜き出したものである。Restricted は目標音、競合音ともに固定。Unrestricted はともに移動。Target_Rest は目標音のみ固定、競合音は移動している状態である。今回のプロットは、目標音提示 0 deg で統一しているため、横軸には競合音提示角度をとっている。

ここまでの仮想空間内での実験同様に、SNR: 0 dB, -6 dB では目標音、競合音の移動に関わらず高い正答率を示しているが、-9.5 dB においては各状態によって一様でない影響を与えた。目標音を固定しての Target_Rest は、目標音の語頭音素が発せられるタイミングで競合音と重なる 0 deg での正答率が著しく低下している。この傾向はわずかながら 0 dB, -6 dB の Restricted においても見られたが、それでも 90 % 以上のスコアである。-9.5 dB の競合音提示 0 deg において正答率が低いのは Target_Rest のみで、Restricted, Unrestricted では正答率が高いという傾向になった。

Target_Rest の受聴状態の特徴としては、目標音と競合音の相対角度が刻々と変化するという点があげられる。この特徴が SNR: 0 dB, -6 dB では全体的な了解度の高さによって気にならなかったが、-9.5 dB となった時に現れたといえる。

低 SNR では目標音と競合音の相対的な位置関係を動かさない方が了解度向上に有効である。逆に、低 SNR においても目標音が正面にあり、かつ競合音との相対的な位置関係が動かなければ高い了解度が得られる。

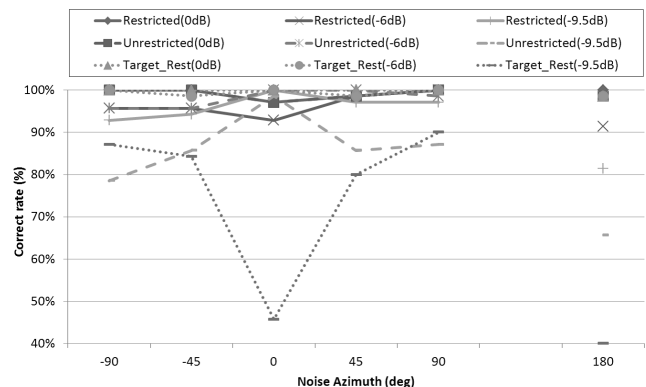


図 10: 目標音 0 deg 提示での DRT 結果

7 まとめ

7.1 結論

競合音のある仮想空間において、頭部運動が音声了解度に及ぼす影響を、実空間の実験も踏まえ、検証してきた。結果として、仮想空間での頭部運動にあたる音像移動は、低 SNR において了解度向上には有効ではなく、むしろ低下させる結果となった。実際の頭部運動と合わせた擬似的ヘッドトラッキングを行っても、有意差は無く、ヘッドトラッキングによる音声了解度向上は得られなかった。

実用的なシステムとして考えた場合、SNR を精々-6 dB 程度まででの運用をするべきであるといえる。しかし-9.5 dB 程度においても、目標音を正面に固定し、競合音との位置関係を動かさなければ十分な音声了解度が得られる。

7.2 今後の課題

本研究は、聴覚情報のみによる音声了解度にだけ焦点をおいて行ってきた。音声了解度に関しては上記のような結果になったが、受聴時の自然性・臨場感といったものに関しては別の研究が必要になる。

音声了解度だけを考えればヘッドトラッキングは必要ないが、やはり3次元音響受聴時、頭部運動を無視して音像が固定されるのは不自然な印象を与える。本研究においても擬似的ヘッドトラッキング時には、実空間で自然に聴き取っている時と同じような傾向が示されているため、臨場感などを考える場合はヘッドトラッキングは避けられない要素であるといえる。

また、本研究では音声了解度と自然性はトレードオフなもののように語ってきたが、これは視覚的な情報と併せることで両立できると思われる。人間の得る情報の大部分は視覚的なものであり、その情報は聴覚情報へも影響を与えうる。発話受聴時、発話者の口・表情・その他の動きを見取り、聴覚情報と併せて認識することでの音声了解度向上が見込まれる。これにより、自然性を高めながら音声了解度も高くすることが可能ではないかと考えている。

参考文献

- [1] 木村, 他:” 音像定位に頭部運動が及ぼす影響に関する考察”, 信学技報 EA2001-44, (2001).
- [2] K. Kondo, T. Chiba, Y. Kitashima, and N. Yano, ”Intelligibility Comparison of Japanese Speech with Competing Noise Spatialized in Real and Virtual Acoustic Space”, Acoust. Sci. & Tech., vol. 31, no. 3, pp. 231-238, 2010.
- [3] 近藤, 他:” 二者択一型日本語音声了解度試験方法の検討” 日本音響学会誌 63 巻 4 号, PP. 196– 204, (2007).
- [4] Rice University Signal Processing Information Base (SPIB)
- [5] Douglas S. Brungart, E.C.: ”The Impact of Head-tracking on Intelligibility in a Multitalker Display” Speech Separation And Comprehension In Complex Acoustic Environments (2003).
- [6] 矢入, 他:” 頭部運動感応型ソフトウェア聴覚ディスプレイの開発” 日本バーチャルリアリティ学会論文誌 437-445 (2006).