異なる向きで発話された音声の知覚
―聴取者に対する発話者の向きの識別―*

☆香嶋陽菜，荒井隆行（上智大），杉本岳大，木下光太郎，中山靖茂（NHK）

1　はじめに

現在，AR（Augmented Reality）や VR（Virtual Reality）等を応用したイマーシブメディアの実現に向け，よりリアルなコンテンツの開発が進められている[1-4]。そのような中，音響情報が担う役割も重要であり，例えば実際の発音源から放射される音波に着目すると，方向ごとに異なる音響特性を再現することが出来れば，発音源に対して多方向からの視聴をよりリアルに行うことができるコンテンツ空間を AR・VR 上で構築できると考えられる[1]。一方，発音源を取り囲む全方位で収録を行い，それを使用して再生すればリアルなコンテンツ空間を構築できるものの，全方位で収録を行うことは設備面からも現実的ではなく，データも膨大となってしまう。そのような理由から，実際には発音源に対して 1 つのマイクロホンのみで特定の方向から収録する場合も少なくない。そこで，仮に全方位による収音が行われなくても，発音源の向きによる音響特性の変化を把握することにより，1 本のみのマイクロホンによる収録から様々な向きの音を再現する手法が提案されている[2]。

そのような背景のもと，杉本ら[4]によって話者の向きが異なる音声に関する主観評価が行われた結果，放射角度の違いが 30° 程度であれば，ヒトがその違いを検知できないことが示された。一方，任意の角度で収録された話者の音声を聞いて，その話者の向きがどの程度分かるかという点についてはまだ十分な検討が進んでいない。そこで，本研究では発話された角度に対するヒトの知覚能力を調べることを目的として実験を行った。

話者の向きが異なるとき，発話の放射特性によって，話者の向きごとにラウドネスは変化している[3]。そのため，ラウドネスを手掛かりにして話者の向きを認識できる可能性が考えられた。そこで，本研究では話者の向きによるラウドネス変化を保った刺激と，話者の向きによらずラウドネスを一定に保ってラウドネスのキューを排除した刺激による，2 つの実験（実験 A・実験 B）を行った。

2　実験

2.1 音声資料

本研究では，NHK 放送技術研究所開発の，複数のマイクロホンを球面上に配置した 3 次元放射特性測定装置[3]を用いて，話者の水平面の右半身 0 度，45 度，90 度，135 度，180 度に置かれたマイクロホンにより同時に収音された音素バランス文の音声を使用した。話者はナレーター経験者で，20，30，40 代の各年代の男女 1 名ずつ，合計 6 名である。

本研究では，放射方向が 0 度の音声の平均ラウドネス値[5]を話者間で正規化し，正規化時のレベルシフト量を話者ごとに 0 度以外の方向の刺激に適用した実験 A と，全刺激の平均ラウドネス値を正規化した実験 B の 2 つを実施した。実験 A では放射方向によるラウドネスの変化が保存されているのに対し，実験 B では放射方向によるラウドネスの変化が排除されている点に留意されたい。

2.2 参加者

実験参加者は日本語母語話者かつ聴覚に異常がない（事前アンケートによって異常の有無を調査），18 歳–23 歳の 50 名を対象とした。そして，実験 A に 25 名が，実験 B に残りの 25 名が参加した。

2.3 実験環境

実験は上智大学荒井研究室の防音室で実施した。参加者の正面にはスピーカ（Genelec, 8020A）を 1 つのみ置き，その下には実験で使用するディスプレイを音声に遮りがないよ

---

* Perception of speech uttered as speaker faces different directions: Identification of speaker's facing direction from the listener, by KASHIMA, Haruna, ARAI, Takayuki (Sophia University), SUGIMOTO, Takehiro, KINOSHITA, Kôtarô, NAKAYAMA, Yasushige (NHK).

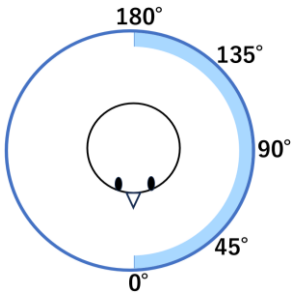

Fig. 1 Directions of a speaker

うに設置した。また，スピーカは 3 方の壁から 1 m 以上離し，スピーカの中心と参加者の受聴点を 1.5 m 以上離した。

2.4 実験手順

実験の前には,「スピーカの位置に話者がいるとし,その話者が向かって 45 度ずつ真後ろ（180 度）まで顔の向きを変えていったときの音声が流れます。それを聞いて，それぞれの音声では話者がどの向きを向いているかを 0 度，45 度，90 度，135 度，180 度の 5 方向でお答えください。」と Fig. 1 を用いて実験内容を教示した。教示は口頭のみで行い，説明の際に刺激は聞かせなかった。

実験 A・実験 B は，いずれも 3 つのセッションから構成された。それらのセッションを以後，実験①，②，③と呼ぶこととする。すべての参加者は，実験①，②，③の順番で実験に参加した。

実験①

実験①では，0 度–180 度のいずれかの向きに対して 20 代女性，20 代男性，30 代女性，30 代男性，40 代女性，40 代男性の順に計 6 名の音声を 1 話者につき 1 方向のみ聞かせ，角度を答えてもらった。

実験②

実験②では，2 つの音声を連続して聞かせた。1 つ目の音声（リファレンス音）は正面（0 度であることを開示）を使用し，2 つ目の音声（ターゲット音）はリファレンス音と同じ話者のある 1 方向を聞かせ，ターゲット音の放射方向を角度で答えてもらった。ビープ音・リファレンス音・ビープ音・ターゲット音の順番で流した。1 話者 5 方向で 6 話者分の全 30 個の刺激を，全参加者に聞かせた。なお，学習効果を防ぐために，前のペアと同じ話者のものは連続して聞かせないことを条件とした。

実験③

実験③では，20 代–40 代，いずれかの年代の男女 2 名の話者それぞれについて 5 方向の音声を聞き比べ，予め開示してある放射方向のうち，どの角度の音声かを回答してもらった。5 方向の刺激は何度でも聞くことができ，時間制限を設けない状況で行った。

3　結果

R (4.3.2) を用いてデータを Confusion matrix で表した。実験 A の実験①，実験②，実験③の結果をそれぞれ A-①，A-②，A-③，実験 B の実験①，実験②，実験③の結果をそれぞれ B-①，B-②，B-③とし，Fig. 2 に示す。

まず実験 A と実験 B の全体を比較すると，実験 A のほうが実験 B よりも全般に正答率が良いことがわかった。このことから，ラウドネスが手がかり（キュー）として発話者の向きの認識に寄与していたと考えられる。さらに，実験 A・実験 B それぞれにおいて，実験①の難易度が一番高く,次に実験②と続き,実験③がもっとも正答率が良かった。このことから，実験①のようにリファレンスがない場合に比べて，実験②や③のように比較対象がある場合に発話者の向きが認識しやすくなることが確認された。

4　考察

実験結果から，ラウドネスが発話者の向きの知覚におけるキューの 1 つとなっていることがわかった。一方，実験 A において 135 度と 180 度の正答率が逆になっていたため，各刺激の intensity を音響分析ソフトウェア Praat を用いて分析した。その結果，実験 B の平均 Intensity はほぼ一定であるのに対し，実験 A の平均 Intensity は 0 度を基準にすると，45 度から順に−1.4, −3.7, −8.5, −5.7 dB と下がる一方，135 度と 180 度が逆転していた。この理由としては，180 度においては話者の頭部を左右に回り込む音波が同位相となり加算される効果によるものと推測される。いずれにしても，ラウドネスの変化が正答率に寄与しており，ラウドネスが発話者の向きの知覚にお

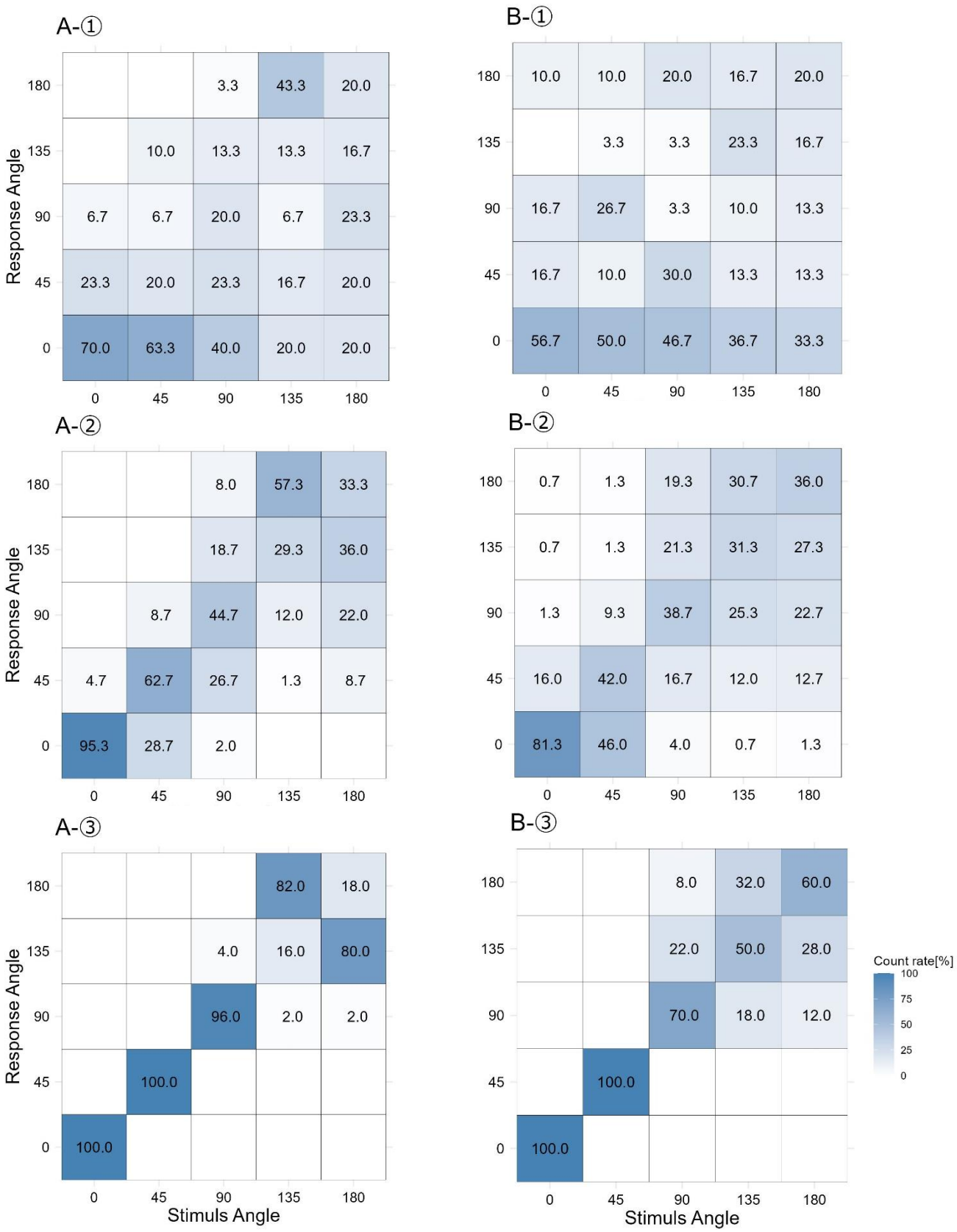

Fig. 2 Experimental results

けるキューの 1 つとなっていることを裏付ける結果となった。

実験 B ではラウドネスのキューを排除したが，実験②では全角度でチャンスレベルを超える正答率が得られ，③では 25 名中 8 名が全問正答した。よって，発話者の方向を判断するとき，ヒトはラウドネス以外に何らかのキューを用いて話者の向きを知覚していることが示唆された。

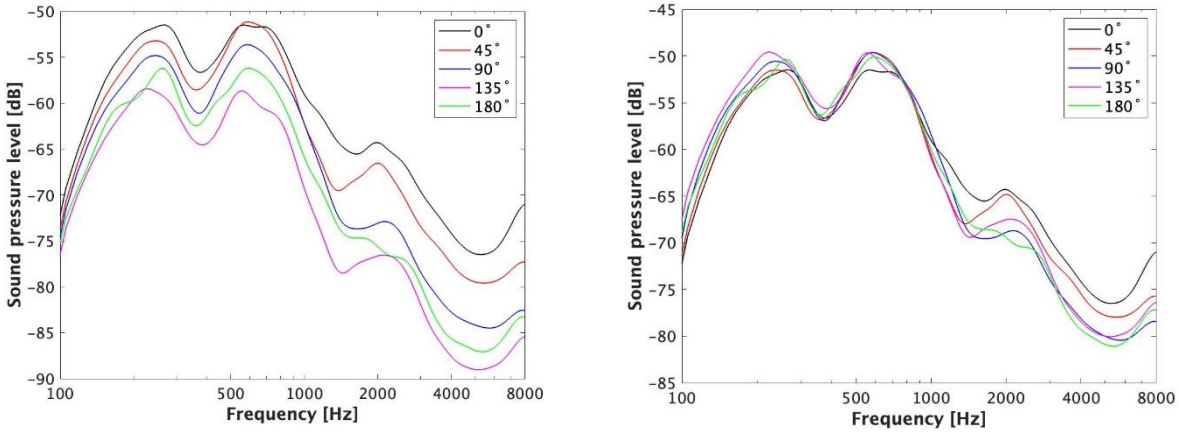

Fig. 3 Long-term average spectra of the sentences for five angles (left: Exp. A, right: Exp. B).

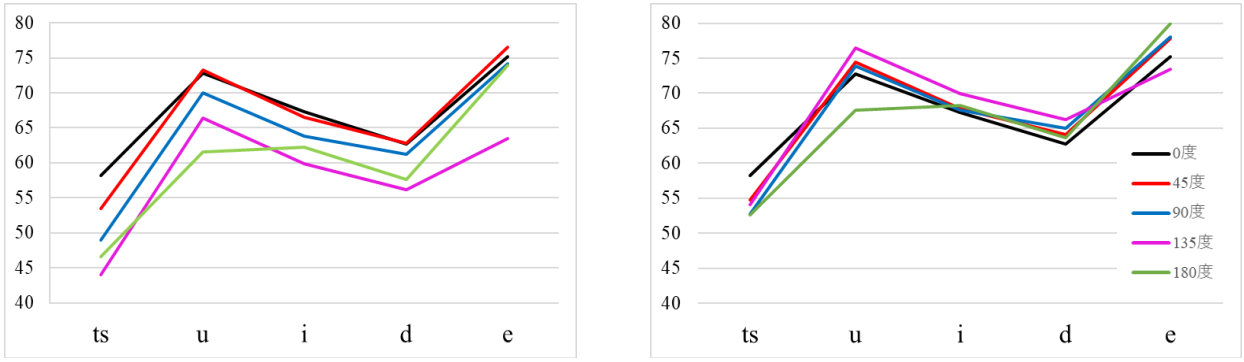

Fig. 4 Intensity variations for each segment (left: Exp. A, right: Exp. B)

また，実験Aと実験Bの共通点として 0 度と 45 度の正答率が高いことが挙げられるが，参加者の内観により，経験として聞きなれている 0 度や 45 度の知覚はできる一方，90 度より後ろを向いている際の判断が難しいことが示唆された。

ところで，例えばヒトは音の到来方向によるスペクトルの変化をキューに，音源定位をしうるという報告もある[6]。ラウドネスというキューが使えない場合，別のキューとして音声の放射特性によるスペクトルの変化を話者の方向判断のキューとして使用していた可能性が考えられる。そこで，以下では各刺激文に対するスペクトルの違いを考察した。

Fig. 3 に，Matlab を用いて角度ごとに長時間平均スペクトル（LTAS）を全話者にわたって平均した結果を示す。また，Praat を使って音素ごとに Intensity を算出した結果を Fig. 4 に示す(図は 20 代女性の「次いで」の Intensity)。これらの図から，平均ラウドネス値の正規化後も主に高周波域に差が認められ，また音素ごとによる違いも確認された。この違いが，キューになっていることが示唆された。

今後，残響下を想定した同様の実験も考えており，その場合は直接音・間接音の関係についても調べてみたい。

謝辞

本研究において，ご協力いただいた辻慎也氏，佐藤夢氏，鎌田光太氏，大塚美緒氏に心より感謝申し上げます。

参考文献

[1] 杉本，木下，音講論 (秋), 807–808, 2021.
[2] 木下他，音講論 (秋), 1067–1068, 2020.
[3] K. Kinoshita and T. Sugimoto, *Acoust. Sci. Tech.*, 44(4), 344–347, 2023.
[4] T. Sugimoto and K. Kinoshita, *Acoust. Sci. Tech.*, 44(5), 360–370, 2023.
[5] Rec. ITU-R BS.1770-5, 2023.
[6] M. M. Van Wanrooij and A. J. Van Opstal, *J. Neurotic.*, 24(17), 4163–4171, 2004.