Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

遠隔多人数会話のための発話音源定位分散の効果

野口康人[†1]　叶璟[†1]　井上智雄[†2]

ビデオ会議システムなどによる遠隔コミュニケーションの普及にも関わらず，その音声環境は現在でもモノラルまたは 2ch ステレオ程度であることが多い．しかし，多人数が遠隔から参加する場合に参加者の音声が同一スピーカを用いて再生されると聴き取りにくいと思われる．本研究では多人数会話における発話音源の定位を数箇所に分散させることの効果について実験的に検討した．同時発話時に認識できる単語数を測定したところ，母語，非母語に関わらず定位を分散させた方が聴き取りの成績がよく主観的にも効果的であること，少なくとも 1 名の話は聴き取れると期待できる同時話者人数の限界が 2 名から 3 名に向上することが分かった．

Effects of Distributed Auditory Localization for Remote Conversation by Multiple Participants

Yasuhito Noguchi[†1]　Kei You[†1]　Tomoo Inoue[†2]

It is often the case that audio is more important than video in remote conversation. Yet most existing remote conversation systems still have simple audio units such as monaural or 2ch stereo output. This is different from face-to-face conversation where different voices come from different locations. Distributed location of multiple voices is considered to make the listening easier but has not been investigated enough. We studied the effects of distributed auditory localization of multiple voices in different locations and in different languages. As a result, it was found that distributed localization is effective in recognizing multiple voices regardless of the languages in both objective and subjective measures.

1. はじめに

インターネットや携帯電話の急速な普及とブロードバンド化を背景に，音声会議システムやテレビ会議システムが複数のメーカーによって開発，流通されており，一般家庭においても日常的に利用できる環境が整えられつつある．このような遠隔会話システムが活用されている分野も遠隔会議，遠隔教育，遠隔医療など多岐に渡り，同時に複数人が参加する遠隔協調作業環境の需要は今後更に増大することが見込まれる．しかしながら現在の多くの遠隔会話システムでは，会話の相手が複数人であっても，音声出力するスピーカはひとつであり，同一方向から聴こえてくるような設計であることが多い．相手の映像を映す位置から音声が聴こえるように工夫されたテレビ会議システムは 1990 年代から研究されているが[1][2][3]，これらのシステムでは相手の人数分だけマイクやスピーカなどの音声機器を用意する必要があり，設備が大掛かりになってしまう．一方，仮想現実感を用いた会議システムでは主に相手の姿の見せ方に着眼が置かれているものの，仮想世界内での相手との位置関係を反映したシステムにおいては，相手と一定距離内に近づいたときに音声が聴こえるように制御したり[4][5]，相手との距離に応じて音声の音量の大きさを制御したり[6]した上で，ヘッドホンを用いて相手の音声を提示している．このように遠隔会話システムにおいては音声の制御も重要な要素のひとつとされており，本研究では遠隔会話において音源定位を分散させた場合の効果について検討する．音源定位をどのように実現するかは古くから検討されており，特に，左右耳における音声の強度差（両耳間強度差：Interaural Intensity difference：IID）と左右耳に到達する時間差（両耳間時間差：Interaural Time Difference：ITD）が重要な要素として多数の研究がなされている[7]．本研究では，IID と ITD を制御することにより，複数の音源を分離して定位させる方法を採用する．これにより複数地点をつなぐ遠隔会話であっても，参加者のいる各地点に 1 台の制御用 PC と 1 台の音声出力装置（ヘッドホン）さえあれば複数の音声の定位方向を分離させて出力することが可能となり，大掛かりな装置を必要としない．対面会話では話者のいる位置に応じてその音声が聴こえるのであり，遠隔会話においても話者ごとの音声が聴こえる方向を分離させた方が個々の音声の聴き取りやすさに良い影響を与える可能性がある．我々は，遠隔地にいる複数人の相手との会話のわかりやすさを向上させるため，遠隔地にいる参加者の音声を個々に定位させることの有効性について実験的に検討する．

†1 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies University of Tsukuba
†2 筑波大学図書館情報メディア系
Faculty of Library, Information and Media Science, University of Tsukuba

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

以下, 2 章で関連研究について述べ, 3 章にて実験内容, 4 章で実験結果について述べる．5 章はまとめである．

2. 関連研究

2.1 複数音声の聴き取り

複数の同時音声の中から個々の音声を聴き取る現象については, 20 世紀中ごろから多くの研究の対象となっている[8]．この現象の要因を解明するために，提示される音声刺激の強さによる違いや提示される音声同士の重なり具合による違い，声の高さやノイズが聴きやすさに与える影響等を明らかにしようする実験が行われてきた．一般に，競合話者が存在する場合には，話者同士の音声が重なる時間が長いほど，目的とする話者の音声を聴き取りづらくなる[9]．しかしながら，目的話者と競合話者が空間的に離れている場合には目的話者の音声を選択的に聴取することが容易になるのも事実であり，このような両耳受聴に基づく現象はカクテルパーティ効果と呼ばれている[10]．そして，この効果を利用することによって，競合話者が存在する場合の単語了解度が向上することが複数の研究によって示されてきた．聴き取り時の単語了解度に与える要因として，音圧の違い[11][12]，音声の周波数スペクトルの違い[13]，複数音声の重なり具合[14][15]，各音声刺激の空間的配置の関係[16]，音声刺激の提示手法[17][18]，空間における音の反響度合い[19][20]，聴取者の注意の分散度合い[21]，聴取者の聴力の正常さ[22]，話し手の性別差[23]等様々なものがあり，それぞれ実験的に検討されている．

2.2 複数音声の空間的分離の効果

複数の音声の空間的配置を分離させることの効果について着目した研究は数多くあり，提示手法も様々である．まず単純に複数の音声を分離する方法して，複数のスピーカを用いて空間的に分離して配置して出力するもの[24][25][26]がある．しかしながら複数のスピーカを用いて参加者ごとの音声出力位置を分離させるような遠隔会話システムは，先に述べたように設備を用意することの負担が大きい．また，バイノーラル録音方式を用いて 3 次元空間の音場をそのまま録音し提示するもの[27]がある．バイノーラル録音方式においては人間の頭部の音響効果を再現するダミーヘッド・マイクロフォン等を利用して録音する必要がある．複数人の音声を分離して定位させるだけであればIIDとITDのみの制御で, より手軽に実現できる. 一方，本研究でも用いる IID や ITD を含む頭部伝達関数（Head-Related Transfer Function：HRTF）を使用して音声刺激を空間的に分離して提示するものがある．これらはモノラル方式の音源をステレオ 2 チャンネル方式で出力する際に仮想的に空間的情報を付与し，提示しようとするものである．しかしながら，これらの実験では，目的話者の音声に対する競合刺激として雑音が用いられていたり[28][29]，まったく同じ話し手の音声が競合刺激として用いられていたりする[30]．現実的に起こりうる競合話者が存在する遠隔会話の場面を想定して実験条件の設定を行う必要がある．

2.3 分散的聴取

音声分離知覚の研究の多くに共通する点として，目的話者の音声を定義する情報をあらかじめ知っているという点がある．例えば，目的話者の音声が呈示される空間的な位置をあらかじめ知っているもの[24][25]，目的話者の音声のみに含まれるキーワードをあらかじめ知っているもの[23][24]，目的話者の音声の音質をあらかじめ知っているもの[16][28]などである．しかしながら，現実的な状況では，聴取者は複数の声のうち聴き取りたいものの特徴をあらかじめ知っているとは限らない．むしろ聴き取る対象を特定せずに複数の音声を聴き取った上で必要な情報を取捨選択したい場合もある．本研究では目的となる音声を設定せず，複数の音声を同時に聴いた場合にどの程度網羅的に聴き取れるかについて検討する．このような，特定の音声を選択的に聴き取ろうとするのではなく，同時複数音声に含まれるすべての声を聴き取るべき対象の音声とする分散的聴取（divided listening）課題について実験を行っている研究もいくつかある[12][21][31]．これらの研究では，複数音源を空間配置的に分離させることが分散的聴取において有効であることを報告している．しかしながら，これらは KEMAR ダミーヘッド･マイクロフォンを用いて空間的に分離させた音源を使用している．これに対して本研究ではダミーヘッド・マイクロフォンのような特殊な機器を用いることなく，一般的な設備環境を想定する．また，これらの研究では同時に呈示される音声の話者人数が 2 名の場合において実験がなされており，3 名以上の人数については検討されていない．しかしながら，遠隔会話システムの利用シーンを考えると，4 名以上が分散している場合の会話において用いられる場合もあり，3 名以上の話者が同時に発話することも想定される．このため，本研究では話者が最大 5 名の場合の分散的聴取についても検討する

話者が 3 名以上の場合の分散的聴取について実験を行った研究は，数は多くないものの，いくつかある[32][33]．これらの研究では 2，3 名程度の聴き分けが可能であることが報告されている．ただし，これらの実験では単純に複数の音声を合成して再生を行っており，複数音声を個別に音源定位をした場合にどの程度聴き取れるようになるかについては検討されていない．本研究では 2〜5 名の話者の音声を，IID と ITD の制御によって分離して定位させた場合の聴き取りやすさへの影響について実験的に検討する．また，多くの先行研究では母語のみを用いて実験を行っているが，近年のビジネスのグローバル化に伴い，母語が異なる者同士のコミュニケーションの必要性も高まっている．本研究では非母語を用いた場合の実験条件も設定し，使用する言語の種類が聴き取りやすさに影響するかどうかにつ

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

いても検討する．

3. 実験

3.1 実験目的

遠隔地に会話の相手が複数人いる場合，発言が衝突した際に会話内容を聴き取りづらくなると考えられる．本実験では，多人数同時発話に対して音声を分散して定位することでそれぞれの発言内容が聴き取りやすくなるか否かを明らかにする．具体的には，モノラル方式で録音した複数の音声を被験者が自分の感覚に合わせて定位した場合，実際にそれらの音声を正しく聴き取れるかについて検証する．

3.2 被験者

被験者は聴力が正常な中国人の成人 20 名（女性 14 名，男性 6 名）である．被験者は全員，日本国際教育支援協会と国際交流基金の主催する日本語能力試験において N2 レベル以上であり，日常的な場面で使われる日本語の理解に加え，より幅広い場面で使われる日本語をある程度理解することができる．在日平均年数は約 1 年であり，平均年齢は 23.0 歳（標準偏差：1.71）である．

3.3 実験条件

本実験では同時に再生する単語数を，中国語は 2、3、4、5 語の 4 条件，日本語は 2、3、4 語の 3 条件を用意した．さらに音源定位の効果を測るため，用いる単語数に応じて複数の音源を均等に分散させて定位する複数音源条件と，用いる単語数に関わらず，全ての音源を分散させずに中央に定位する単数音源条件を用意した．実施した実験条件をまとめると，単語数要因は中国語 4 条件，日本語 3 条件，音源数要因は 2 条件を有する．すなわち 2 要因 8 または 6 条件の被験者内実験である．表 1 に実験条件を示す．本研究では音源を仮想的に分散させることの効果を検証することが目的であるため，必ずしも 3 次元的に音源を分離させる必要はなく，左右方向の 1 次元的な分離のみを扱う．

3.4 実験環境

図 1 に実験環境を示す．事前に本実験用に開発したソフトウェアの入ったパーソナルコンピュータおよびキーボード，マウス，密閉型のヘッドホン（JVC 社の HA-XS10X）を用意した．実験は騒音のない環境で実施した．被験者は本実験用に開発したソフトウェアを用いて，両耳間の音量バランス，時間差を調整し，自身の手で任意に音源定位が可能である．ソフトウェア構築環境として，OS は Windows 7，数値計算言語には MathWorks 社の MATLAB 7.0[34]を使用した．図 2 に音源定位のパラメータ設定時の画面を示す．被験者は画面中央の図（図 2 中の図は最も左側から聴こえるように設定することを指示している）に従い，左右のスライダを上下に調整することで音源定位することができる．

表 1　実験条件

Table 1　Experiment Pattern

音源数 / 単語数	複数音源	単数音源
2 単語	2 単語 2 音源条件：左，右に定位	2 単語 1 音源条件：中央に全て定位
3 単語	3 単語 3 音源条件：左，右，中央に定位	3 単語 1 音源条件：中央に全て定位
4 単語	4 単語 4 音源条件：左，右，左と中央の中間，右と中央の中間に定位	4 単語 1 音源条件：中央に全て定位
5 単語	5 単語 5 音源条件：左，右，左と中央の中間,右と中央の中間，中央に定位	5 単語 1 音源条件：中央に全て定位

左側のスライダを調整することで，IID バランス（左：右）を 0：10～10：0 の整数 11 段階で設定できる．また，右側のスライダを調整することで，ITD を 0ms～5ms の範囲で 0.05ms 刻みの 100 段階で設定できる．右のスライダを一番上まで引き上げると 5ms となり，一番下まで引き下げると 0ms となる．

ITD の代表的なモデルとして，Kuhn や Woodworth と Schlosberg の計算式がある[35][36]．これらではそれぞれ人の頭部の大きさを変数の一つとして組み込んで計算式を提示している．さらに，3D 化の効果を高めるために ITD を大き目に設定した音声システムもある[37]．このように，頭部の中心から左右耳の延長線上に音源を定位する場合の適切な ITD の値についても諸説あり，一つの値に定めるこ

図 1　実験時の被験者の様子

Figure 1　Actual Scene of the Experiment

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

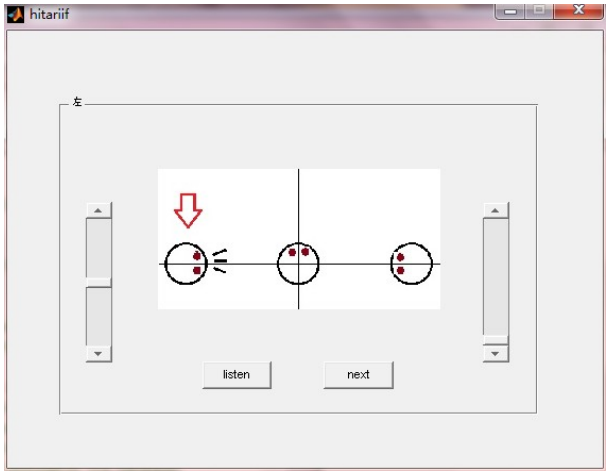

図 2　音源定位用パラメータ設定時の画面
Figure 2　The Interface of for Setting Up the Individual Audio Parameters

とはできない．頭部の大きさや音の聴こえ方は個人差もあるため，本ソフトウェアでは各被験者が自身の感覚に合わせて設定できるようにした．

使用する中国語，日本語の単語は 4 モーラで統一した．中国語の単語は高校一年生の語学教科書，人教版全日制普通高中教材第一冊[38]に記載されている四字熟語を採用した．音声ファイルは母国語を中国語とする男性留学生 5 名の音声を録音した計 250 語を使用した．日本語の単語は，単語了解度リスト[39]（全 4380 語）にある単語を採用した．単語了解度は，多くの音声通信システムや補聴器，客席の音声伝達性能の評価に用いられており，その中でも了解度が比較的高いとされる親密度 7.0～5.5 のリスト（全 1095 語）から選定した．日本語の音声ファイルは，日本語 NHK アクセント辞典[40]の付録である男性アナウンサー4 名の音声を録音した計 130 語を使用した．被験者に提示される音声の音量は，同時に再生される単語数に関わらず同一にするため，音量正規化処理を施した．

3.5 実験手順

被験者には機器の操作方法について説明し，事前準備，実験の順に取り組ませた．事前準備では左右方向の音源定位パラメータを設定させ，次に日本語の 2～4 単語実験をランダムな順番で実施し，最後に中国語の 2～5 単語実験をランダムな順番で実施した．

被験者は，図 2 の画面中の図に従い，まず左方向に音源定位する場合のパラメータについて，左右に配置されたスライドを動かすことによって設定する．被験者には 2 つのスライダがどのような意味をもつかについては知らせず，「画面中の矢印が指す方向，つまり最も左側または最も右側から聴こえるように自由に調整してください」とのみ伝えた．「listen」ボタンを押すと，現在のパラメータ設定で定位された音声ファイルが流れる．ボタンを押下することで何度でも聴こえ方を確認し，設定し直すことができる．

被験者は，左方向の設定を終えたら「next」ボタンを押下し，右方向の音源定位設定に移行する．左右両方の音源定位の設定が完了したら実験に移行する．

各単語数の実験について，複数音源条件，単数音源条件を 5 回ずつ連続して実施した．複数音源条件を先に実施する被験者の人数と単数音源条件を先に実施する被験者の人数を同一にすることでカウンターバランスをとった．条件が切り替わる際には，合図として画面に「attention!!」の文字列を表示した．被験者にはあらかじめ合図が表示された前後で聴こえ方が変わることのみ教示しておいた．実験では音声ファイルは「listen」ボタンを押下したタイミングの一度のみ再生することができる．被験者は聴き取れた単語を入力フォームに入力した．図 3 に 2 単語実験の場合の入力フォーム画面を示す．単語数が増えるに従い，「a, b」と示されている単語入力欄も「a, b, c」,「a, b, c, d」と増える．被験者に提示される音声ファイルは，ソフトウェアが「どの話者のどの単語か」をランダムに選定する．同一被験者に同一の音声ファイルが 2 度以上用いられることのないように制御した．

3.6 質問紙調査方法

質問紙調査では，事前準備時の音源定位の操作性，使用した単語の難易度，再生速度，音源の分散状況による聴こえ方の違いに関する印象，音源の分散状況の差による聴こえやすさへの影響について調査した．質問項目は計 8 項目で構成し，回答は「1:全く当てはまらない」,「2:当てはまらない」,「3:あまり当てはまらない」,「4:どちらでもない」,「5:やや当てはまる」,「6:当てはまる」,「7:よく当てはまる」の 7 段階のリッカート尺度で行い，それぞれ 1～7 点を対応付けた．質問項目は，Q1「左右の音源定位の設定が難しかった」，Q2「単語が難しかった」，Q3「再生速度が速かった」，Q4「合図の画面前後で聴こえ方が違った」，Q5「2 語の場合，合図を表示する前の音声の方が聴きやすい」，Q6「3 語の場合，合図を表示する前の音声の方が聴

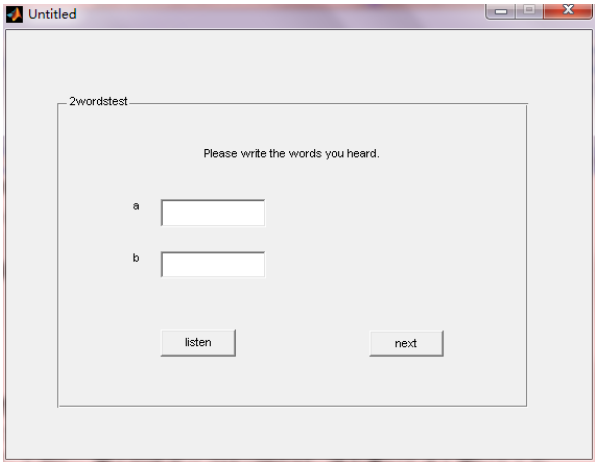

図 3　2 単語実験時の画面
Figure 3　The Interface of Two-word Experiment

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

きやすい」，Q7「4 語の場合，合図を表示する前の音声の方が聴きやすい」，Q8「5 語の場合，合図を表示する前の音声の方が聴きやすい」である．Q1～Q7 を日本語実験後，Q2～Q8 を中国語実験後にそれぞれ実施した．Q1 は音源定位パラメータ設定時の操作性に関する質問であるため，先に実施した日本語実験後のみ，Q8 は 5 単語実験に関する質問であるため，中国語実験後のみで実施した．

4. 実験結果

4.1 音源定位用パラメータ

事前準備時の，左右方向の音源定位については被験者 1 名に対し 3 回ずつ実施し，20 名の被験者で 60 サンプルを得た．その結果，左方向の音源定位のパラメータの平均値は，IID バランスが 8.8：1.2（左：右，標準偏差（以下，SD と表す）：0.9），ITD（右耳への音声出力時刻から左耳への音声出力時刻を引いた時間）が 1.3ms(SD：1.5)であった．同様に，右方向の音源定位のパラメータの平均値は，IID バランスが 1.3：8.7（左：右，SD：1.0），ITD（左耳への音声出力時刻から右耳への音声出力時刻を引いた時間）が 2.1ms（SD：1.8）であった．

4.2 正答率

図 4 に平均正答率を示す．エラーバーは標準誤差を示す．図中では全ての単語を正確に入力できた場合を正答率 100%とする．例えば，2 単語実験の場合，2 問中 2 問正解で 100%，2 問中 1 問正解であれば 50%，1 問も正解できなければ 0%となる．各言語，各単語数実験の複数音源条件と単数音源条件の結果を Wilcoxon の符号付順位検定を用いて比較した．この結果，中国語 5 単語実験，日本語 4 単語実験を除く全ての実験について，複数音源条件の方が単数音源条件よりも有意水準 5%で正答率が高かった（中国語：2 単語 Z=-3.582, p=0.0003, 3 単語 Z=-3.002, p=0.003, 4 単語 Z=-2.227，p=0.026，5 単語 Z=-0.954，p=0.340，日本語：2 単語 Z=-3.744，p=0.0002，3 単語 Z=-2.42，p=0.015，4 単語 Z=-1.386，p=0.166）．また，中国語，日本語の両言語実験それぞれについて，単語数が多くなればなるほど正答率が低下することが分かった．

4.3 質問紙調査結果

質問紙調査の結果について，図 5 に質問項目毎の平均得点を示す．エラーバーは標準誤差を示す．質問項目 Q5～Q8 は，被験者が複数音源条件と単数音源条件のどちらを先に聴いたかによって回答の意味合いが反対となるため，先に単数音源条件を受けた被験者の回答を反転項目として取扱い，集計した．よって Q5～Q8 の結果は，複数音源条件の方が単数音源条件よりも聴きやすいと感じたかどうかの結果を示しているといえる．

図 5 の結果から，事前準備時の音源定位の操作性，使用した単語の難易度，再生速度に関する Q1～Q 3 の平均得点は「どちらでもない」という回答である 4 点の前後 1 点以内の数値であり，難しすぎず簡単すぎず，早すぎず遅すぎず適度であったと言える．Q4 の音源の分散状況による聴こえ方の違いに関する印象については，中国語での実験時に 6.1，日本語での実験時に 5.7 という高得点であり，印象的であったと考えられる．また，Q 5～Q 8 について，複数音源条件の方が単数音源条件よりも聴きやすいかどうかについて，両言語において「どちらでもない」という回答の 4 点を上回る結果となった．このことから，被験者は使用する言語や同時に発声される単語数に関わらず，複数音源の方が単数音源よりも聴きやすいと感じていることが分かった．

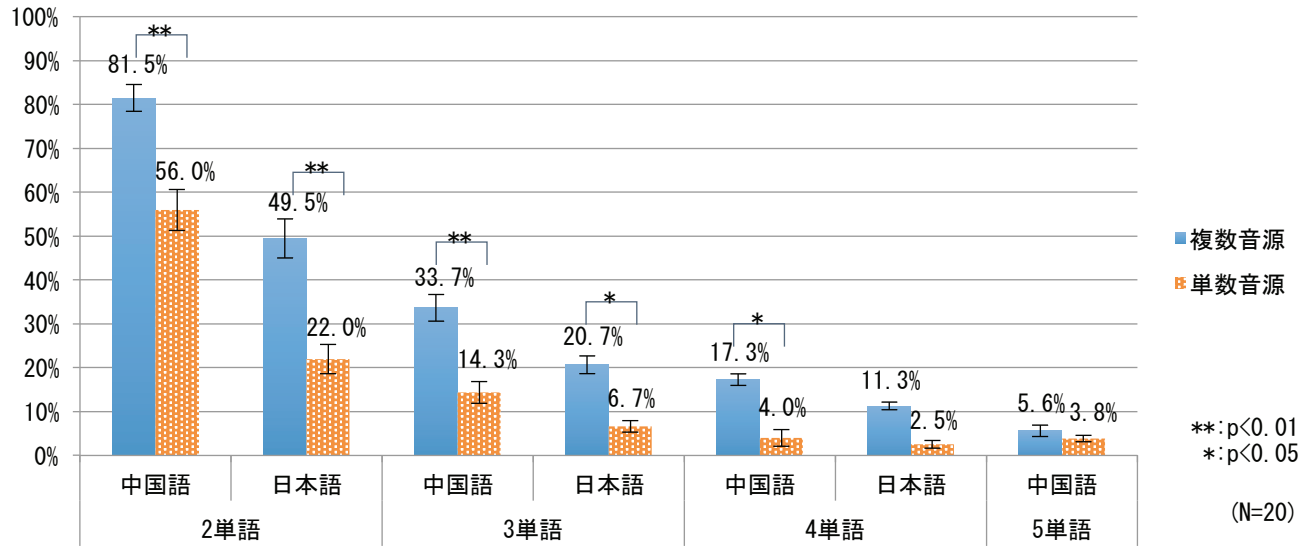

図 4　平均正答率

Figure 4　The Rate of Correct Answers

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

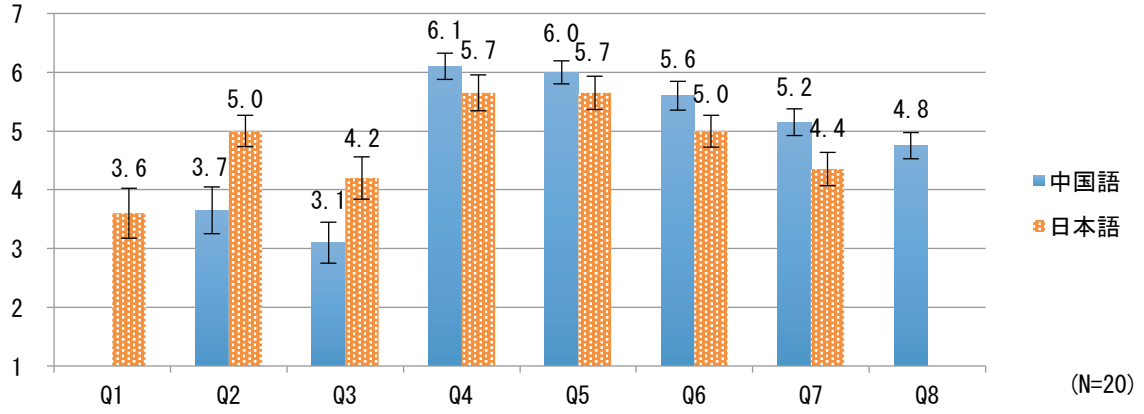

図 5　質問項目毎の平均得点

Figure 5　The Average Point of Questionnaire

4.4 検討

4.4.1 正答率

中国語 5 単語実験，日本語 4 単語実験を除く全ての実験について，複数音源条件の方が単数音源条件よりも正答率が高かった．特に，より聴き分けがしやすいと考えられる少単語数実験においてその差は顕著であった．中国語の結果の方が日本語の結果よりも総じて正答率が高いのは，被験者の母語が中国語であったためだと考えられる．一方，中国語では 5 単語，日本語では 4 単語実験において両条件の正答率に差が認められなかった．母語では 4 単語まで，非母語では 3 単語までにおいて，音源を分離して定位することの有効性を示すことができた．

2.3 節においても記述したが，複数音源を空間配置的に分離させることが分散的聴取における聴き取りに有効であることは先行研究でも言及されている[12][21][31]．これらの研究では KEMAR ダミーヘッド･マイクロフォンを用い，空間的に分離した 2 つの音源を同時に聴いた場合について検討している．本研究においては 2 単語実験における複数音源条件と単数音源条件の正答率の比較に相当する．本研究においては IID，ITD を用いた音源定位で複数の音源を分離させており，手法の上で違いはあるものの，音源の定位を分離させることの有効性については同様の結果が得られた．Shinn-Cunningham らは刺激の音量の強さが強いほど聴き取りの正答率が高まることを報告している[21]．本実験における音声刺激は用いる音源の数によって正規化されており，音源の数が多くなればなるほど各音源の音量は小さくなる．本実験における単語数が多くなればなるほど正答率が低くなっているという結果は，各音源の音量の大きさも関係している可能性がある．また，Best らは音声刺激の定位の分離角度をより広げた方が聴き取りの正答率が高まることを報告している[31]．本実験においては 2 単語実験が最も大きく分離されており，5 単語実験が最も小さく分離されている．音量と同様，本実験における単語数が多くなればなるほど正答率が低くなっているという結果は，各音源の分離程度の大きさが関係している可能性がある．

話者が 3 名以上の分散的聴取に関しては，柏野らが複数話者により同時発声された単語を提示し，被験者に話者人数を回答させる実験を行っている．その結果，課題の正答率が，話者が 2 名時にはほぼ 100%であるのに対し，3 名以降急速に低下し，話者が 3 名の場合に，話者人数が 3 名であると正しく回答できた割合はたかだか 2 割であると報告している[32]．川島らは同時複数音声再生の後に，その中から無作為に選択する単独音声を提示し，複数音声の中に単独音声が含まれていたかどうかを選択させる方法で分散的聴取における知覚限界を探った．その結果，分離知覚に関わる認知的処理効率の限界（注意の限界）が話者数 2，3 名であることを推測している[33]．これらの実験における音声の呈示は本実験での単数音源条件にあたり，母国語での正答率の結果は 2 単語の場合に 56.0%，3 単語では 14.3%，4 単語では 4.0%，5 単語では 3.8%と 3 単語以上の場合に正答率は大きく低下した．関連研究における実験ではそれぞれ，話者数と特定音声の有無の判定であるので評価対象が本研究とは異なっているが，同時発話数が増加するにつれて単語の認識率が低下するという傾向は一致した．

中貝らは遠隔協調作業環境を想定し，画面に映った 3 名の話者が同時に発声した別々の単語のうち，あらかじめ指定された話者の単語を選択的に聴き取らせるという実験で，バイノーラル録音再生方式による単語了解度がモノラル方式，2 チャネルステレオ方式よりも高いことを報告している[41]．本研究ではバイノーラル方式ではなく，IID，ITD による定位を用いているが，音源を分散して定位させた場合の単語了解度の有効性を示している点は一致している．ただし，本研究では 3 点定位のみならず，2 点から 5 点ま

Vol.2015-GN-93 No.41
Vol.2015-CDS-12 No.41
Vol.2015-DCC-9 No.41
2015/1/27

で評価し，了解度の変化を検討することで音源定位の適当な分散数を得ている．

図 4 から，母語を用いた場合の複数音源条件の正答率は，2 単語で 81.5%，3 単語で 33.7%，4 単語で 17.3%，5 単語で 5.6%であった．同時話者数によって正答率の意味合いは異なり，聴き取れた話者数が 1 名の場合は 2 単語実験では 50%，3 単語事件では 33.3%，4 単語実験では 25%，5 単語実験では 20%となる．また，2 単語実験での正答率が 75%を超えているということは，聴き取れた単語数が 2 個であるケースの方が 1 個であるケースよりも多いことを意味する．つまり，同時話者数が 2 名の場合には聴き取れた話者数が 2 名であった場合の方が 1 名であった場合よりも多いといえ，同時話者数が 3 名の場合には聴き取れた話者数はおよそ 1 名であったといえる．同時話者数が 4 名の場合には聴き取れた話者数が 1 名であった場合の方が 0 名であった場合よりも多いといえ，同時話者数が 5 名の場合には聴き取れた話者数が 1 名であった場合の方が 0 名であった場合よりも少ないといえる．一方，単数音源条件では 2 単語で 56.0%，3 単語で 14.3%，4 単語で 4.0%，5 単語で 3.8%であった．つまり，同時話者数が 2 名の場合には聴き取れた話者数がおよそ 1 名であったといえ，同時話者数が 3 名の場合には聴き取れた話者数が 1 名であった場合の方が 0 名であった場合よりも少ないといえる．また，同時話者数が 4 または 5 名の場合には正答率は大きく低下し，多くのケースで 1 名も聴き取れなかったことが分かる．以上のことから，音源を複数に分散させることによって，特に 2～4 名が同時に発話する場面において聴き取ることのできる話者の人数が増加することが分かった．また，単数音源条件では話者 1 名の発話内容も聴き取れていないケースが半数以上を占めるのが同時話者 3 名以上の場合であると推察されるが，複数音源条件では同時話者 4 名の場合においても話者 1 名の発話内容が聴き取れているケースが全く聴き取れていないケースよりも多いと考えられる．複数音源の定位を分散させることにより，少なくとも 1 名の話は聴き取れると期待できる同時話者人数は 2 名から 3 名に向上したといえる．

4.4.2 質問紙調査結果

図 5 から，主観的評価においても複数音源を分離させることが聴きやすさに効果的であることが分かった．また，2 語の場合の音源を分散させることの印象を問う Q5 では，中国語実験で 6.0，日本語実験で 5.7 という高得点であるのに対し，3 語の場合を示す Q6 ではそれぞれ 5.6，5.0，4 語の場合を示す Q7 ではそれぞれ 5.2，4.4 とその得点が低下している．同時話者の人数が多くなるに従って，複数音源の分散の効果の印象も弱まる傾向にあると言える．

その一方で，中国語，日本語の結果を比較したところ，質問項目 Q2「単語が難しかった」，Q3「再生速度が速かった」については日本語の結果の方が高得点であった．被験者は中国人であるため，非母語の日本語を用いた実験の方が母語の中国語を用いた実験よりも，使用した単語を難しいと感じていたと推察する．また，Q3 の結果から，中国語実験と日本語実験で再生速度は同一にしていたが，再生速度の感じ方に差が生じたことが分かった．これは，使用する言語による聴き取り能力の差に起因するものと考えられる．また，複数音源を分散して定位させることの効果を問う質問項目について，2 単語の場合に関する Q5 では両言語間でさほど差がないことに対し，3 単語の場合に関する Q6，4 単語の場合に関する Q7 では中国語の結果の方が高得点である傾向にあった．2 単語よりも 3 単語，さらに 3 単語よりも 4 単語実験において被験者は中国語実験の方が日本語実験よりも音源を分散させることの効果をより強く感じていたことが分かる．このことから，聴き取りやすい条件よりも多少聴き取りづらい条件においては，非母語を用いた場合よりも母語を用いた場合の主観的評価に，音源定位の分散がより効果的であったと推測される．

5. まとめ

本研究では，音声の定位を分離することが，複数音声の聴き分けに効果的かどうかは明らかでないため，実験のためのソフトウェアを実装した上で検討した．

この結果，複数の音声を個々に定位し複数位置から聴こえるように同時再生した場合は，個々に定位せずに同時再生した場合に比べ，同時に聴き取れる話者の人数が増大することが分かった．また，複数音源の定位を分散させることにより，少なくとも 1 名の話を聴き取れると期待できる同時話者人数が 2 名から 3 名に向上することが分かった．一方，質問紙による主観的な評価でも，複数の音声は個々に定位することが「聴きやすさ」に大きく影響し，より聴きやすいという印象を与えることが示唆された．また，これらの知見は母語，非母語に関わらず有効であることが分かった．

謝辞 本研究の一部は，科学研究費補助金 26330218 の支援による．ここに記して謝意を表す．

参考文献

1) A. J. Sellen: Speech patterns in video-mediated conversations, Proc. ACM CHI '92, pp.49–59, (1992).
2) K. Okada, F. Maeda, Y. Ichikawa, and Y. Matsushita: Multiparty videoconferencing at virtual social distance: MAJIC design, Proc. ACM CSCW'94, pp.385-393, (1994).
3) 井上智雄，岡田謙一，松下温: 空間設計による対面会議と遠隔会議の融合 ：テレビ会議システム HERMES，電子情報通信学会論文誌．D-II，情報・システム，II-情報処理 Vol.J80-D-2，No.9，pp.2482-2492，(1997).
4) S. Benford and L. Fahlén: A Spatial Model of Interaction in Large Virtual Environments, Proc. of the Third European Conference on CSCW (ECSCW'93), pp.109-124, (1993).
5) C. Greenhalgh and S. Benford: MASSIVE: a collaborative

virtual environment for teleconferencing, ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 2, No. 3, pp. 239–261, (1995).
6) 田尻 哲男, 島村 和典: サイバースペースにおける通信サービスの一提案, 情報文化学会誌 Vol.3, No.1, pp.76-80, (1996).
7) 黒住幸一, 大串健吾: 音響信号の両耳間相関関数に基づく音像定位の予測モデル, 日本音響学会誌 Vol.44, No.10, pp.726-734, (1988).
8) E. C. Cherry: Some experiments on the recognition of speech, with one and with two ears, J. Acoust. Soc. Am. Vol.25, pp.975-979, (1953).
9) G. A. Miller: The masking of speech, Psychol. Bull. Vol.44, pp.105-129, (1947).
10) A. W. Bronkhorst: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions, Acustica Vol.86, pp.117-128, (2000).
11) J. P. Egan, E. C. Carterette and E. J. Thwing: Some factors affecting multi-channel listening, J. Acoust. Soc. Am. Vol.26, pp.774-782, (1954).
12) A. Ihlefeld and B. G. Shinn-Cunningham: Spatial release from energetic and informational masking in a divided speech identification task, J. Acoust. Soc. Am. Vol.123, pp.4380-4392, (2008).
13) Festen JM, Plomp R.: Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing, J Acoust Soc Am. Vol.88, No.4, pp.1725-36, (1990).
14) J. C. Webster and P. O. Thompson: Responding to both of two overlapping messages, J. Acoust. Soc. Am. Vol.26, pp.396-402, (1954).
15) R. Carhart, T. W. Tillman and E. S. Greetis: Perceptual masking in multiple sound backgrounds, J. Acoust. Soc. Am. Vol.45, pp.694-703, (1969).
16) D. R. Begault and T. Erbe: Multichannel spatial auditory display for speech communication, J. Audio Eng. Soc. Vol.42, pp.819-823, (1994).
17) R. Drullman, A. W. Bronkhorst: Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation, J Acoust Soc Am., Vol.107, No.4, pp.2224-2235, (2000).
18) N. W. MacKeith and R. R. A. Coles: Binaural advantages in hearing of speech, The Journal of Laryngology & Otology, Vol.85, No.03 pp.213-232, (1971).
19) J. P. Moncur, D. Dirks: Binaural and monaural speech intelligibility in reverberation, J Speech Hear Res., Vol.10, No.2, pp.186-195, (1967).
20) A. K. Nábělek and J. M. Pickett: Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids, J. Acoust. Soc. Am. Vol.56, pp.628-639, (1974).
21) B. G. Shinn-Cunningham and A. Ihlefeld: Selective and divided attention：Extracting information from simultaneous sound sources, Proc. of the International Conference on Auditory Display, 2004.
22) A. J. Duquesnoy: Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons, J Acoust Soc Am., Vol.74, No.3, pp.739-743, (1983).
23) D. S. Brungart, B. D. Simpson, M. A. Ericson and K. R. Scott: Informational and energetic masking effects in the perception of multiple simultaneous talkers, J. Acoust. Soc. Am. Vol.110, pp.2527-2538, (2001).
24) R. L. Freyman, U. Balakrishnan and K. S. Helher: Spatial release from informational masking in speech recognition, J. Acoust. Soc. Am. Vol.109, pp.2112-2122, (2001).
25) R. L. Freyman, U. Balakrishnan and K. S. Helher: Effect of number of masking talkers and auditory priming on informational masking in speech recognition, J. Acoust. Soc. Am. Vol.115, pp.2246-2256, (2004).
26) T. L. Arbogast, C. R. Mason and G. Kidd, Jr.: The effect of spatial separation on informational and energetic masking of speech, J. Acoust. Soc. Am. Vol.112, pp.2086-2098, (2002).
27) W. A. Yost, R. H. Dye and S. Sheft: A simulated 'cocktail party' with up to three sound sources, Percept. Psychophys. Vol.58, pp.1026-1036, (1996).
28) J. Peissig and B. Kollmeier: Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners, J. Acoust. Soc. Am. Vol.101, pp.1660-1670, (1997).
29) A. W. Bronkhorst, R. Plomp: Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing, J Acoust Soc Am. Vol.92, No.6, pp.3132-3139, (1992).
30) M. L. Hawley, R. Y. Litovsky and H. S. Colburn, Speech intelligibility and localization in a multi-source environment, J. Acoust. Soc. Am. Vol.105, pp.3436-3448, (1999).
31) V. Best, F. J. Gallun, A. Ihlefeld, B. G. Shinn-Cunningham: The influence of spatial separation divided listening, J Acoust Soc Am, Vol.120, No.3, pp.1506-1516, (2006).
32) 柏野牧夫, 平原達也：一度に何人の声を聞き分けられるか？, 音講論集, pp.467-468, (1996).
33) 川島尊之, 佐藤隆夫: 同時複数音声の分散的聴取における知覚限界, 日本音響学会誌, No.65, Vol.1, pp.3-14, (2008).
34) MathWorks: http://www.mathworks.co.jp/products/matlab/, (2014/12/24 参照)
35) G. F. Kuhn: Physical Acoustics and Measurements Pertaining to Directional Hearing, Directional Hearing, pp.3-25, (1987).
36) J. W. Kling and L. A. Riggs: Woodworth & Schlosberg's experimental psychology, Holt, Rinehart and Winston, (1971).
37) V. Hardman and M. Iken: Enhanced Reality Audio in Interactive Networked Environments, in Proceedings of the Framework for Interactive Virtual Environments, pp.55-66, (1996).
38) 人民教育出版社中学語文室: 人教版全日制普通高中教材第一冊, 人民教育出版社, (2002).
39) 東北大学 電気通信研究所: http://www.ais.riec.tohoku.ac.jp/lab/wordlist/index-j.html, (2014/12/24 参照).
40) NHK 放送文化研究所: NHK 日本語発音アクセント辞典 新版, 日本放送出版協会, (1998).
41) 中貝順一, 小澤 賢司：音の再生方式と高能率符号化が競合話者存在下での単語了解度に及ぼす影響, 電子情報通信学会論文誌. A, 基礎・境界 Vol.J88-A, No.9, pp.1026-1034, (2005).