

遠隔複数人会話をわかりやすくするための音像定位の検討

叶環^{†1} 野口康人 井上智雄^{†2}

遠隔協調作業や遠隔会議では音声映像よりも重要であることも多い。近年、遠隔会話システムが一般に普及してきたが、多くの遠隔会話システムでは、電話と同様に1対1の会話を想定しており、音声は1箇所から提供される。対面会話では複数の話者がいる場合には、話者の位置に応じてその音声が聞こえるが、このようなシステムでは会話相手が複数であっても、音声は同一箇所から聞こえてくる。本研究では、複数話者の音声は同一箇所から聞こえる場合と、複数箇所から聞こえる場合で、音声の聞き取りやすさがどのように異なるかを、複数単語を同時に提示してその聴取結果を調べる実験により検討した。その結果、音源の定位箇所を複数与える方が、聴取しやすいことが分かった。さらに実験結果にもとづいた遠隔音声会話システムについても論じた。

Study of auditory positioning for remote conversation by multiple participants

KEI YOU^{†1} YASUHITO NOGUCHI TOMOO INOUE^{†2}

1. はじめに

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、音声は同一方向から聞こえてくるような設計であることが多い。しかしながら、対面会話では話者のいる位置に応じてその音声が聞こえるのであり、遠隔会話においても話者ごとに音声が聞こえる位置を分離させた方がよい可能性がある。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検討する。ただし、本研究では遠隔地にいる参加者の音声を、複数のスピーカを用いて出力するのではなく、一般的なステレオスピーカと音声処理技術によって定位する手法を想定する。本稿ではまず、遠隔複数人会話における音像定位の基礎的研究として、複数の音声を個々に定位することが、音声の聞き取りやすさに効果的かどうか実験的に検討する。

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 音声会話支援手法

2.1.1 音声速度

言葉を「ゆっくり」、「はっきり」話すことが、補聴手段として有効であることは知られており[1]、音声速度に関する研究は数多くある。ラジオ放送のニュース音声に対し、話速変換を行うことで音声再生速度を下げる手法[2]や、超早口音声に対応するため、単語親密度を用いて会話速度調整を行う手法[3]が提案されている。単語親密度(word

familiarity)とは、ある単語がどの程度なじみがあると感じられるかを表した指標であり、単語親密度ごとの単語理解度リストが配布されている[4]。単語理解度は、多くの音声通信システムや補聴器、オーディオリアムの音声伝達性能の評価に用いられている。

2.1.2 音声明瞭度

ノイズレベルを下げる、あるいは音量を上げることで、音声明瞭度を高め、会話を支援することができる。近くの雑音を消して、遠くの声を強調するという手法も提案されている[5]。また、書類をめくる音や机をたたく音のような卓上で生じる突発性の雑音を高精度に除去することが可能な遠隔会議システムも提案されている[6]。

2.1.3 音像定位

音像定位をどのように実現するかは古くから検討されている。特に、音声の左右耳における強度差、音声は左右耳に到達する時間差が重要な要素であり、多数の研究がある[7]。

音声・音響の臨場感を作り出すためには音像定位が重要である[8]。例えば、人間の頭部をモデルに作成したダミーヘッドの耳の部分にマイクを埋め込んだダミーヘッドマイクを用いるバイノーラル録音では、3Dの音響効果が録音できる。バイノーラル録音再生方式はモノラル方式や2チャンネルステレオ方式と比べ単語理解度が高まることが明らかにされている[9]。

反対に3D音響を再現するためには、ダミーヘッドの特徴を数式化した頭部伝達関数(Head-Related Transfer Function, HRTF)を用いるが、HRTFは頭や耳介の形状の影響を受けるため、個人差が大きいことが知られている[10]。

一方、視覚と聴覚のクロスモーダル知覚を用いることで、聴覚刺激のみを与えたときと比較して、音像の位置をより強く認識できることが明らかにされている[11]。

^{†1} 筑波大学大学院図書館情報メディア研究科
Department of Library, Information and Media Science, 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

2.2 遠隔音声会話支援

遠隔音声会話支援を行う研究としては、参加者ごとに AV 機器を用意し、現実の会話場面に似せた環境を構築する Hydra がある[12]。対面会話と同様、参加者ごとに音声が出力されるスピーカが異なるため、複数の参加者が同時に発声しても誰が発言したか把握しやすい。しかしながら、参加者の人数分のマイクとスピーカが必要となり、それらの設備を各参加者が用意することは現実的ではない。

また、立食パーティのような、会話空間に複数の話題が混在する場合を想定した研究[13]があり、参加している会話の聞き取りやすさと参加していない会話の概要把握に有効な手法について述べられている。しかしながら、複数の会話をどのように音像定位するべきかについては十分に明らかでない。

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聞き取りづらくなると考えられる。本実験では、発言が衝突する際に、音声を個々に定位することでそれぞれの発言内容が聞き取り易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聞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 人である。被験者は全員、日本語能力試験 N2 以上であり、在日平均年数は約 1 年である。平均年齢は 23.8 歳である。

3.3 実験概要

複数人が会話する状況を想定し、2 人もしくは 3 人が同時に発声した際、話した単語の聴取状況を検証する実験を行った。本実験で使用するものは、実験用に構築した、被験者自身の手で任意に音像定位が可能なソフトウェアである。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[14]を使用し、両耳間の音量バランス、時間差が調整可能となる機能を実装した。

図 1 に本ソフトウェアの音像定位用パラメータ設定時の画面を示す。中央に音像定位するパラメータが設定されている。左側のスライダは音量バランス調整用、右側のスライダは時間差調整用として機能する。中央に音像定位する場合、音量バランスは 5 : 5、時間差は 0ms に設定される。左のスライダを一番上まで引き上げると左右の音量バランスが 10 : 0 (左耳用音量割合 : 右耳用音量割合) となり、一番下まで引き下げると 0 : 10 (左耳用音量割合 : 右耳用音量割合) となる。設定可能な値は 0~10 までの整数である。左耳用音量と右耳用音量の割合の合計値は常に 10 となる。つまり、左耳用音量を 8 とした場合、右耳用音量は 2 とな

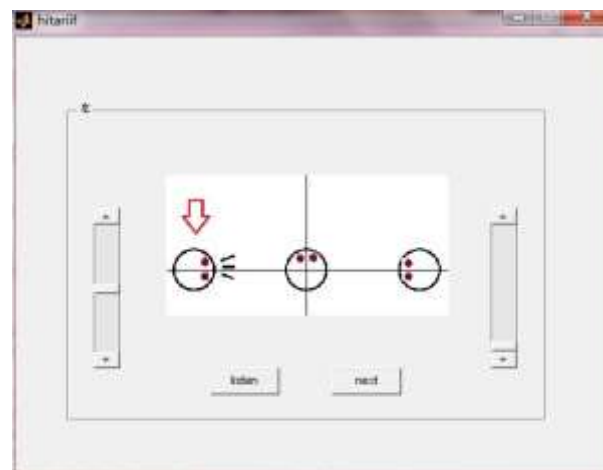


図 1 音像定位用パラメータ設定時の画面

る。

右のスライダを一番上まで引き上げると、時間差が 5ms となり、一番下まで引き下げると時間差が 0ms となる。ここでいう時間差とは、音声は左右耳に到達する時間差のことを指す。時間差の代表的なモデルとして、Kuhn[15]の計算式がある。ここでは、真横に音像定位する場合の時間差は 0.53ms と 0.73ms とされている一方、Woodworth と Schlosberg[16]の近似式での時間差は 0.68ms と示されている。さらに、3D 化の効果を高めるために時間差を 2.73ms に設定した音声システム[17]もある。本ソフトウェアでは各被験者が感覚に合わせて設定できるよう、0ms から 5ms と設定可能な時間の幅をやや大きめに設けた。

使用する日本語の単語は、単語理解度リスト[4]にある単語を採用した。日本語の音声ファイルは、日本語 NHK アクセント辞典[18]の付録を採用した。アナウンサーが読み上げたものが収録されている。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[19]に記載されている四字熟語を採用した。音声ファイルには附録の録音ファイルを用いた。日本語の音声ファイルが「たなばた」など 4 つの発音で構成されている単語であることに對し、中国語の音声ファイルも 4 つの発音で構成されている単語である。

多人数が同時に発話した場合、およそ 2 人程度の発話内容までしか聞き取ることができないことが明らかにされているため[20]、本実験では同時に再生する単語数を 2 語と 3 語の 2 パターンを用意した。さらに音像定位の効果を測るため、2 単語を用いる場合は、中央で 2 単語を同時に再生した場合と左右にそれぞれ音像定位して再生した場合の 2 パターンを用意した。3 単語を用いる場合は、中央で 3 単語を同時に再生した場合と左右中央それぞれで音像定位して再生した場合の 2 パターンを用意した。

3.4 実験手順

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、ヘッドホ

ンを用意した。ヘッドホンにはオーディオテクニカ社のATH-WS55を使用した。実験は騒音のない静かな環境で実施した。被験者には機器の操作方法について説明し、事前準備、実験の順に取り組みさせた。

はじめに、事前準備である左右方向の音像定位パラメータを設定させ、日本語を用いた2単語実験、3単語実験、中国語を用いた2単語実験、3単語実験の順で実施した。

被験者は、図1の画面中の図に従い、まず左方向に音像定位する場合のパラメータについて、左右スライドを動かすことで設定する。「listen」ボタンで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、何度でも確認し直し直することができる。被験者が左方向から「最も感度がいい」と判断した場合に、「next」ボタンを押下し、右方向の音像定位設定に移行する。

左右方向の音像定位の設定が完了したら2単語実験および3単語実験に移行する。図2に2単語実験時の画面を例示する。実験開始後は「listen」ボタンを押下したタイミングで音声ファイルが再生されるが、一度押下すると「listen」ボタンは消え、再度再生することはできない。被験者は聞こえた単語についてローマ字で入力フォームに入力する。次の単語を聴く準備ができたなら「next」ボタンを押すことで次の音声ファイル再生に移行することができる。再生した音声ファイルの識別番号および被験者の入力した文字列はExcelファイルに出力され、保存される。実験時の被験者の様子を図3に示す。

3.5 2単語実験

2単語実験では、2つの単語を同時に再生した場合の音源数の違いによる影響について比較する。比較する条件を表1に示す。本実験では2単語2音源条件および2単語1音源条件の2条件を用意した。2単語2音源条件では、音声ファイル2つを左右で同時に再生し、2単語1音源条件では、音声ファイル2つを中央で同時に再生する。左右の音像定位は前節で述べた事前準備の結果得られた、被験者



図2 2単語実験時の画面

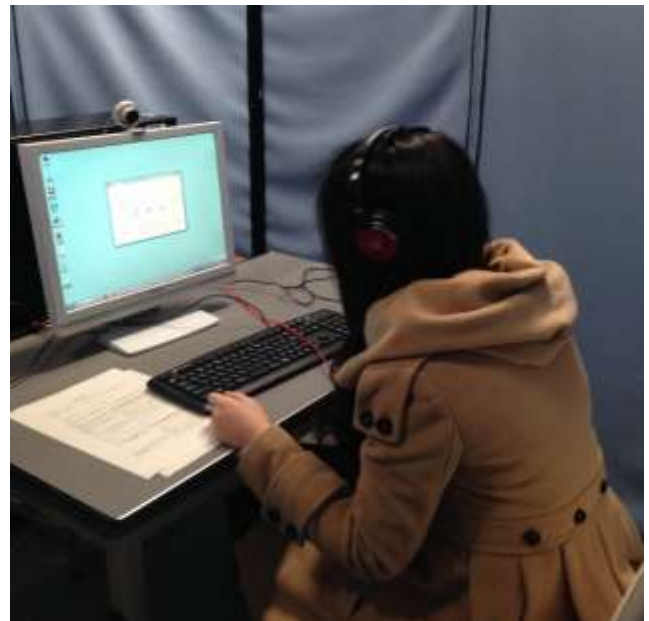


図3 実験時の被験者の様子

表1 2単語実験、3単語実験の条件パターン

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： 中央に定位

本人が「最も感度がいい」と判断した音量バランス、時間差のパラメータを用いる。

実験は各条件5回ずつ連続で行い、参加する両条件の順番はカウンターバランスをとった。被験者に提示される音声ファイルは、ソフトウェアがランダムに出力する音声ファイル2語である。条件が切り替わる際には、画面に「attention!!」の文字列を表示した。音声ファイルは一度きり再生することができ、被験者は聞き取れた単語のみ入力フォームに入力した。

以上の実験について、再生する音声ファイルに日本語を用いた場合、中国語を用いた場合の2パターンについて、日本語、中国語の順に実施した。

3.6 3単語実験

3単語実験では、3つの単語を同時に再生した場合の音源数の違いによる影響について比較する。比較する条件を表1に示す。本実験では3単語3音源条件および3単語1音源条件の2条件を用意した。3単語3音源条件では、音声ファイル3つを左右中央で同時に再生し、2単語1音源条件では、音声ファイル2つを中央で同時に再生する。左右の音像定位は前々項で述べた事前準備の結果得られた、被験者本人が「最も感度がいい」と判断した音量バランス、時間差のパラメータを用いる。

実験は各条件 5 回ずつ連続で行い、参加する両条件の順番はカウンターバランスをとった。

3 単語実験で被験者に提示される音声ファイルは、2 単語で使用されていない単語からランダムに選択された 3 語である。2 単語実験と同様、条件が変わる際には、画面に「attention!!」の文字列を表示した。音声ファイルは一度きり再生することができ、被験者は聞き取れた単語のみ入力フォームに入力した。

以上の実験について、再生する音声ファイルに日本語を用いた場合、中国語を用いた場合の 2 パターンについて、日本語、中国語の順に実施した。

3.7 質問紙調査方法

2 単語実験と 3 単語実験の両方を実施した後、被験者に質問紙調査を行った。質問紙調査では、実験準備の音像定位の操作性、使用した単語の難易度、再生速度、音源数による聞こえ方の違いに関する印象、音像定位の差による聞こえやすさへの影響について調査した。質問項目は計 6 項目で構成し、回答は「1:全く当てはまらない」から「7:よく当てはまる」までの 7 段階のリッカート尺度で行った。質問項目は、No.1「左右の音像定位の設定が難しかった」、No.2「単語が難しかった」、No.3「再生速度が速かった」、No.4「合図の画面前後で聞こえ方が違った」、No.5「2 語の場合、合図を表示する前の音声の方が聞きやすい」、No.6「3 語の場合、合図を表示する前の音声の方が聞きやすい」である。カウンターバランスをとったため、質問項目中の「合図を表示する前」の条件が意味するのは、複数音源の場合が 10 名、単数音源の場合が 10 名となる。

4. 実験結果

事前準備の左右方向の音像定位については被験者 1 人に対しそれぞれ 3 回実施し、20 人の被験者で 60 サンプルを得た。その結果、左方向の音像定位時のパラメータの平均値は、音量バランスが 8.52 : 1.48 (左耳用音量 : 右耳用音量)、時間差 (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 0.85ms であった。同様に、右方向の音像定位時のパラメータの平均値は、音量バランスが 8.33 : 1.67 (右耳用音量 : 左耳用音量)、時間差 (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 1.99ms であった。

4.1 2 単語実験

2 単語実験から得られた正解数を比較した。正解数とは、被験者が提示された音声ファイルを聞き取り、正しく入力できた単語の数を指す。例えば、2 単語同時に再生された時に 2 単語とも正しく入力できた場合の正解数は 2、1 単語のみ入力できた場合の正解数は 1、2 単語とも入力できなかった場合の正解数は 0 となる。

図 4 に、2 単語 2 音源条件、2 単語 1 音源条件について、日本語、中国語それぞれを用いた場合の平均正解数を示す。

エラーバーは標準偏差の幅を正方向に示す。

日本語使用時の平均正解数は、2 単語 2 音源条件では 1.08、2 単語 1 音源条件では 0.53 であった。中国語使用時の平均正解数は、2 単語 2 音源条件では 1.55、2 単語 1 音源条件では 0.96 であった。

それぞれのデータに対して、Shapiro-Wilk の正規性検定を行ったところ正規分布しているとはいえないデータであったため (全ての条件 : $p=0$)、2 音源と 1 音源の結果の比較に Wilcoxon の符号付順位検定を用いた。この結果、2 音源条件の方が 1 音源条件よりも正解数が多いことが分かった (日本語使用時 : $p=6.39E-07$, 中国語使用時 : $p=1.31E-07$)。

4.2 3 単語実験

3 単語実験から得られた正解数を比較した。図 5 に、3 単語 3 音源条件、3 単語 1 音源条件について、日本語、中国語それぞれを用いた場合の平均正解数を示す。エラーバーは標準偏差の幅を正方向に示す。

日本語使用時の平均正解数は、3 単語 3 音源条件では 0.57、3 単語 1 音源条件では 0.27 であった。中国語使用時の平均正解数は、3 単語 3 音源条件では 0.99、3 単語 1 音源条件では 0.55 であった。

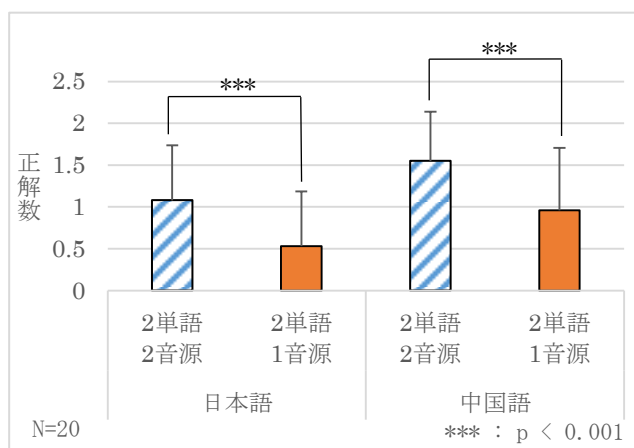


図 4 2 単語実験の正解数

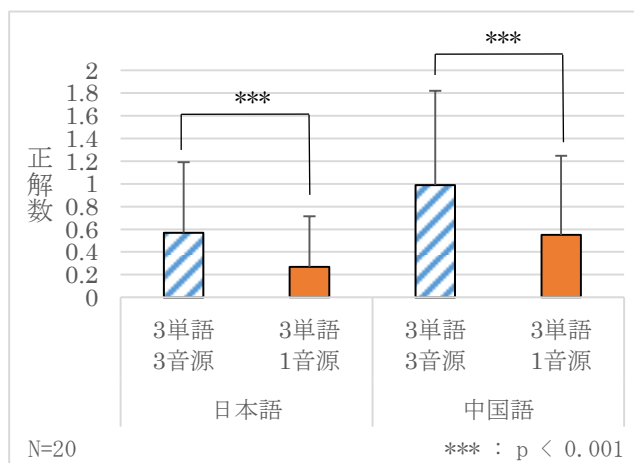


図 5 3 単語実験の正解数

表2 質問項目1~3の結果

No.	質問項目	平均得点
1	左右の音像定位の設定が難しかった	3.75
2	単語が難しかった	4.45
3	再生速度が速かった	3.95
4	合図の画面前後で聞こえ方が違った	5.55

(N=20)

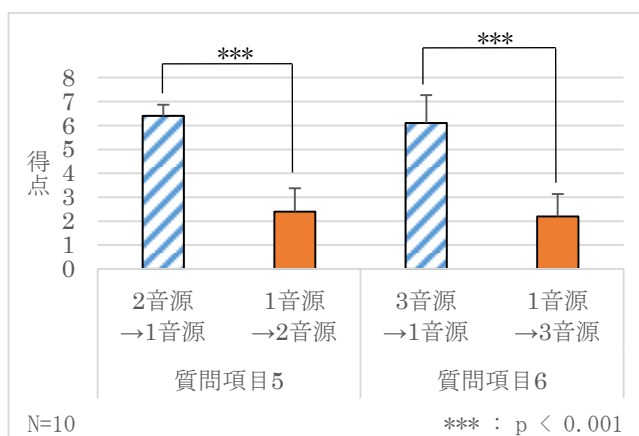


図6 質問項目4, 5の結果

それぞれのデータに対して、Shapiro-Wilkの正規性検定を行ったところ正規分布しているとはいえないデータであったため(全ての条件: $p=0$)、3音源と1音源の結果の比較にWilcoxonの符号付順位検定を用いた。この結果、3音源条件の方が1音源条件よりも正解数が多いことが分かった(日本語使用時: $p=0.0002$, 中国語使用時: $p=2.55E-05$)。

4.3 質問紙調査結果

質問紙の回答の「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」に対し、それぞれ1点から7点までを対応付け、質問項目ごとに平均点を求めた。表2に質問項目1~4、図6に質問項目5, 6の結果を示す。

質問項目1~5については、それぞれ平均得点が3.75, 4.45, 3.95, 5.55であった。質問項目5, 6はそれぞれ「2語の場合、合図を表示する前の音声の方が聞きやすい」、「3語の場合、合図を表示する前の音声の方が聞きやすい」であったが、被験者によって「合図を表示する前の音声」が異なるため、図6のように示す。図6には、質問項目ごとの、複数音源、単数音源の順に実施した被験者と単数音源、複数音源の順に実施した被験者の評価結果の平均値をそれぞれ示す。質問項目5については、2音源、1音源の順に実施した被験者の回答の平均得点は6.4、1音源、2音源の順に実施した被験者の回答の平均得点は2.4であった。これらは対応のある順序尺度のデータのため、得点の比較に

Wilcoxonの符号付順位検定を用いたところ、両者間で有意差が認められた($p=0.003$)。質問項目6については、3音源、1音源の順に実施した被験者の回答の平均得点は6.1、1音源、3音源の順に実施した被験者の回答の平均得点は2.2であった。これらは対応のある順序尺度のデータのため、得点の比較にWilcoxonの符号付順位検定を用いたところ、両者間で有意差が認められた($p=0.004$)。

4.4 検討

図4, 図5の結果から、日本語を使用した場合、中国語を使用した場合いずれの場合においても、2単語実験、3単語実験の両者において複数音源条件の方が単数音源条件に比べて正解数が有意に多くなることが分かった。

また、図6の結果から、被験者の意識としては、複数音源の方が単数音源よりも聞きやすいと感じていることが分かった。この結果は図4, 図5の正解数による比較結果を支持するものであり、主観的にも客観的にも複数音源条件の方が単数音源条件よりも単語の聴取がしやすいことが明らかとなった。

また、表2の結果から、左右の音像定位の設定難易度については平均得点が3.75点であった。「どちらでもない」という回答の4点に近い数値であり、本ソフトウェアの操作性が難しいとは言えない。本実験で使用した単語については4.45点再生速度については3.95点という結果であった。いずれも4点前後が平均値の値となるため、音像定位の方法や使用した単語は、難しすぎず簡単すぎず、再生速度は早すぎず遅すぎず適当であったと考える。一方、質問項目「合図の画面前後で聞こえ方が違った」については平均得点が5.55点であり、1人が「全く当てはまらない」、2人が「あまり当てはまらない」と回答したが、その他の被験者は「やや当てはまる」、「当てはまる」、「よく当てはまる」のいずれかを回答した。一部の人のにとっては音像定位の効果が薄かったものと考えられるが、概ね全員が音像定位による聴取の違いを感じていたことが分かる。

4.5 提案システム

これまでの音像定位を用いた研究には、遠隔地にいる会話相手の位置関係を含めた高臨場感を目指す研究[8]もあるが、本研究では遠隔地間での会話のわかりやすさ向上を目指す。中貝らは、バイノーラル録音再生方式の有効性を検証する過程で、話者の発声した単語を選択的に聴取する実験を行った[9]。これに対し、本研究では同時に再生された音声全ての聞きやすさについて検証した。遠隔にいる話者の発言が衝突した際に発言内容がわかりやすいシステムを目指す。

本稿では複数の音声の定位を分離することが、音声の聞きやすさに影響するかどうかを検討するための実験を実施した。その結果、定位を分離させて複数位置から聞こえるように同時再生した場合は、定位を分離させずに同時再生した場合に比べ、聞き取れる単語数が増大することが分か

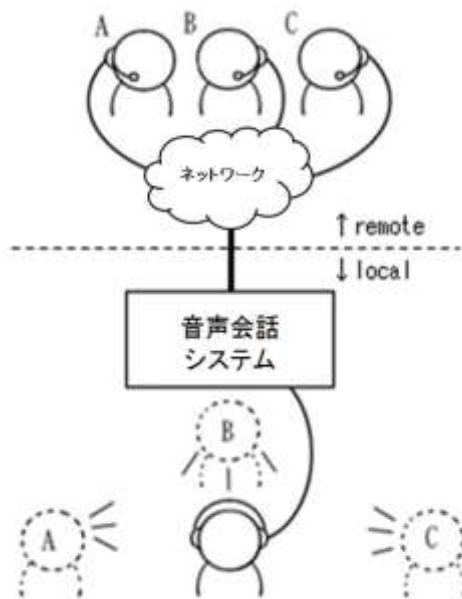


図7 提案システムイメージ

った。このことから、本研究で提案するシステムでは、遠隔地にいる複数参加者の音声の定位を分離することで会話のわかりやすさを向上させるシステム設計を提案する。図7にシステムイメージを示す。遠隔地にいる会話参加者の音声をマイクで取得し、音声会話システムを通してローカルの参加者に伝達する。このとき、ローカルの参加者はヘッドホンをかけて会話に参加するが、遠隔会話参加者の複数の音声に対し、複数の音像定位を行うことで、異なる場所からそれぞれの音声が聞こえるようにする。図7ではAさんの音声は左側から、Bさんの音声は正面から、Cさんの音声は右側から聞こえるようにする。このように音声会話システムが音声の定位を分離させることにより、ローカルの参加者にとって会話の相手の声が識別しやすいシステムとなる。

5. まとめ

本研究では、複数相手の音声の定位を分離することが、音声の聞き取り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聞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、聞き取れる量が増大することが分かった。また、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「聞きやすさ」に大きく影響し、遠隔地にいる複数人の会話相手の音声に対し、より聞きやすいという印象を与えることが示唆された。これらの結果を考慮し、本稿では、遠隔の複数相手との会話をわかりやすくするために、相手ごとに音声の聞こえる位置を定位させる音声会話システムを提案した。

今後は、頭部伝達関数を用い、前方の円形定位を実現し、

受聴者が聞き分けることの可能な角度について追調査していく。また、ヘッドホンの左右の音量を別々に設定できるように機能拡張を行うなど、人間にとってよりやさしいシステムの開発を目指す。

参考文献

- 1) 伊福部達: 高齢難聴者のための音声情報変換方式, 電子情報通信学会誌 Vol.84, No.5, pp.325-328, (2001).
- 2) 今井篤, 都木徹, 黒住幸一, 小山隆二, 島津敏雄: 話速変換技術を利用したインターネットでのラジオニュースサービス, 映像情報メディア学会誌: 映像情報メディア Vol.59, No.2, pp.265-270, (2005).
- 3) 西本卓也, 渡辺隆行: 単語親密度を統制した超早口音声の聴取に対する慣れの検討, 電子情報通信学会論文誌. D, 情報・システム Vol.J94-D, No.1, pp.209-220, (2011).
- 4) 東北大学 電気通信研究所 先端音情報システム研究所: <http://www.ais.riec.tohoku.ac.jp/index-j.html>, (2014/2/18 参照).
- 5) Boll, S. F.: Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process. Vol.ASSP-27, pp.113-120, (1979).
- 6) 戸上真人, 川口洋平, 菅沼俊輔, 橋本孝幸, 小窪浩明, 大淵康成: 垂直配置マイクロホンアレイを利用した卓上突発音除去機能を備える遠隔会議システム, 電子情報通信学会論文誌. D, 情報・システム Vol.J93-D, No.10, pp.2069-2084, (2010).
- 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) 榎並和雅, 勝本道哲, 西村竜一: 超臨場感(Ultra-Reality)の現状と将来像, 日本音響学会誌, Vol.64, No.5, pp.322-327, (2008).
- 9) 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理解度に及ぼす影響, 電子情報通信学会論文誌.
- 10) 山本健一郎, 望月研二, 苗村健, 齊藤隆弘, 原島博: 空間共有通信における個人差を考慮した音像の提示, 電子情報通信学会技術研究報告. EA, 応用音響, Vol.102, No.398, pp.69-74, (2002).
- 11) 王夢, 小川剛史: 視覚と聴覚のクロスモーダル知覚を用いた音像定位システムに関する基礎検討, 情報処理学会研究報告, Vol.2013-DCC-4, No.7, (2013).
- 12) Sellen, A.: Speech patterns in video-mediated conversations, Proceedings ACM CHI'92 Conference on Human Factors in Computing Systems, pp.49-59, (1992).
- 13) 大橋純広, 瀧崇宏, 河合栄治, 藤川和利, 砂原秀樹: 視覚情報により強化された3Dサウンド場による共有型多人数音声チャットシステムの設計と実装, 情報処理学会研究報告. CSEC, コンピュータセキュリティ Vol.2006, No.26, pp.227-232, (2006).
- 14) MathWorks: <http://www.mathworks.co.jp/products/matlab/>, (2014/2/18 参照)
- 15) Kuhn, G. F.: Physical Acoustics and Measurements Pertaining to Directional Hearing, in Yost, W. A. and Gourevitch, G. (eds), Directional Hearing, Springer-Verlag, pp. 3-25, (1987).
- 16) Woodworth, R. S. and Schlosberg, R.: Experimental Psychology, Holt, Rinehard and Winston, pp. 349-361, (1954).
- 17) Hardman, V. and Iken, M.: Enhanced Reality Audio in Interactive Networked Environments, Framework for Interactive Virtual Environments (FIVE) Conference, December (1996).
- 18) 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).
- 19) 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).
- 20) 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).