

# 音源定位結果と音声認識結果を HMD に統合呈示する 聴覚障害者向け音環境理解支援システム

徳田 浩一

駒谷 和範

尾形 哲也

奥乃 博

京都大学大学院 情報学研究科 知能情報学専攻

## 1. はじめに

会議や授業に参加する聴覚障害者のための情報保障手段に要約筆記がある。最近では PC 上で支援者がタイピングを行うものや、音声認識を用いた文字入力を試みられているが、作成された字幕は、利用者の近くのディスプレイか前方スクリーンに呈示されることが多い。このような固定設置型の字幕呈示手法では、字幕と字幕以外の注視対象(資料、黒板、テレビ等)が離れているので、視線移動が多くなるという問題が生ずる。

また、聴覚障害者支援で呈示する情報は、音声認識結果である言語情報だけでは不十分である。聴覚障害者とその場の音文脈を周囲の人と共有するためには、音環境中の非言語情報の呈示も重要である。例えば、音源到来方向の呈示は、音源位置の視認を促すので、字幕に反映されない音環境情報の取得が容易となる。

我々は、光学透過型ヘッドマウントディスプレイ (HMD) に音源方向と字幕を呈示する、聴覚障害者の目で見える音環境理解支援システムを開発している。本稿では、開発システムの概要と音源到来方向呈示の有効性に対する実験について報告する。

## 2. 聴覚的知覚レベルに基づいた設計

聴覚障害者の聴能訓練で用いられる聴覚的知覚のレベルには、1. 聴覚的検知能、2. 聴覚的弁別能、3. 聴覚的識別能、4. 聴覚的理解能の 4 つの段階がある(表 1) [1]。これら各レベルで開発すべき機能を以下にまとめる。

### 1. 聴覚的検知能: 「音の有無の検知」

#### 1) 実時間での音源情報の呈示

音文脈のライブ感の共有や音源の視認のためには、音の存在を実時間で把握できることが必要である。本システムでは字幕呈示の前に音源存在情報を呈示し、音声認識結果の修正に伴う遅延を回避する。

#### 2) 利用者の視線方向に依存しない機器への情報の呈示

固定設置の呈示機器では、利用者の視線が機器から外れている場合に情報の検知漏れを招くおそれがある。このため、利用者の視線方向に非依存な呈示機器を適用する。

### 2. 聴覚的弁別能: 「音の異同の判断」

#### 1) 音声尤度の呈示

存在の確認できた音源が、あるカテゴリーに属するか否かの情報を提供する。本システムでは、検知できた音源が音声か非音声かを利用者が知ることができるよう、GMM に基づく音声/非音声識別結果の、周囲分離音に対する音声尤度情報を音源アイコン表示に反映させる。

### 3. 聴覚的識別能: 「複数のカテゴリーへの照合と同定」

#### 1) 音声認識機能の組み込み

音声認識結果を取得し、字幕の原型を作成する。

Hearing-Impaired-Supporting System in Understanding Auditory Scenes by Presenting Sound Source Localization and Speech Recognition Results in Integrated Manner on HMD: Koichi Tokuda, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

表 1: 聴覚的知覚のレベルと実装する機能

知覚のレベル	知覚の内容	実装する機能
聴覚的検知能	音の on/off	音源存在情報呈示
聴覚的弁別能	音の異同	音声尤度呈示
聴覚的識別能	既存カテゴリーへの同定	音声認識・音源定位結果呈示
聴覚的理解能	意味内容の了解	認識結果修正, キーワード強調表示, 信頼度表示

#### 2) 音源定位結果の呈示

音文脈の共有のためには、利用者が音源発生位置を把握し、視認を行うことが有効である。そのため、周囲分離音の音源到来方向を同定し、呈示する。

### 4. 聴覚的理解能: 「弁別や識別に基づく意味内容の了解」

#### 1) 認識結果の修正

音声認識結果には認識誤りが含まれるため、支援者が人手で修正・整形を行う。

#### 2) キーワードの強調表示

話し言葉が字幕化された文章は読みにくく、重要な単語を見逃すおそれがあるので、登録済みキーワードを強調表示し、利用者の字幕の読解を支援する。

#### 3) 低信頼度単語の灰色表示

修正者を配置できない場合は、認識結果の単語信頼度の低い単語が目立たないように、灰色表示する。

### 3. 音環境理解支援のための呈示システム

前章で述べた設計に基づいて、音声認識を用いて作成した字幕と周囲音の音源定位結果を、視線移動の少ない装置に呈示するシステムを実装した。図 1 に示したように、音声認識部、音源定位部、認識結果修正部、情報出力部から構成され、それらは LAN で相互に接続される。音声認識部では、話者マイクより音声を取得し、大語彙連続音声認識システム Julius [2] を用いて音声認識を行う。認識結果修正部では、PC 要約筆記ソフト IPtalk [3] を用いて、認識結果を人手により修正する。音源定位部では、マイクロホンアレイを用いて周囲環境音を収集し、音源定位、音源分離、音源同定を行う。音源定位には Multiple Signal Classification (MUSIC) を、音源分離には Geometric Source Separation (GSS) を用いる [4]。

情報出力部では、認識結果修正部から送信された字幕と、音源定位部から送信された音源定位結果を受信し、呈示用画面に逐次反映させる(識別能の支援)。図 2 に情報出力部の GUI のスクリーンショットを示す。字幕は画面下半分の字幕表示領域に表示される。事前に登録したキーワードは強調表示される(理解能の支援)。修正者を配置できない場合の起動モードでは、キーワード強調表示機能は無効とし、設定値未満の単語信頼度の単語を灰色表示する(理解能の支援)。

音源定位結果は画面上半分の 3D 画面中のマイクロホンアレイに向かってビーム表示され、GMM に基づく音声/非音声識別結果の音声尤度に応じて音声アイコンが表示される(弁別能の支援)。話者マイクを装着した話者が話している間は、3D 空間の事前に登録した位置に話

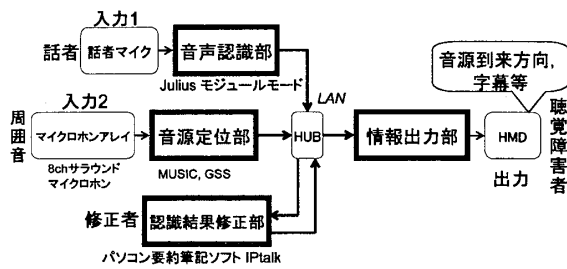


図 1: システムブロック図

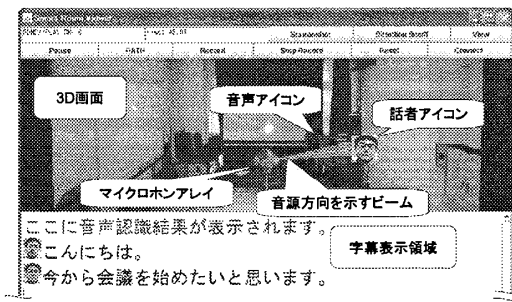


図 2: 情報出力部の GUI

者アイコンが表示される (検知能の支援). 呈示機器に透過型 HMD を用いて両方の視野を重ね、システム情報と他の注視対象の位置が離れるのを防ぐ (検知能の支援).

#### 4. 実験: 音源到来方向呈示機能の有効性

本システムで開発した音源到来方向呈示機能の有効性について評価する. なお, HMD の有効性については, [5] で報告した通り, 字幕とそれ以外の注視対象を重ねることで, 視線移動負荷軽減の可能性が示唆された.

##### 4.1 実験内容

音源到来方向呈示機能により利用者が素早く音源位置を視認できるか評価する. 音声イベント情報を (1) 音源名, あるいは (2) 音源到来方向をビームで呈示し, 視認できた内容を被験者に記録させた. なお, (1) は従来の支援者による音声イベント情報の字幕追加挿入に相当する. (1) (2) の各条件で呈示される情報に意識を向けさせるため, 次の (a) (b) の方針で実験を構成した.

(a) 音声イベント発生タイミングの視覚による判断を強制: 実際には音は発生させず, 実験システム内部でディスプレイの呈示内容とタイミングを制御することで, 仮想的に音声イベントを発生させる.

(b) 音源発生位置の素早い視認を強制: 仮想音声イベント発生直後に視認させる情報は, 音源位置に設置したディスプレイに短時間に更新表示される 4 数字とする. これにより, 視認に要した時間の短さを正答数の多さに反映させる.

##### 4.2 方法

被験者は視覚に障害のない学生 18 名である. ディスプレイを, 被験者の座る周囲 1.5m の位置に 4 か所 (右前, 右後, 左前, 左後) 被験者の向きに設置し, それぞれに音源名 (警報機, ドア, 電話, やかん) を対応させた. タスク教示の際, 被験者に 4 音源の名称と場所を確認させ, HMD を装着させた. 仮想音声イベント発生時, 被験者の HMD 上に以下のどちらかの条件で情報呈示した.

- (1) 字幕表示領域に音源名 (字幕)
- (2) 3D 画面上に音源到来方向 (ビーム)

表 2: 呈示情報の種類と視認できた数字の正答率 (%)

	前方向	後方向
(1) 音源名呈示 (従来手法)	90.3	56.5
(2) 音源到来方向呈示 (本手法)	94.2	73.6

両条件とも被験者 HMD への情報呈示の 30msec 後から, 当該音源に対応するディスプレイにランダムな 1 桁数字を 500msec 毎に 4 数字切り替えて表示した. 被験者には視認できた数字を表示順に記録させ, 10 秒程度経過してから次の試行へ移った. 2 条件と 4 方向の呈示順はランダムとし, 計 24 試行実施した.

#### 4.3 結果

表 2 に, 各条件下での数字書き取り正答率を示す. 音源の方向を前と後に分けて正答率を算出した. 両方向とも, 音源名を呈示されたときより, 音源到来方向を呈示されたときの正答率の方が高かった. 後方向については, 1%水準で (1) (2) 条件間の有意差が見られた.

#### 4.4 考察

音声イベントの発生から視認までに時間がかかると, 音声イベントを見逃すことにつながる. 表 2 より, 音源到来方向をビーム呈示することで, 素早い視認を可能にすることが確認できた. 特に後方向は音源名と位置が結びつきにくいため, 音源方向を伝達することが素早い視認に寄与したと考えられる.

本システムの想定する利用状況では, 利用者の注視対象は, 話者の顔, 手元の資料, 音源到来方向等, 動的に変化する. したがって, それら注視対象とシステム呈示情報間の視線移動負荷の少ない HMD の有効性が活用できる. 特に, 利用者が音源方向を視認する際にどの方向を向いていても, システム呈示情報を常に確認できることの有効性は大きいと考えられる.

音源到来方向の呈示に関しては, 不要な視認を防ぐための支援として, 方向や音源種類によるフィルタリング機能を実現することが考えられる.

#### 5. おわりに

本研究では, 聴覚障害者が周囲の人と音文脈を共有して積極的にその場の活動に参加するためには, 言語情報である字幕に加えて非言語情報の呈示も重要であるとの立場から, 音源到来方向を呈示する音環境理解支援システムを設計し, 実装した. 被験者実験の結果, 音源到来方向の呈示で, 音声イベント発生時の音源方向の素早い視認が可能なることを確認した. また HMD へのシステム情報呈示で, 他の注視対象との視線移動の負荷が軽減できるようになった. 今後, プロソディ・単語信頼度の色による呈示の効果についても評価する.

謝辞 本研究の一部は, 科研費, GCOE の支援を受けた.

#### 参考文献

- [1] 大沼: 教師と親のための補聴器活用ガイド, コレール社, 1997.
- [2] Julius development team: <http://julius.sourceforge.jp/>
- [3] 栗田: <http://iptalk.hp.infoseek.co.jp/>
- [4] Yamamoto et al.: Design and Implementation of A Robot Audition System for Automatic Speech Recognition of Simultaneous Speech, Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU-2007), pp.111-116, 2007.
- [5] 徳田他: 聴覚障害児の授業支援のための HMD による音声認識結果呈示システムの設計, 第 69 回情報処全大, 3ZB-4, 2007.