6ZB-2

音を視覚化する録音再生システム

吉田 雅敏 † 海尻 聡 ‡ 山本 俊一 ‡ 中臺 一博 * 駒谷 和範 ‡ 尾形 哲也 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学大学院 情報学研究科 知能情報学専攻

* (株) ホンダ・リサーチ・インスティチュート・ジャパン

1. はじめに

ディジタル技術の進展と共に，人間生活の全場面の映像や音響をディジタル化し，アーカイブをしようというライフログの研究が活発化している [1]．会議や講義のアーカイブ化，建物監視のアーカイブ化，などデータは爆発的に増えつつある．アーカイブデータが映像の場合には早送りあるいはサムネイル自動付与などにより，所望の映像部分を探索し，再生することは，比較的容易である．一方，アーカイブデータが音響信号，特に，混合音である場合には，早送りによって所望の信号部分を探索し，再生することは，音情報にはサムネイルなどの時間的一覧性機能が未確立であり，映像ほどには容易でない．

実環境では，音源は私達の周囲に遍く広く分布している可能性があるので，特定の方向の音だけでなく，全方位の音響信号を録音する必要がある．本稿では音源の存在情報を利用して，3D 次元での混合音に対する時間的一覧性機能を有する録音再生システムの設計と実装について報告する．

2. 音情報視覚化の設計方針

Ben Shneiderman [2] は情報を分かりやすく提示するための情報視覚化の設計方針を『*Over first, zoom and filter, then details on demand.*』(以下，*O-ZF-D* と略す) という 3 レベルで表現している．つまり，ユーザーに全体の概略を示しつつ，さらに重要な部分情報を提示する．ユーザーはこれらを手がかりに必要な情報がある箇所を探し出して求める情報を得るというものである．

2.1 *O-ZF-D* に基づいた音情報視覚化の機能

音情報の視覚化の *O-ZF-D* の各レベルを次に示す:

1) *Over first* (*O* レベル) ⇒ 音情報の全体像を提示.
   (a) 音源の存在する方向を時系列に従って一覧表示する.
2) *Zoom and filter* (*ZF* レベル) ⇒ 音源の存在を提示.
   (a) 音源の存在する方向を提示する.
   (b) マイクの音そのものを再生する.
3) *Details on demand* (*D* レベル) ⇒ 音源の情報を提示.
   (a) 音源方向を選択することでその方向にある音源の音を再生する.

図 1 を用いて各レベルの機能を説明する．図は会議の議事録として録音を行っている状況を想定している．時々音楽も演奏される．

図 1 の左部分は，時間軸に沿って会議参加者それぞれが発話した時間帯を示したものである．この様な形で，音源が存在するおおよその時間帯を提示することで音情報全体の概略が解る (*O* レベルの機能)．これにより聞きたい音がある時間帯を限定することが出来る．

Recording and Playback System with Auditory Scene Visualization: *Masatoshi Yoshida, Satoshi Kaijiri, Shunichi Yamamoto (Kyoto Univ.), Nakadai Kazuhiro (HRI-JP), Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

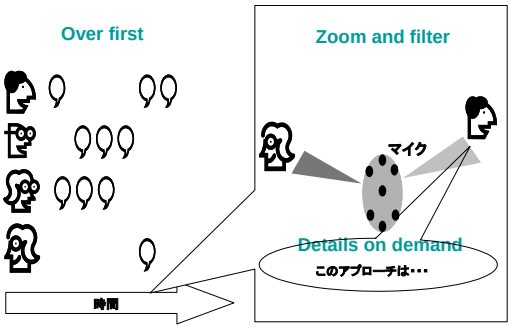

図 1: 会議録アーカイブ再生での *O-ZF-D* 設計

図 1 の右部分は，特定時刻の発話者を存在する方向に従って提示したものである．この様な形で音源の存在ををより詳細な形で提示し，またその時刻での録音データを再生する (*ZF* レベルの機能)．音から音源の判別が出来なくても，音源の位置がある程度限られるならばその音源が何なのか判断しやすくなる．従って，このレベルの機能により音の意味の判断が容易になる．

ユーザーは *ZF*，*D* レベルの機能によって提示された情報から聞きたい音源を探し出す．システムはユーザーの要求に従ってその音源の音を再生する (*D* レベルの機能)．音源ごとに音を再生するため特定音源の発音内容を明確に把握することが出来る．従ってこのレベルの機能により複数の音が同時に発せられている場合でも音を容易に理解することができる．

2.2 音の視覚化の関連研究

音を視覚化するものとして，音カメラ [3]，Noise Vision[4] などがある．これらは音源の存在する方向と音源からの音の音圧，あるいは周波数などの複数の情報を一画面にすべて同時に提示するものである．これは *ZF* レベルの機能のみを実現したものである．この様な視覚化手法は音情報の時間的一覧性の無さ，および弁別の難しさを解決できないため，長時間の音情報を扱うことは出来ない．

3. 音の視覚化機能付き録音再生システム

3.1 システム構成

機能 1a，2a の機能を実現するためには音源が存在する方向を特定する必要がある．この音源定位にはビームフォーミングによる定位をカルマンフィルタにより精度を上げる村瀬ら [5] が開発した手法を採用した．機能 3a の機能を実現するためには音源ごとの音を分離する必要がある．この音源分離は Valin ら [6] により提案されている，Geometric Source Separation による音源分離を Post-Filter によって雑音抑圧処理する手法を採用した．

本システムの構成を図 2 に示す．システムはクライアント・サーバーシステムで構成される．クライアントシステムは，通常のマイク 7ch とサラウンド音用マイク 1ch からなる 8ch マイクを用いて録音をする．録音は各チャ

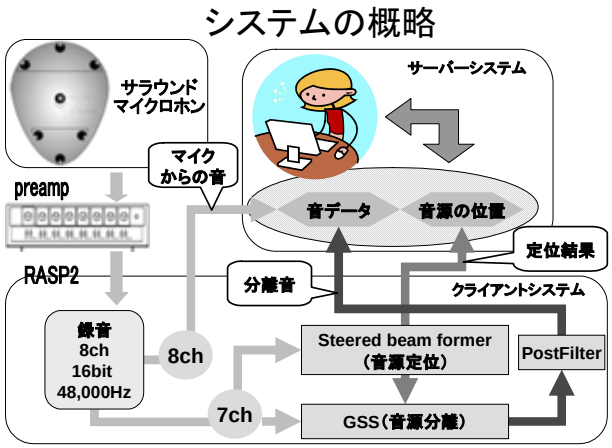

図 2: システム構成

ンネルすべて 16bit，サンプリングレートは 48kHz で行う．マイクからの 8ch の音と音源方向の情報，音源ごとの分離音はサーバーシステムへ送られる．サーバーシステムではそれらの情報を再生表示する．

3.2 インターフェース

本システムの操作画面を図 3 に示す．操作パネル (①) には PLAY (再生)，PAUSE (一時停止)，STOP (停止)，RECORD (録音) ボタンがあり，通常の音再生機器と同じ感覚で使用することが出来る．

タイムグラフ (④) には音源方向を時間軸に沿って示したグラフを表示する．横軸は時間，縦軸は水平方向角度を示す．これは機能 1a を実現するため機能である．

PLAY ボタンをクリックすることにより再生される音は録音された音そのものである．これは機能 2b に対応している．中央の画面では，再生される音に同期してその音源方向をマイク (②) を中心にしてビーム (③) で提示する．これが機能 2a を実現したものである．ビームの上に表示される ID 番号は音源ごとに一意に振られ，色はタイムグラフ (④) と対応している．

機能 3a の実現のためにはユーザが音源ごとの音を指定するインターフェースが必要になる．音源からの音を出力させるには，中央画面のビーム上に表示されている ID をマウスでクリックする．それにより，音源を指すビームは強調表示され再生音が録音された音から指定する音源の音に切り替わる．ある特定範囲にある音源の音を再生する場合は，水平方向の角度範囲，垂直方向の角度範囲を指定する．図 4 では再生する音源方向の範囲を四角形からなる球体 (⑤) を用いて表示している．角度範囲指定は球体の四角形をマウスでクリックすることで行う．

4. 考察

実現したシステムでは音情報を *O-ZF-D* を方針として視覚化を行った．現段階のシステムでは音源の方向情報と音源ごとの音を用いることで，音情報を効率良くユーザーに提示できるようになった．

しかし，音情報からは音源の方向だけでなく，音量や音源の種類の情報を生成できる．音源の種類が人声ならば音声認識，環境音ならば擬音語に変換することによりさらに詳細な情報を得ることができる．これらは音源方向などと同じく視覚化可能であり，*O*，*ZF*，*D* の各レベルで様々な形で提示できる．例えば，音量はタイムグラフ上の定位結果グラフの太さで提示できる．また，音源の種類を取得することにより，定位結果グラフに音源の種類を示すラベルを付加することができる．さらに，音声認識結果や環境音認識結果を用いることで音の内容を詳細に提示できる．

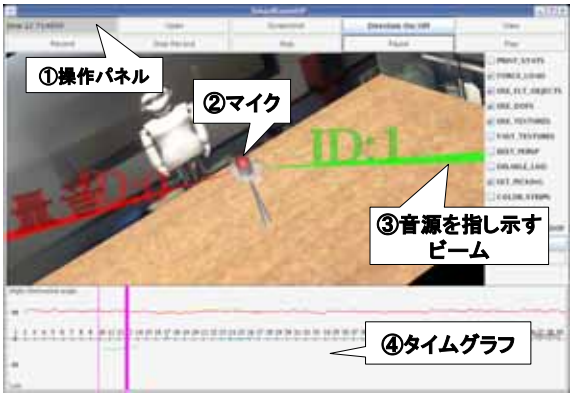

図 3: マウスにより傾けた再生画面

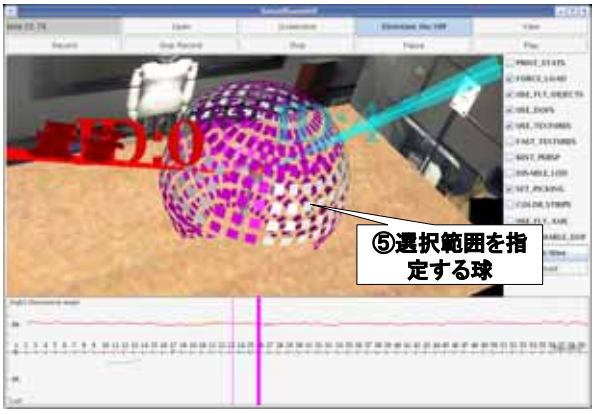

図 4: 範囲指定時の再生画面

5. おわりに

本研究では混合音をサラウンドマイクロフォンで収録し，収録した混合音の再生を高度に行うため，*O-ZF-D* に基づいて音を視覚化する手法を開発した．音源方向を提示することにより，音をブラウジングしたり，必要とする音情報がある時間的空間的位置を容易に探し出すことが可能となった．この結果，混合音であっても，音を弁別しやすいシステムとなった．今後，分離音に対して，その音源同定結果の提示や，分離音が音声であった場合には，その音声認識結果の提示などの機能を付加して，より使いやすいシステムへと発展させていく．

謝辞 本研究は，科研費，21 世紀 COE の支援を受けた．

参考文献

[1] S. Cherry: Total recall, *IEEE Spectrum*, **42**:11 (Nov. 2005.) 24–30.

[2] B. Shneiderman: *Designing the User Interface (3rd Ed)*, Addison-Wesley, 1998.

[3] 中部電力 (株), (株) 熊谷組, 信州大学: 音カメラ, 2001, http://www.aea.ne.jp/data01.html

[4] 日東紡音響エンジニアリング (株): Noise Vision, 2006, http://www.noe.co.jp/system/nsvision.html

[5] M. Murase, *et al.*: Multiple Moving Speaker Tracking by Microphone Array on Mobile Robot, *Proc. of Interspeech-2005*, 249–252.

[6] J-M. Valin, *et al.*: Enhanced Robot Audition Based on Microphone Array Source Separation with Post-Filter, *Proc. of IROS-2004*, 2123–2128.