2B4-9

連続音声録音を用いた会話体験の探索

A Browsing Interface for Dialogical Experience using Continuous Recording Audio Data

河村 竜幸 Tatsuyuki Kawamura　中西 英之 Hideyuki Nakanishi　石黒 浩 Hiroshi Ishiguro

大阪大学大学院工学研究科
Graduate School of Engineering, Osaka University

In this paper we propose a browsing interface in continuously recorded audio data using auditory features. We have implemented memory prosthesis with a bone conduction microphone and an IC voice recorder. Conventional voice recorder tends to cause an inversion of privacy when it is used in informal settings. On the other hand, the user with our proposed system can refer to recorded his or her own voice with a low inversion of privacy. In this paper we first introduce the memory prosthesis device and its interface. We then propose a discrete-continuous user browsing model. Lastly, we explain browsing methods using auditory features.

1. はじめに

近年，ユーザの日常生活をカメラやマイクで常時記録する体験記録の研究が注目を浴びている [相澤 03][Gemmell 06]．また，ユーザの体験を記録した映像や音声を用いてユーザの記憶想起活動を支援しようという試みも盛んである [Jebara 96][Kawashima 02]．しかし，ユーザ視点の映像をカメラで記録する方式ように，外界の情報を積極的に記録する方式では他者にとってプライバシー侵害の可能性がある．このプライバシー侵害の可能性から，外界の情報を積極的に記録する方式が社会で導入されることに対する心理的障壁は非常に高いと考えられる．この問題に対し，著者らは骨伝導マイクを用いることでユーザの発話だけを常時記録する方式を提案している [河村 06]．しかし，ユーザの発話を常時録音してゆくことで蓄積される記録が膨大となってゆくため，連続的に録音された音声データの参照に適したインタフェースを実現することもまた，システム導入の障壁を下げる重要な課題となる．そのため本研究では，骨伝導マイクを用いて連続的に録音されたユーザ発話の記録データ（以後，連続音声録音データ）を探索するためのインタフェースを試作した．また本研究では，音声データの特徴量を計算し，閾値処理することで再生区間を決定する探索方式を採用した．

2. 連続音声録音による会話の再利用

会議の場面で見られるように，言語的情報を相互に交換する会話には重要な内容が含まれることが多く，この会話を音声記録として記録・蓄積するというニーズは過去より存在している．近年，IC レコーダの大容量化・小型軽量化・低価格化による，音声の録音技術の進歩が急速に進んでいる．この録音技術の進歩により，連続音声録音を用いる記憶補助システムの研究が注目を浴びている [Vemuri 04][Hayes 04]．

連続音声録音による記憶補助は重要な技術であると考えられている一方で，ユーザの発話と共に会話対象者の発話を同時に収録する場合，会話対象者のプライバシーを侵害する恐れがある．たとえユーザが私的利用が会話の録音目的であることを会話対象者に表明したとしても，近年ネットで多発する情報漏洩の問題から，会話対象者にとってプライベートな会話を録音されることに対する心理的障壁は高いと考えられる．そこで本研究では外界の音を抑制しユーザの発話音声のみを取得することを目的として開発された骨伝導マイクを採用している．骨伝導マイクを採用することで会話対象者の発話を理解できるレベルで録音することは困難となる．しかし著者らは，会話に対してユーザが積極的に参加している場面では，ユーザ発話の連続音声録音データのみで会話の概要を想起することは可能ではないかという立場である．また，ユーザの発話箇所だけで再利用可能なアプリケーションが存在するのであれば，骨伝導マイクを用いた連続音声録音は十分に意味があるものと考えている．例えば著者らが有効ではないかと考えているのが以下の利用例である．

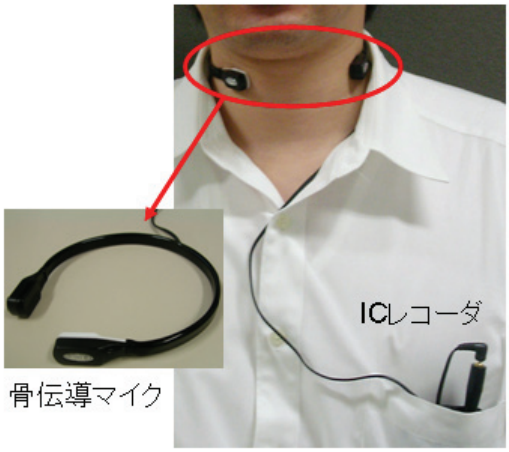

図 1: 連続音声を録音する記憶補助装置

- 言った/言わない論争
- 発言の一貫性
- 自己反省

これらは全て自己の発言を振り返ることが前提となっている．一番目の “言った/言わない論争” では，ユーザ自身が会話対象者に対して実際に作業を指示したか，また指示内容と作業内容が一致しているかを明らかにすることが可能となる．二番目の “発言の一貫性” では，日常のユーザの発話を録音しておくことで，ユーザの体験を顧みることが可能となり，ユーザ自身の人間性を維持・向上させるための判断材料となる．最後の “

連絡先: 河村 竜幸，大阪大学大学院工学研究科，大阪府吹田市山田丘 2-1，Phone 06-6879-7970，Fax 06-6879-7969，kawamura@ams.eng.osaka-u.ac.jp

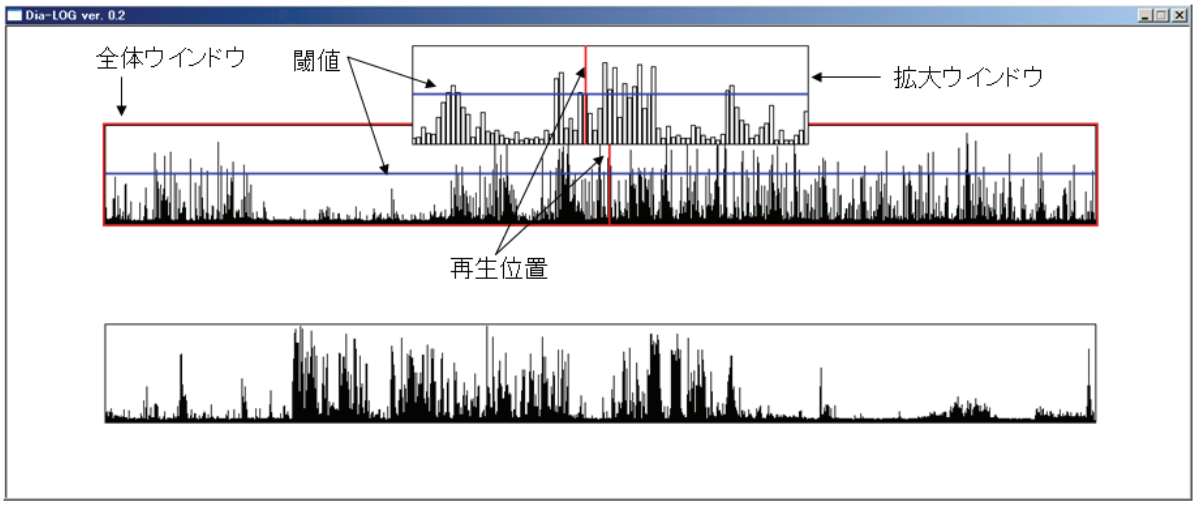

図 2: 連続音声録音データ探索のためのインタフェース

自己反省" では，日常会話で発生するユーザの悪い癖を反省する材料として，また宴会の席のように記憶が曖昧となり思い出せない泥酔状態中の自己を改めて冷静に観察することが可能になると考えられる．このように会話体験が適切に再利用されることで，ユーザ自身だけでなく，ユーザ周辺の人々へ良い影響が与えられる可能性がある．

3. 連続音声録音を用いた記憶補助システム

本研究で提案する記憶補助システムはユーザの発話を連続的に録音するハードウェアと連続音声録音データを参照・探索するためのインタフェースで構成される．本節ではこれら 2 点をそれぞれ説明する．

3.1 連続音声録音用ハードウェア

本研究では，ユーザの発話のみを取得するための骨伝導マイクと長時間の音声録音が可能な IC レコーダを用意した（図 1）．骨伝導マイクには，ユーザの発話以外の外部音声を収録しにくく，かつユーザにとって装着負荷が低く長時間装着可能な FireFox Technologies 社製を採用した．IC レコーダはユーザにとって装着負荷の低い軽量型 (46g，アルカリ電池含む) で長時間の連続音声録音（ステレオ HQ モードで約 35 時間 25 分）が可能な OLYMPUS 社の Voice Trek V-50 を採用した．実利用時には三洋電機社の eneloop を用い，最大一日に 1 回の電池交換が必要であった．

3.2 探索インタフェース

本研究では，連続音声録音を探索するためのインタフェースを開発した．図 2 にインタフェースの概観を示す．ウインドウは全体ウインドウと拡大ウインドウによって構成されている．図では全体ウインドウは 2 つで構成されているため，一度に 2 種類の連続音声録音データをユーザに提示することが可能となっている．ただしユーザは一度に一つの連続音声録音データしか再生することができない．選択また音声データは固定サンプル長を 1 ブロックとする音声ブロック単位で管理されている．

音声ブロック: 固定サンプル長を 1 ブロックとして音声信号を管理している．後に説明する探索に用いる特徴量は音声ブロック単位で計算される．例えば図中の拡大ウインドウに存在する複数の矩形は音声ブロック単位で計算された音量を示している．ここでは矩形が縦に長いほど音量が大きく，短いほど音量が小さいことを意味している．本研究では 16kbps のデータに対して 100,000 点のサンプルで 1 ブロックと定めた．これにより 1 ブロック当り再生時間は 6.25 秒となる．

全体ウインドウ: 全体ウインドウでは 1 連続音声録音データ全体の波形を表示している．ただし，全体ウインドウで表示可能な音声ブロックには限界があるため，ユーザに提示されるのは一部のブロックのみである．例えば図の全体ウインドウの横方向は 1,000 ピクセルで設定されているが，8 時間程度の連続音声録音データの場合は音声ブロックの数が 5,000 程度になるため，全体の 1/5 が全体ウインドウでユーザに提示されていることになる．しかしながら図 2 からわかる通り，ユーザは上の全体ウインドウで提示されている連続音声録音データと，下の全体ウインドウで提示されている連続音声録音データとの違いを理解することが可能となっている．

拡大ウインドウ: 拡大ウインドウは，ユーザが全体ウインドウ上で移動させたマウス座標を中心とした周辺の音声ブロックを提示するため，ユーザは全体ウインドウでは可視化されていなかった音声ブロックを参照することが可能である．

再生位置: ユーザは再生させたい音声ブロックを全体ウインドウ上または拡大ウインドウ上でマウス左クリックにより直接指定するか，マウスホイールを操作することで再生位置を時系列方向に変更させることが可能である．

閾値: ユーザは閾値を変更することで，離散的に連続音声録音データを再生することが可能である．閾値は全体ウインドウ上または拡大ウインドウ上でマウス右クリックにより直接指定することが可能である．例えば音量を特徴量としている場合，閾値を上げることで音量が大きい音声ブロックのみを参照することが可能である．

4. 音声特徴量を用いた探索

本研究では，連続音声録音データの探索方法として，音声データ自身が持つ特徴量に着目した．この音声データが持つ特徴量に対して閾値処理を実行することで参照区間の削減を目指す．

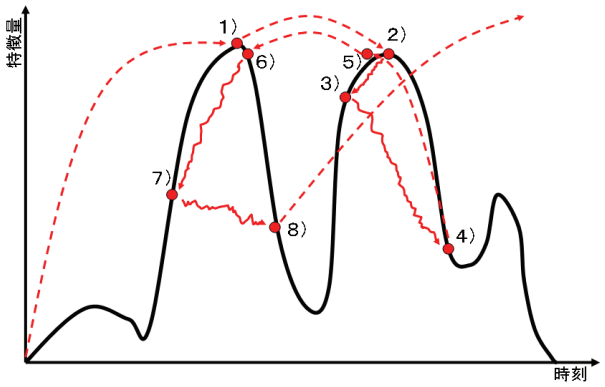

図 3: 離散連続遷移型ユーザ探索モデル

4.1 離散連続遷移型ユーザ探索モデル

ユーザが長時間録音された連続音声録音データを参照する時，個々の発話を詳細に参照すると，膨大な参照時間を必要とする．そのため，連続音声録音データ中の発話を離散的に参照可能である場合，ユーザは参照した一部の発話からその発話が含まれる会話の話題を推定し，参照した発話付近の他の発話を詳細に参照するべきかを判断してゆくプロセスが参照効率を向上させると考えられる．本研究では，この離散的な参照と連続的な参照を遷移させてゆく探索モデルを離散連続遷移型ユーザ探索モデルと呼ぶことにする．離散連続遷移型ユーザ探索モデルでは，離散探索時に参照する発話として話題を推定することが可能な情報を持つ特徴的なものが選択されることが望ましい．

図 3 は，ユーザの連続音声録音データに対する離散連続遷移型ユーザ探索モデルの具体的な実行例を示したものである．実線は連続音声録音データの特徴量を計算した結果の曲線を意味する．破線は再生位置のスキップを意味する．波打った線は巻き戻しまたは再生を意味する．以下で図中の 1) から 8) を説明する．ここでユーザは閾値の設定が可能であるとする．またユーザは音声の早送り・巻き戻しの操作が可能であるとする．システムは閾値を超えている区間のみ音声の再生が可能であるとする．

1) 高い閾値を設定する．最初に現れる特徴的な点（閾値を越える点）にスキップする．ユーザは最初，この特徴的な点周辺の話題に興味を示さなかった．

2) 次に現れる特徴的な点にスキップする．ユーザは特徴的な点周辺の話題に興味を持った．

3) 閾値を下げ，興味を持つ範囲の話題が開始された最初の点に戻る．

4) 閾値を下げつつ，興味を持つ範囲を視聴する．

5) 前に現れた特徴的な点周辺の話題が現在の話題と関わることに気付く．閾値を上げる．

6) 前に現れた特徴的な点にスキップする．ユーザは特徴的な点周辺の話題に興味を持った．

7) 閾値を下げ，興味を持つ範囲の話題が開始された最初の点に戻る．

8) 閾値を下げつつ，興味を持つ範囲を視聴する．

このように本研究で提案する探索方式は，ユーザは特徴的な点の大域的なスキップと，特徴的な点周辺の視聴を繰り返すプロセスとなっている．図 4 は閾値の変更に伴い再生区間が変化することを示している．図中の特徴量は後述する音量である．

4.2 音声特徴量

本研究で採用した特長量は音量と音声信号の分散である．各特長量は音声ブロックごとに計算される．各特長量が持つと考えられる意味を以下に示す．

音量: 音量が非常に小さい値または 0 を示す音声ブロックでは，ユーザは発話していないと考えられる．平均的な音量よりも大きい値を示す音声ブロックでは，ユーザの特徴的な発話，ユーザの主張が強く示されていると考えられる[*1]．そのため，閾値を高く設定するほど，再生区間としてユーザの主張が強く長時間継続した区間が選択されやすくなると考えられる．

音声信号の分散: 音声信号の分散が非常に小さい値または 0 を示す音声ブロックでは，ユーザは発話していないと考えられる．また音声信号の分散を特徴量としているため，分散特徴量が小さいまたは大きい場合は，音量特徴量と比較的相関が高くなる．しかし平均的な値を示す音声ブロックでは音声ブロック内での音量の変化も評価対象となる．図 5 に音量特徴量と相違がある箇所の例を示す．上図が音量の特徴量を示したものである．下図が分散の特徴量を示したものである．図の囲みで音声特徴量の上位 4 つの音声ブロックの値が，分散特徴量では順位が変化していることがわかる．

4.3 先行研究の探索方式との比較

連続音声録音を扱った先行研究として Arons の SpeechSkimmer が挙げられる [Arons 97]．SpeechSkimmer で採用されている方法は音声プレイヤーの早送り・巻き戻しの機能を拡張したものである．無音声区間から発話や話題の切れ目を自動検出し，発話単位・話題単位で再生位置をスキップするシステムを開発している．図 6 に本提案手法と SpeechSkimmer とのスキップ方式の概念的な相違を示す．ここでは横軸を時刻とし，縦軸を音量とする．また有音声区間は音量から計算されるものとする．SpeechSkimmer ではデジタル的に有音声区間と無音声区間を区別している．また，参照時間を減少させるために，有音声区間内の再生速度を上げている．連続音声録音データ中の有音声区間の長さは固定であり．単純に推定される参照時間は有音声区間の長さを再生速度で割った値で示される．有音声区間の長さは固定であるため，参照時間は再生速度を上げることで短縮できる．しかし再生速度を極端に速くすると，連続音声録音を参照するユーザが再生された音声を理解することが困難となる．対して本研究では，有音声区間・無音声区間の明示的な区別をせず，単純にある時刻の音量がユーザの設定する閾値を越えていればその時刻の音声を再生する方式を採用している．本提案方式では，ユーザの設定する閾値によって再生区間の長さは動的に変化し，単純に推定される参照時間は再生時間の長さそのものとなる．連続音声録音の参照時間を短くするには設定する閾値を上げればよい．

*1 音声ブロックのサンプル長が短い場合，瞬間的に音量が大きくなるクシャミも特徴的な区間として取り上げられる．本研究では音声ブロックのサンプル長を約 6 秒に設定しているため，単発的に発声する音声は特徴的な区間として取り上げられない．

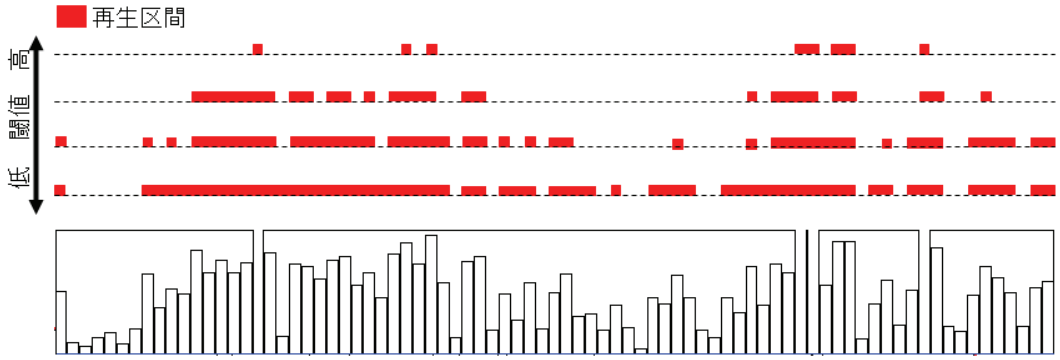

図 4: 閾値を変更することによる再生区間の変化

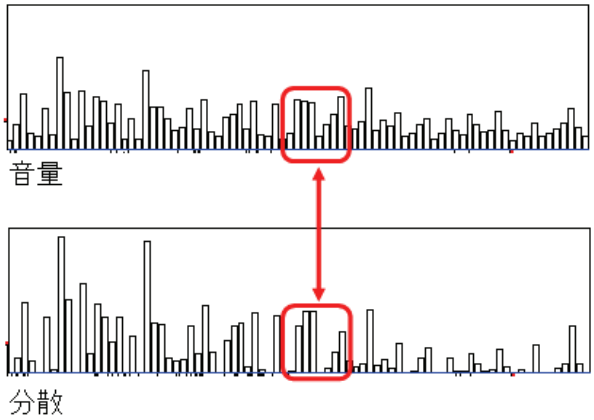

図 5: 音量特徴量と分散特徴量の相違

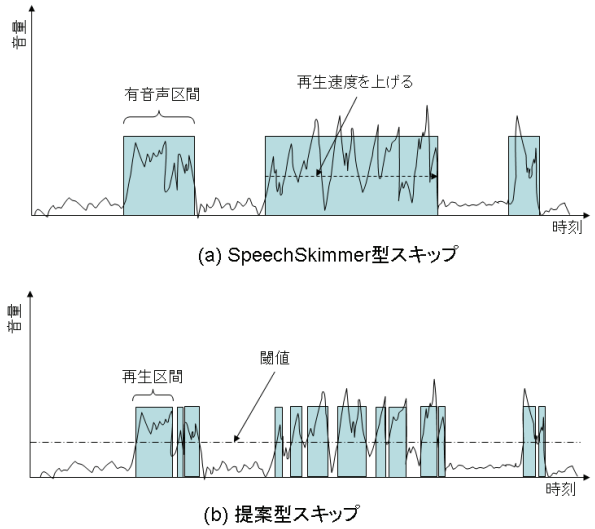

図 6: 連続音声録音のスキップ方式の概念的相違

5. おわりに

本稿では，骨伝導マイクを用いて録音された連続音声録音データを探索するためのインタフェースと音声特徴量を用いた探索方法を提案した．今後，実験協力者を用いて長期運用実験を行い，は利用効率の客観的評価を得る必要がある．

謝辞

本研究の一部は，文部科学省科学技術振興調整費「先端融合領域イノベーション創出拠点の形成：ゆらぎプロジェクト」の研究助成によるものである．ここに記して謝意を表す．

参考文献

[相澤 03] 相澤清晴，石島健一郎，椎名誠: ウェアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み，電子情報通信学会論文誌, Vol. J86-DII, No. 6, pp. 807–815 (2003).

[Gemmell 06] Gemmell, J., Bell, G. and Lueder, R.: MyLifeBits: a Personal Database for Everything, *Communications of the ACM*, Vol. 49, Issue 1, pp. 88–95 (2006).

[Jebara 96] Jebara, T., Schiele, B., Oliver, N. and Pentland, A.: DyPERS: Dynamic Personal Enhanced Reality System, *MIT Media Laboratory, Perceptual Computing Technical Report#463*, (1996).

[Kawashima 02] Kawashima, T., Nagasaki, T. and Toda, M.: Information Summary Mechanism for Episodic Recording to Support Human Activity, *Proc. International Workshop on Pattern Recognition and Understanding for Visual Information Media*, pp. 49–56 (2002).

[河村 06] 河村竜幸，中西英之，石黒浩: 骨伝導マイクと IC レコーダを用いた記憶補助装置の構築, 情報処理学会第 121 回ヒューマンインタフェース研究会報告書，2006-HI-121, pp. 35–42 (2006).

[Vemuri 04] Vemuri, S., Schmandt, C., Bender, W., Tellex, S. and Lassey, B.: An Audio-Based Personal Memory Aid, *Proc. 5th International Conference on Ubiquitous Computing*, pp. 400–417 (2004).

[Hayes 04] Hayes, G.R., Patel, S.N., Truong, K.N., Iashello, G., Kientz, J.A., Farmer, R. and Abowd G.D.: The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid, *Proc. 6th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 168–179 (2004).

[Arons 97] Arons, B.: SpeechSkimmer: A System for Interactively Skimming Recorded Speech, *ACM Trans. on Computer-Human Interaction*, Vol. 4, No. 1, pp. 3–38 (1997).