Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

# イベントの内容、感情、話者情報をロギングする リッチなサウンドライフログの実装と評価

綾部櫻子[†1]　田野俊一[†1]　市野順子[†1]　岩田満[†2]　橋山智訓[†1]

ライフログとは「人間の生活を記録するツール」である．しかし検索が困難であることや UI が使いづらいことなどの問題から，未だ普及していないと言える．そこで本研究では「音」のみに着目したサウンドライフログを提案し，その実装と評価を行った．本研究の特徴は，音声認識，重要語抽出，感情認識，話者識別の技術によって音から①イベントの内容②その時の感情③話者を抽出し，それらの情報を活用して，④音の提示や検索に適した UI を検討することである．音のみを記録し，UI によって画像が主流である現在のライフログの問題点を解消することで，効率的で利便性の高いライフログを目指す．

# Implementation and evaluation of Sound Lifelog that can log contents of the event and the emotion and the speaker

SAKURAKO AYABE[†1]　SHUN'ICHI TANO[†1]
JUNKO ICHINO[†1]　MITSURU IWATA[†2]　TOMONORI HASHIYAMA[†1]

The lifelog is the tool to record the life of human. However, the lifelog has not been used many people, because it is difficult to search and the UI is difficult to use. This paper proposes Sound Lifelog that focuses on the sound or the voice only. This paper is to extract contents of the event, and the emotion from the sound by using the speech recognition and the emotion recognition. This paper proposes the UI which is appropriate to presentation and the search of the sound using of these data. By focusing on the sound, this paper aims to fix some issues of the current lifelog.

## 1. はじめに

ライフログとは「人間の生活を記録するツール」である．記録することによって，いつでも客観的に自分の日常を振り返ることができる．しかし，記録した膨大な情報の中から検索することは困難であることや，UI が使いづらいことなどの問題からライフログは未だ普及していないと言える．

そこで本研究では，「音」に着目した新しいライフログ，サウンドライフログを提案する．本研究の特徴は，日常の音声のみを常に自動的に記録し，取得した音声から音声認識，感情認識，重要語抽出，話者識別の技術によって，①イベントの内容②その時の感情③話者情報を抽出することである．そしてそれらの情報を活用し，④音の提示や検索に適した UI を検討する．

## 2. 関連研究とその問題点

関連研究のその問題点について述べる．

### 2.1 関連研究

ライフログの始まりは「Memex」[1, 2]であると言われている．「Memex」は，ドキュメントを中心に格納されている個人データを保管したり，検索，見出しを付けたりすること支援するシステムの概念である．現在では一生分のデータの保管，検索が可能となった．

現在商品化されているライフログツールとして，SenseCam(Microsoft)[3]がある．これは，小型のウェアラブルカメラであり，首から下げて利用する(図 1)．搭載されているセンサに反応，あるいは 30 秒ごとに自動的に画像を取得してくれる．このように，現在のライフログの主流は画像を利用したものとなっている．

図 1 SenseCam

画像以外のデータを利用しているライフログに，ライフ顕微鏡[4]がある．これは腕時計型のセンサであり，3 軸加速度，脈波センサ，温度センサにより腕の動きなどを 24 時間計測する．取得したデータは「ライフタペストリー」

†1 電気通信大学大学院　情報システム学研究科
Graduate School of Information Systems, The University of Electro-Communications
†2 東京都立産業技術高等専門学校　ものづくり工学科
Tokyo Metropolitan College of Industrial Technology

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

という形式で表すことができ，ユーザは自分の習慣や生活リズムを把握することができる(図 2).

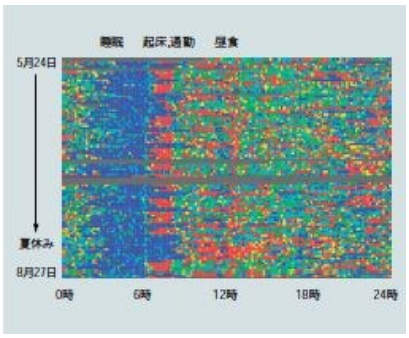


図 2　ライフタペストリー

また，ライフログデータの 1 つである音声の可視化に関する研究も行われている[5, 6, 7]．研究[5]では，電話での会話を声の音量, ピッチ, コンテンツを利用し可視化する．

### 2.2 関連研究の問題点

以上の関連研究の問題点をまとめる[8].

(a)従来のライフログの問題点

①画像に大きく依存しているため情報量が多く，効率的な閲覧・検索に適した UI が不足している

②適したソフトウェアや画像処理技術がなく，組織化が上手くできていない

③不要なデータはユーザが判別して手動で消さなければならなく，負担が大きい

④そもそも人はデジタルデータへのアクセスや管理に時間を割かない

⑤画像以外のデータは情報がプアなために詳細な思い出しを支援するのが難しい

(b)音声の可視化に関する研究の問題点

⑥短時間の会話の視覚化しかされていない

⑦一目で分かる情報が少なく，過去の会話を再現することが難しい

⑧検索に対応していない

これらの問題点を踏まえ，本研究では新しいライフログ提案する．

## 3. 本研究のアプローチ

### 3.1 アプローチ

人間は五感(視覚・聴覚・触覚・嗅覚・味覚)により外界から情報を得ている．一般に視覚から獲得する情報量が全体の約 8-9 割であり，次に聴覚が 1 割ほどの情報量であると言われている．視覚情報，即ち画像を利用したライフログは多く存在しているが，聴覚情報である音声のみを利用したライフログは未だ存在していない．音声は情報量が少ないが，少ないからこそデータ処理が容易などというメリットも考えられる．そこでライフログデータとしての音声の可能性について考えてみた．

### 3.2 音声の可能性

音声はハンズビジーな環境下でも気軽に記録できるという利点があるため，音声を利用したアイデアメモ[9]や音声対話システム[10]は多く存在している．研究[10]では，音声から感情を抽出し活用することにより，運転手と車のリッチな音声インタラクションを目指している．

人が過去を想起するときに，手がかりとして利用される出来事の 4 つの W と 3 つの性質がある[11]. 4 つの W とは Who-What-Where-When を指す．その中の Who-What は，現在の画像処理技術では記録することは難しい．しかし音声ならば，声から誰と話していたか，また環境音や会話内容から何をしていたかを記録することが可能であると考えられる．

出来事の 3 つの性質とは出来事の稀さ，感情喚起度，快不快度を指す．これらも同様に，画像処理では記録することが難しい．しかし音声ならば，声の抑揚などから感情を記録することができ，話者や会話内容の稀さを出来事の稀さとして記録することができると考えられる．

### 3.3 音声利用によるメリット

以上のことを踏まえ, また音声と画像を比較してみると，ライフログに音声を利用することによって次のようなメリットが考えられる．

①思い出すのに重要な 3 つの情報を記録できる

i.その時のイベントの詳細な内容が分かる

ii.その時の感情が分かる

iii.その時に交流していた話者が分かる

②デバイスを身につける当事者の負担の軽減

③記録されているという当事者，第三者の負担の軽減

よって本研究では，ライフログデータとして音声のみを利用することを提案する．

### 3.4 サウンドライフログの提案

これより，音声のみを利用したライフログをサウンドライフログとする．関連研究の問題点と音声利用のメリットを踏まえ，本システムの機能とデバイスの要件を述べる．

#### 3.4.1 機能の要件

(A)　音声から 3 種類の情報を抽出

取得した音声から①会話内容②感情③話者の 3 つの情報を抽出する．

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

(B)　音声の一覧表示

音声はアイコン化し，抽出した 3 つの情報にはそれぞれ色を割り当てる．表示の際，一生分を一目で把握できるように一覧表示にする．

(C)　過去の出来事へのトリガー

過去の思い出に積極的に触れられるよう，最近の音声から抽出した情報を利用し，それらと一致する過去の音声を強調・点滅させる．

(D)　音声の検索

音声から抽出した会話内容，感情，話者を元に，過去の音声を検索することができる．

(E)　シームレス表示

誰もが直感的に操作できるよう，シームレス表示とする．

#### 3.4.2 デバイスの要件

デバイスは，どこでも思い出にアクセス可能にするため，携帯端末であるスマートフォンで実現する．また，一瞬の思い出を逃さないよう音声は常に取得することとする．

## 4. 実証実験

### 4.1 目的

音声のみから，会話内容・感情・話者がどのくらい分かるのか検証する．

### 4.2 実験手順

①IC レコーダで音声を取得

被験者は，本研究室の学生 4 名と他研究室の学生 1 名，合わせて 5 名の男子学生である．IC レコーダ(DS-800, OLYMPUS)を渡し，5 人全員で集まって会話をしている様子の音声を取得してもらった．

②Wizard of Oz 法で発話の分類

今回取得した音声の 1 時間 6 分 31 秒のうち，任意に選択した 3 分 51 秒間の会話，100 発話分を抽出し，Wizard of Oz 法(人間がシステムの代わりに処理を行う実験方法)によって会話の文字起こし，感情・話者の分類を行った．感情については平常・喜び・怒り・悲しみの基本的な 4 感情とする[12].

③分類の正誤確認

音声の分類後は，発話した本人へ正誤確認を行った．実際に音声を聴いてもらい，会話内容，感情，話者が誤っていた場合，訂正してもらった．

### 4.3 実験結果と考察

表 1 に会話内容，話者，感情，そしてその他に抽出された笑い声に関して，筆者が認識・分類できた発話数と，被験者本人の認識と一致していた発話数を示す．

表 1　分類した発話数と一致した発話数

| 分類 | | 分類した発話数 | 一致した発話数 | 一致率 |
|---|---|---|---|---|
| 会話内容 | | 100 | 67 | 67% |
| 話者 | | 100 | 84 | 84% |
| 感情 | 平常 | 90 | 90 | 100% |
| | 喜び | 1 | 1 | 100% |
| | 怒り | 0 | 0 | - |
| | 悲しみ | 0 | 0 | - |
| 笑い | | 9 | 9 | 100% |

(a)会話内容に関して

会話内容の音声認識に関して，以下のような場合，認識が難しいことが分かった．

①発話の語尾

②新しい流行語

③発話が多人数で重なる

しかし文章全体ではなく単語であれば，多くを認識することができた．単語だけ抽出できれば，そこから会話内容を推測することは可能であると考える．

(b) 話者に関して

今回の実験では5名中4名が本研究室の学生であるため，その 4 名に関しては筆者が聞いたことがある声である．しかし，筆者は音声取得現場にはおらず，また 5 名全員が男性であったため，声の判別が難しい場面が多少あった．特に短い発話や，いくつかの発話が重なる場合の認識は困難であった．

よって話者認識には，予め学習させること，発話が重なっても認識できる技術が必要であると考える．

(c)感情に関して

「喜び」の感情は，全体的に声の抑揚が小さかったため，抑揚だけでの認識が困難であった．一方で，「笑い」は多くが認識できた[13]．その理由として，笑い声は抑揚の変化が大きいためであると考えられる．

以上のことから，笑い声に加え，怒鳴り声，泣き声などの抑揚が大きく変化する音声を認識すれば，それが感情として分類できると考える．ため息やくしゃみ，しゃっくり等の音声も同様に認識できる可能性がある．

以上の結果より総合的に判断すると，サウンドライフログの機能の要件は工夫次第で実現可能であると考えられる．

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

## 5. システムの設計と実装

本システムの設計と実装について述べる．

### 5.1 システム構成

図 3 が本システムの構成図である．

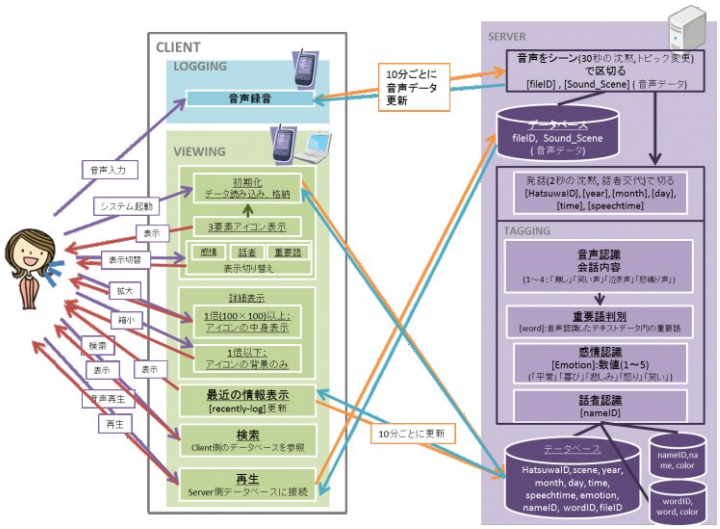


図 3　システム構成図

サーバ側では, 取得した音声を認識技術によって分類し, データベースに格納していく．クライアント側では取得した音声のレビュー，検索が可能となる．

データベースはユーザを大学生と仮定し 20 年分を, 一部は実データをもとに，一部は仮想的に作成した．

### 5.2 ハードウェア

本システムの開発には以下のハードウェアを用いた．

①lenovo ThinkStation D20 ( Microsoft Windows7 PC)

②富士通 Windows7 ケータイ　F-07C (図 4)

③wacom Cintiq 24HD touch(Microsoft Windows7)


図 4　②F-07C(富士通)

### 5.3 認識方法

本システムで必要な認識方法は次の 4 つである．認識については本研究の主題ではないため，簡易に実現した．

(1)　音声認識：音声から重要語を抽出するため，会話内容を音声認識によりテキスト化する．

(2)　重要語抽出：重要語の定義は，①会話のトピックとなる頻出単語，②聞きなれない単語，初めて聞いた単語とする．テキスト化された会話から抽出する．

(3)　感情認識：周波数分析，音量，抑揚から笑い声や怒鳴り声などの特徴的音声を判別し，「平常」「喜び」「怒り」「悲しみ」「笑い」の 5 種類に分類する．

(4)　話者識別：周波数分析により，男女の違いのみを判別する．

### 5.4 GUI デザイン

本研究では，GUI を中心に設計・実装を行った．

#### 5.4.1 起動画面

図 5 に本システム起動画面を示す．図中のアルファベットは 3.4.1 項で述べた機能を指している．

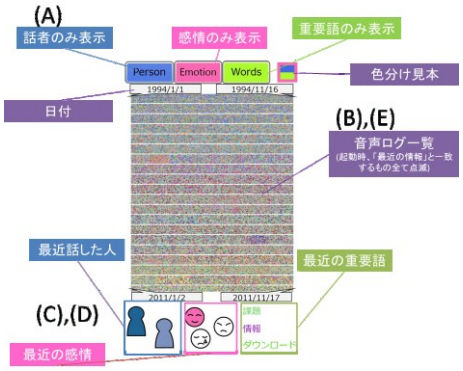


図 5　起動画面デザイン

画面は, 3.4 節で述べた機能の要件を満たすようなデザインになっている．画面内は大きく 3 つに分かれており，上部には表示切り替えボタン，真ん中には音声一覧，下部には最近 10 分間の音声から抽出した感情，話者，重要語の表示画面/検索画面となっている．

#### 5.4.2 音声表示

音声は発話(1 人の話者が一息で発言した言葉)と会話 1 シーン(複数の発話が連続しているもの)の 2 種類に分けられ，それぞれアイコン化する．発話アイコン，シーンアイコン共に基本的なデザインは図 6 に示す．また，会話 1 シーンには複数の発話が含まれていることから，最終的にシーンアイコン内にそのシーンに含まれる発話アイコンを並べ，1 つのアイコンとなる(図 7).

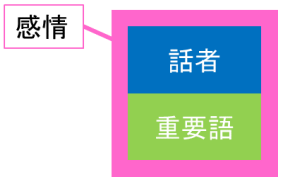


図 6　　基本的なアイコンデザイン

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15


図 7　発話アイコンを含むシーンアイコン

アイコン内にはそれぞれ音声から取得した感情，話者，重要語情報を載せる必要がある．本研究ではそれらの情報を一目で把握することを可能にするため，色を利用した．それぞれの発話における感情には，意味を赤＝怒りのように意味を連想させる色を使用した．また話者，重要語はランダムに色を付けた．その 3 色を発話アイコンに割り当てていく．

シーンアイコンの色は，含まれている複数の発話の割合が多い色，つまりその会話を代表する要素の色とした．

色が付けられたシーンアイコンは図 5 の音声一覧画面に時間順に並ぶ．左上が一番古く，右下が一番新しい音声となる．1 行は 1 年を表している．アイコンに色を割り当て，一生分を一度に並べることで，一生を通しての変化を一目で把握することが可能となる．

音声一覧画面はシームレス表示となっており，拡大していくと図 8 のようになる．シームレス表示により，誰でも直感的に音声間を行き来することが可能となる．

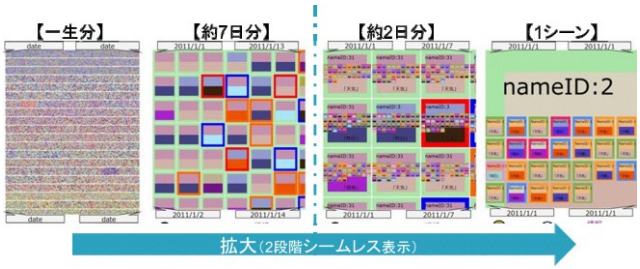


図 8　シームレス表示

### 5.4.3 表示切り替え

シーンアイコンの色は，俯瞰して見た時に認識できる色である．よって感情，話者，重要語において重要視したい要素を選択することで，色の表示を切り替えることができる．切り替えに応じてシーンアイコンのデザインは図 9 のように変化し，選択した色を強調させる．

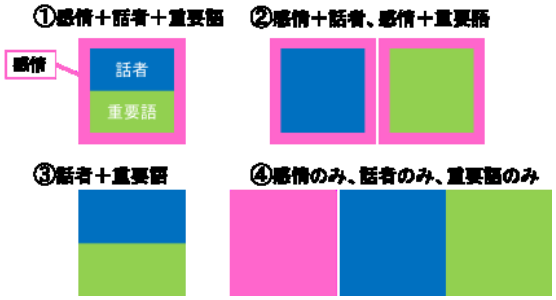


図 9　シーンアイコンの表示切り替え時のデザイン

### 5.4.4 過去の出来事へのトリガー

過去の出来事に積極的に触れられるように，最近の音声から抽出した情報を活用した．最近の音声から感情，話者，重要語を抽出し，それらの情報と一致する情報を持つ過去の音声アイコンを点滅させた．

まず最近の音声と共通の情報を持つ過去の発話アイコンを点滅させ，そこから点滅している発話アイコンを含むシーンアイコンを点滅させた．システム起動時には，最近の音声から抽出した 3 つの情報それぞれを含む，過去の音声全てを点滅させている．これにより，目的なく起動しても点滅が目に留まり，予期していなかった新たな思い出しや発見を支援できると考えられる．

### 5.4.5 検索

最近の音声からの情報が表示されているエリアを使用して，過去の音声を検索することができる．

最近の音声からの情報，例えば話者を示すアイコンを長押しすることによって，今まで交流したことのある話者リストが表示される(図 10)．ユーザはその中から好きな話者を選択すると，先程長押しした話者アイコンの色が変わる．再度話者アイコンをタッチすることで選択した話者による発話が含まれる，シーンアイコンが点滅する．

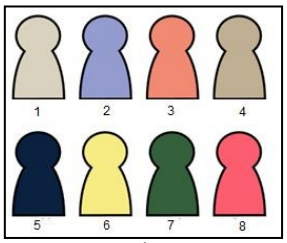


図 10　話者リスト

## 5.5 動作例(表示切り替え)

表示切り替え機能を使用すると，感情，話者，重要語から変化を見たい情報を自由に選択できる．

(A)　3 要素表示[感情，話者，重要語]

感情，話者，重要語全てを表示すると図 11 のようになる．

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

範囲 A を見てみると，少しオレンジが目立っているのが分かる．ユーザはそれを見て，オレンジは感情の喜びを表していることから何か楽しいイベントが起こったのではないかと推測できる．また範囲 B は様々な色が混ざったように見える．話者あるいは重要語の種類が増え，色々な出来事があったのではないかと推測できる．

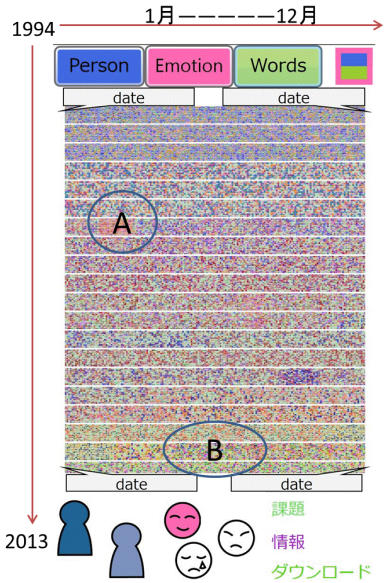


図 11　3 要素(感情，話者，重要語)表示画面

(B)　話者のみ

図 12 は話者のみの色にした画面である．範囲 C を見ると，色合いが大きく変化していることが分かる．このことから，自分を取り巻く環境が大きく変化したことが分かる．また範囲 D は他と比べて色の種類が多いのが分かる．ユーザは大学生になり，多くの人と交流を持ち始めたことを読み取ることができる．

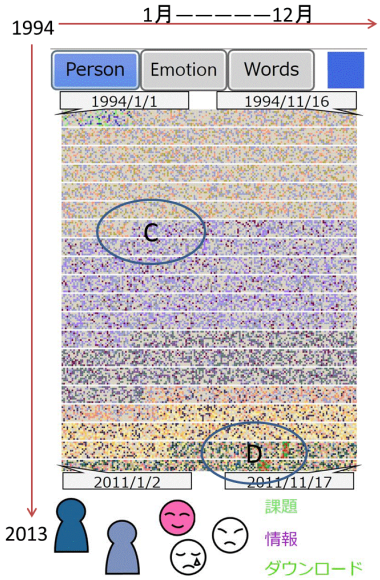


図 12　話者表示画面

(C)　感情のみ

感情のみの色に切り替えると図 13 のようになる．範囲 E を見ると赤が目立っている．この時期はユーザが小学生の頃であるので，感情を抑えずよく怒っていたこと，または周りに怒られていたことが推測できる．範囲 F は緑が目立つ．緑は感情の平常を表す色である．よってこの時期は特別なイベントがほとんど無かったということが分かる．

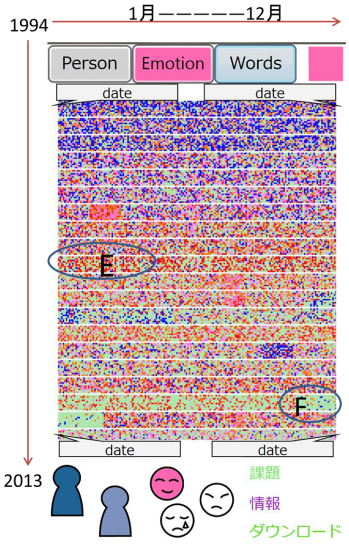


図 13　感情表示画面

(D)　重要語のみ

重要語のみの色に切り替えると図 14 のようになる．範囲 G を見ると，紫が多いことが分かる．よってこの時期はユーザが小学生であるので，小学生時代約 3 年間に渡って流行していた言葉があることが分かる．範囲 H を見ると，色合いが変化しているのが分かる．この頃ユーザは大学受験期であるので，受験期によく発言していた言葉(テスト，勉強など)があったのだと読み取ることができる．

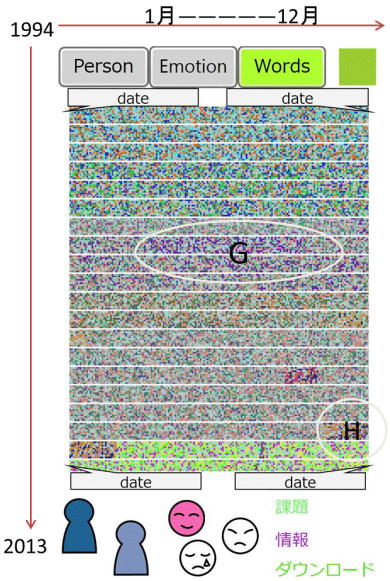


図 14　重要語表示画面

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

## 6. 評価実験

実装した本システムの有用性を評価するため，3 つの実験を行った．

### 6.1 シナリオ実験

3 人の被験者に，用意したシナリオ(18 ステップ)に沿ってシステムを操作してもらった．

その結果，被験者は全てのシナリオを達成可能(達成率 96%)であり，操作性の容易さに対して評価を得ることができた．また，被験者は個々の機能を上手く使えていた．

### 6.2 課題解決実験

3 人の被験者に，本システムの機能を組み合わせて解く問題を 12 問出題し，ユーザが自分自身で本システムを利用できるか検証した．

その結果，平均正答率は 81%であった．また，被験者は本システムの機能を自然に使いこなせていたことから本システムの操作性の容易さが示された．そして音声一覧表示を利用しての平均正答率は 75.1%，検索機能を利用しての平均正答率は 83.5%から各機能の有用性が示された．

### 6.3 自由利用実験

本システムはライフログであるため，実験には長期的な時間の確保が求められる．しかし長期的な時間の確保は難しい．また個人ごとに経験は大きく異なるため，条件を統一できず，公平な評価ができない．これらの理由から自由利用実験の実現は難しい．よって本研究では，新しい評価手法である「共有体験を利用した実験方法」を考案した．

この評価方法の手順は次の通りである．

①被験者全員にドラマ・映画を観てもらうなどして体験を共有する
②時間をおく
③被験者全員共通の問題を出題
④被験者にライフログを用いて解答してもらう
⑤解答結果を定量的に分析する

今回は 2 人の被験者に同じドラマを 60 分視聴してもらうことで共有体験してもらった．2 日後，本システムを利用し 7 問の問題を解答，その後自由に利用してもらった．

その結果，平均正答率は 85.7%であった．日常をテーマにしたドラマの何気ないシーンでも，本システムを利用し正答することができていた．本システムを利用することで，それまで忘れていた記憶を思い出す姿も見られた．

図 15 に，本システムの機能別利用回数を示す．被験者は色々な機能を駆使して思い出すことができていた．

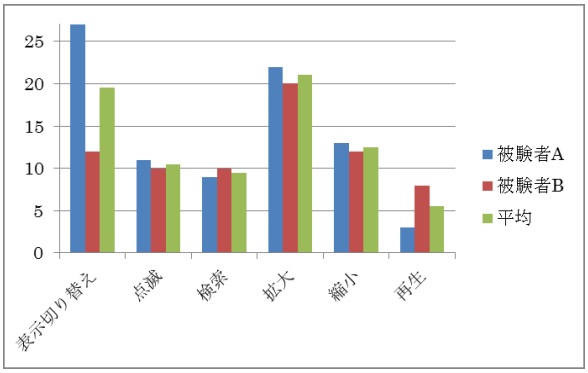


図 15　機能別利用回数

また，図 16 より，被験者は表示切り替えにおいても様々な組み合わせを駆使していたことが分かる．

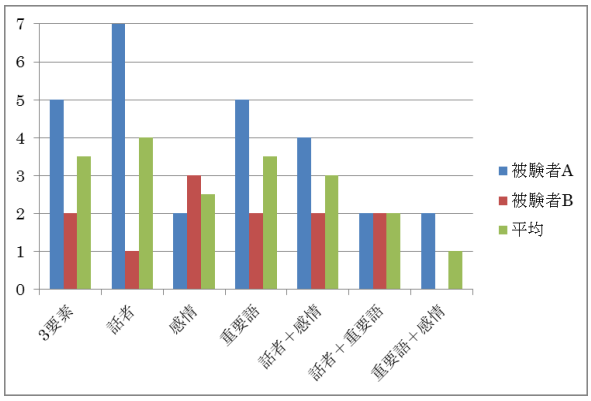


図 16　表示切り替え　種類別利用回数

以上より，本システムのライフログとしての有効性，そして機能の有用性が示された．

## 7. まとめと今後の課題

本研究では，音のみに着目した新しいライフログであるサウンドライフログを提案した．特徴となる機能は次の通りである．

①音声から 3 種類の情報を抽出［感情・話者・重要語］
②音声の一覧表示
③過去の出来事へのトリガー
④音声の検索
⑤シームレス表示

従来ライフログの問題解決に必要な機能を分析，サウンドライフログシステム全体の設計を行い GUI 中心に実装した．その後 3 つの評価実験を行い，本システムの機能の有用性とライフログとしての記憶支援の有効性が示された．

今後の課題としては，次の点が挙げられる．

①システム実装の強化
②実際のデータを利用しての実験

Vol.2013-MBL-65 No.31
Vol.2013-UBI-37 No.31
2013/3/15

## 参考文献

1) Bush, V.: As we may think, Atlantic Monthly Vol. 176, No. 1, pp.101–108, 1945
2) Abigail Sellen, Steve Whittaker: Beyond total capture: A constructive critique of lifelogging, Communications of the ACM, Vol. 53, No. 5, pp.70-77, 2010
3) Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Bulter, Gavin Smyth, Narinder Kapur, Ken Wood: SenseCam: A Retrospective Memory Aid, Microsoft Research, Ubicomp 2006, LNCS 4206, pp. 177-193, 2006
4) 矢野和夫: センサは Web を超える―省力化から知覚化へ―,情報処理, Vol. 48, No.2, pp.160-170, 2007
5) Pooja Mathur, Karrie Karahalios: Visualizing Remote Voice Conversations, CHI EA '09, pp. 4675-4680, 2009
6) Tony Bergstrom, Karrie Karahalios: Conversation Clock: Visualizing audio patterns in co-located groups, Proceedings of the 40th Annual Hawaii International Conference on System Sciences, pp. 78, 2007
7) Tony Bergstrom, Karrie Karahalios: Conversation Clusters: Grouping Conversation Topics through Human-Computer Dialog, CHI '09, pp. 2349-2352, 2009
8) 綾部櫻子，田野俊一，市野順子，岩田満，橋山智訓: イベントの内容，感情をロギングするリッチなサウンドライフログの提案, 日本知能情報ファジィ学会, FSS, 2012
9) 川西康介, 大平茂輝: 研究ノートに基づく音声ログの整理とその利用, 名古屋大学工学部電気電子・情報工学科, 研究報告知能システム(ICS), Vol. 2012-ICS-167, No.3, pp. 1-6, 2012
10) Christian Martyn Jones, Ing-Marie Jonsson: Automatic Recognition of Affective Cues in the Speech of Car Driven to Allow Appropriate Responses, School of Mathematical and Computer Sciences Heriot-Watt University, OZCHI '05 Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future, pp. 1-10, 2005
11) 海保博之, 加藤隆: 認知研究の技法, 福村出版, pp.138-141, 1999
12) 柴崎晃一，光吉俊二: 抑揚からの感情認識の評価―感性制御技術(ST)と，人間の感情の評価法について―，電子情報通信学会技術研究報告.TL, 思考と言語 105(291), pp.45-50, 2005
13) 松村雅史，辻竜之介: 笑い声の無拘束・長時間モニタリング―爆笑計―電子情報通信学会技術研究報告.WIT，福祉情報工学 , 105(372), pp.7-12, 2005