2-404-53
2017/3/2

音環境比較による会話場検出と応用システムの提案

遠山 魁[1,a)]　角 康之[1]

概要：会話グループや同じコンテキストを共有しているグループの検出の手法として、近年では音環境比較を用いた研究が多く提唱されている。しかし、検出精度のさらなる向上を目的とした研究が多く、音環境比較を用いることで得られる情報の活用についての議論までに至っていない。そこで本研究では、音環境比較による手法を基盤とした応用システムの例の提案を行う。応用システムとしてミーティングキャプチャと 3D 空間上での会話場可視化の二つを提案する。

Propose the System Depending on Conversation Groups Based on Auditory Similarity

Kai Toyama[1,a)]　Yasuyuki Sumi[1]

Abstract: The method based on Auditory similarity was proposed as co-location detection instead of previous method in recently. But, most research's purpose is increasing precision of method based on auditory similarity. Therefore we think, we should discuss applying this method to systems or services. In this paper, we propose system that use method based on auditory similarity. We show two system that called meetingcapture and Visualize conversation filed in 3D virtual space.

1. はじめに

本研究の目的は、音環境比較による会話場検出手法を用いたサービスや応用システムの例を提案することである。応用システムとは、得られる会話場の情報を基にサービスを行うものを指す。しかし従来手法によるものの多くは装着する計測デバイスの拘束性の強さや、部屋のサイズや密度などによっては身体間の距離や向きが大きく異なるため正しく検出できないといった問題を抱えている。そこで、本研究では音環境比較による会話場検出を用いて応用システムの提案を行う。音環境比較による会話場検出とは、音を特徴量として会話しているかどうかを判定する検出手法である。この手法で得られるコンテキスト情報を基にサービスを行うようなシステムとして、ミーティングキャプチャと 3D 空間上での会話場可視化の提案を行う。

[1] はこだて未来大学
Future Uniersity Hakodate
[a)] k-toyama@sumilab.org

2. 関連研究

2.1 従来手法による研究

従来手法とは会話グループ検出を身体性から得る手法のことを指す。人類文学者の Edward Hall[1] が対人距離は一定であると定義しており、従来手法のアイディアはこの Hall の定義に基づいたものである。具体的には赤外線 [2, 3, 4]、Bluetooth[5] や WiFi[6, 7] などの電波、超音波 [8]、それらを複合的に用いることで身体間の距離や向きを計測し、会話グループの判定を行う。また、従来手法を用いて計測した会話グループをグループウェアやライフログに活用している。ただし、図 1 に示すような部屋のサイズや人の密度が異なる状況下では Hall が提唱した定義では不十分であることが考えられる。また、壁越しである場合にうまく判定できないことが問題点として挙げられる。以上の問題点から物理的距離による計測から会話場を得ることは難しいと考える。そこで、本研究では音環境比較による会話場検出手法による応用システムの提案を行う。

© 2017 Information Processing Society of Japan

図 1　部屋の大きさや密度によって身体間の距離が違う様子
Fig. 1　Distance between bodies depending on population density and room sizes

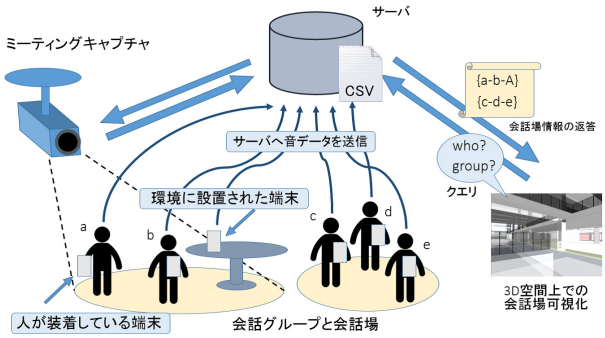

図 2　音環境比較による会話場検出と応用システムの概要図
Fig. 2　Overview of Proposal System

2.2　音環境比較による検出手法の研究

従来手法に対して、音環境比較による会話場検出を行ったのが中蔵らの Neary[9] である。同じ体験を共有している人同士はお互いの声や周りの物音を聴いているというアイディアに基づいたもので、音の周波数成分の類似度から会話グループの判定を行う。具体的にはマイクが搭載された端末を参加者が持ち、会話や周囲の環境音を 6 秒ごとに計測する。そして計測された音の内、有音区間をを高速フーリエ変換 (FFT) し、得られた周波数特性のうち 50Hz から 1600Hz までの 1Hz 毎の周波数スペクトルをベクトルとみなしてコサイン類似度を求めるというもので、算出された類似度が閾値 0.775 を超えた場合、会話グループとみなす。適合率 96.6 パーセントで、再現率 67.9 パーセントで検出することができる。しかし、Neary では端末による P2P 通信を行って会話場の判定を行っているため、応用システムがアクセスするにはこの方式では不適切である。また、Neary では誰といるかという点に着目しており、場所の情報については考慮していない。

Neary と同様に音を特徴量として近接性を図る研究としては佐藤らの CoHear[10] や W.-T. Tan の The Sound of Silence[11] などがあげられる。The Sound of Silence では定常音と非定常音の抽出を行うことによって、マイクの周波数特性を吸収している。また、CoHear ではマイクの周波数特性の吸収のほかに、音信号の復元を困難にすることでプライバシーに考慮したり、サーバ方式にすることでスケーラビリティの向上を図っている。しかし、音環境比較で得られた結果を用いてシステムに活用することについては述べられていない。

そこで本研究では以上のような検出精度の向上ではなく、音環境比較による検出手法を用いた応用システムの提案を行う。また、応用システムが会話場情報にアクセスする必要があるため、CoHear のようなサーバ方式による会話場判定を行う。場所の情報については音を計測する端末を環境側に固定することで解決する。

3.　音環境比較による会話場検出と応用システム

本節では、本研究が提案する音環境比較による会話場検出と応用システムの全体像について述べる。図 2 に提案するシステムの概要図を示す。本研究が提案するシステムでは、個人が持つ携帯端末 (以下、人ノード) や、場所に固有な環境に埋め込まれた携帯端末 (以下、環境ノード) によって各々の会話音声や環境音を取得する。計測された環境音をサーバ上で計算し会話場の判定を行い、会話場の情報をサーバに蓄積していく。そして、応用システムであるミーティングキャプチャや 3D 空間上での会話場可視化がサーバに定期的にアクセスし、会話場情報を得る。会話場の状況によってサービスを行っていくというものである。以下 4 章では音環境比較による会話場検出手法について、5 章で応用システムについて詳細を述べる。

4.　音環境比較による会話場センシング

4.1　端末による音環境の計測

本節では、音環境比較による会話場の検出手法について述べる。音環境の計測方法については人ノード、環境ノードともに Nexus5 を用いた。Nexus5 に内臓されたマイクを使って音の計測を行う。また、マイクの周波数特性を考慮して同一の端末を用いた。端末の装着や設置については図 3、図 4 に示す通りである。人ノードは内臓マイクを上向きにして胸ポケットなどに装着する (図 3)。環境ノードを任意の地点、物に設置することで、人ノードがどこで会話しているかといった大まかな位置を知ることができる。図 4 では机の上に端末が置かれており、この机に設置された環境ノードと人ノードがグループであると判定されることで、ある人物が机の近くで会話しているということをシステムが認識できるようになる。

4.2　音環境比較のアルゴリズム

本節ではアルゴリズムについての詳細について述べる。本システムで用いるアルゴリズムは主に Neary のアルゴリズムを参考としている。図 5 に音環境比較のアルゴリズムのブロック図を示す。以下に端末側とサーバ側、サーバへのアクセスについてそれぞれについて述べる。

© 2017 Information Processing Society of Japan

図 3　端末を装着している様子
Fig. 3　Wearing device with microphone upward

図 4　環境側に設置された端末
Fig. 4　Installing device on environment or things

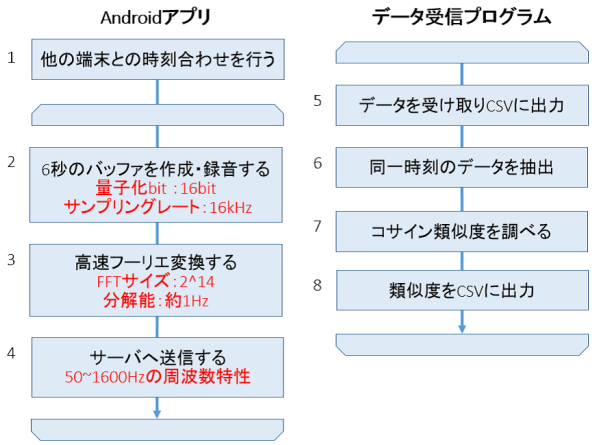

図 5　音環境比較手法によるアルゴリズム
Fig. 5　Algorithms of method based on auditory similarity

4.2.1　端末側処理

端末側処理としては、データの録音と信号処理、サーバへの送信を行う。まず全ての端末の時刻合わせを行う (図 5：1)。これは、端末ごとの時刻に誤差があるためである。そこで、各端末間で User Datagram Protocol(UDP) 通信を行い、通信にかかった時間から逆算し時刻誤差を求め、時刻合わせを行う。次に、6 秒ごとのバッファを作成し録音を行う (図 5：2)。録音した 6 秒の音声データを FFT によって 50Hz から 1600Hz の約 1Hz ごとの周波数特性を得る (図 5：3)。そして得られた計 1551 個のデータに加え、計測時刻、端末 ID をサーバに送信する (図 5：4)。Neary のアルゴリズムでは 6 秒バッファ中の有音区間 1 秒を抽出していたが、時間ずれに弱い点と、再現率を下げる要因になると考えたため 6 秒分のデータを比較に用いた。サーバに送信後は計測を中止するまで図 5 中の 2 から 4 の処理を繰り返す。

4.2.2　サーバ側処理

サーバ側の処理としては、端末から送られてきたデータの記録と類似度の計算を行う。まず、端末側から送信されてきたデータを受け取り逐次 CSV へ書き出す (図 5：5)。次に、同一時刻の周波数特性のデータを全端末分抽出する (図 5：6)。抽出したデータ同士のすべての組み合わせでコサイン類似度を算出する (図 5：7)。算出された類似度と計測時刻、ペアとなっている端末 ID を別の CSV に書き出す (図 5：8)。またこれらの処理はバッチ処理で行っており、計測を停止するまで以上の処理を繰り返す。

4.2.3　サーバへのアクセスと会話グループの判定

本システムでは、会話場の判定はバッチ処理で行っておらず、5 章で述べる応用システムがアクセスしたタイミングで判定を行い、結果を返す。以下、サーバへのアクセスと会話グループ判定について述べる。

サーバへのアクセスは HTTP 通信を用いる。また、得たい情報ごとにクエリを実装した。実装したものとしては”who?”と”group?”の二つであり、”who?”は会話場にいるメンバー、”group?”は会話グループの構成員と位置を示す環境ノードの情報を返す。また、クエリに含めるパラメータとして会話グループを検索したい時間や検索の判定にかける時間窓も設定した。パラメータについては会話グループの判定処理で詳細を述べる。

会話グループの判定処理の方法は、応用システムがアクセスしたタイミングで計算を行う。先ほど述べたクエリとパラメータを元に計算を行う。サーバには端末のペアと閾値が蓄積されており、パラメータである時間と時間窓にしたがって必要な情報を抽出する。具体的には検索したい時間から時間窓分遡った範囲を抽出する。そして抽出されたデータのうち、閾値 0.75 を超えた回数が多ければ会話グループ、そうでなければ会話グループではないとみなす多数決制で判定をする。また、計算に用いる検索時間や時間窓については応用システムに応じて使い分ける必要があるため、適切な時間窓を調べるために会話場の視覚化ツールも作成した (図 6)。この視覚化ツールでは人ノードを人型のアイコン、環境ノードを四角のアイコンで表現し、グルーピングを線でつなぐことで追体験しやすくした。

4.3　音環境比較による会話場検出の実例

以上で述べた音環境比較手法を用いた会話場検出の実例

© 2017 Information Processing Society of Japan

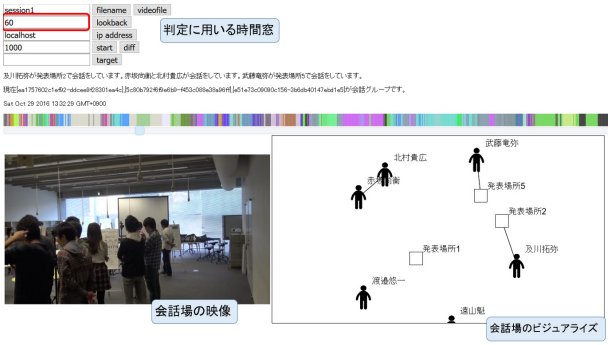

図 6　会話場のビジュアライザー
Fig. 6　Visualizing tool for method based on auditory similarity

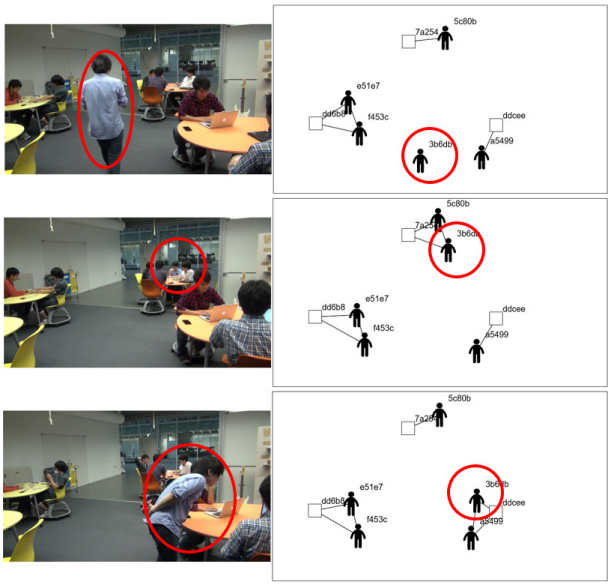

図 7　会話グループの判定結果：人の移動の変化
Fig. 7　Result of distinguish conversation groups and detect moving

を示す。

図 7 は会話場間を移動する様子を示したものである。図 7 中の左側は実際の映像、右側は音環境比較による手法で得られた判定結果を元に視覚化したグラフである。人型のアイコンが人ノード、四角が環境ノードをあらわしている。図 7 の例では会話グループが 3 つ存在しており、右側、左側、奥にそれぞれ 2、3 人が会話している。また、環境ノードがそれぞれの机の上に設置されており、人ノードは椅子に座っている人の一部が装着している。図 7 中の上段では歩いている男性が赤丸で囲われている。この赤丸で囲まれた男性は 3b6db という ID の端末を持っており、この ID と対応している。上段右の判定結果によるとこの段階ではどの会話グループにも属していない。次に、中段では奥の会話グループに参加しており、判定結果も環境ノード (7a254) とその付近で会話をしている参加者 (5c80b) とグルーピングされている。下段では右手前の会話グループへ移動し、会話をしており、判定結果も右手前テーブルの環境ノード (ddcee) とグルーピングされていることがわかる。このような移動の遷移も音環境比較によって検出することができる。また、環境ノードと人ノードが結びつくことで、誰がどのあたりで会話しているかといった情報を得ることができる。

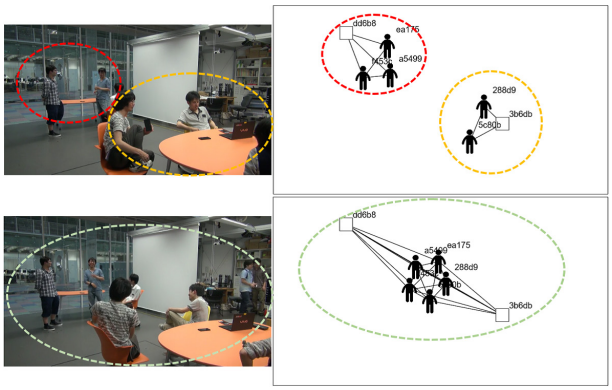

図 8　会話グループの判定結果：会話場の広がり
Fig. 8　Expand conversation field

図 8 は会話場が広がる様子を表したものである。図 8 の例では奥と手前に会話グループが計 2 つ存在している。また、机の上には環境ノードが設置されており、人ノードは装着している人としていない人が混じっている。上段は 2 つの会話グループが存在しており、判定結果も 2 つのグループが存在していることを示している。次に、下段は 2 つの会話グループ間で会話をはじめている様子で、判定結果も環境ノードを含めた大きな会話場であるとみなしている。以上のことから、会話場が統合するような状況も検出することができる。また、物理的に離れている環境ノード同士がグルーピングされることで大まかな会話場のサイズを認識できると考える。

5.　応用システム

応用システムについて述べる。応用システムとは、音環境比較による検出手法で得られた会話場情報を元に、サービスなどを行うシステムのことを指す。会話場情報へのアクセスについては 4.2.3 章で述べたとおりで、クエリや時間、時間窓を含めた HTTP 通信によってアクセスをする。提案する応用システムとしてはミーティングキャプチャと 3D 空間上での会話場可視化の 2 つである。

5.1　ミーティングキャプチャ

ミーティングキャプチャとは、会話場情報を用いたライフログシステムである。記録をしたい参加者がいる会話場を常に追跡し、追跡対象とグルーピングされた環境ノード付近に向けてカメラを動かすことで、より体験に近い議事録を取ることができると考える (図 9)。また、カメラは会話場に向けてフォーカスを動的に変えたいため、レンズを動かすことができ、かつズーム機能のあるものを選んだ。音環境比較による手法を用いることで、Hall の定義から漏

© 2017 Information Processing Society of Japan

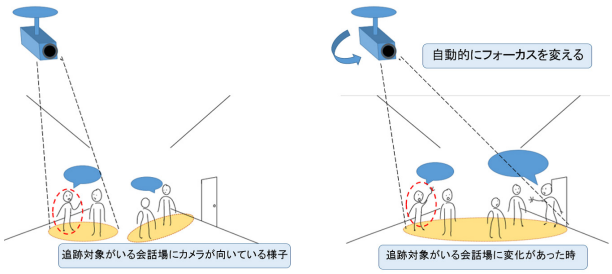

図 9 ミーティングキャプチャのアイディア
Fig. 9 Idea of meeting capture system

図 10 3D 空間への会話場の可視化
Fig. 10 Visualize conversation field in 3D virtual space

れるような状況下や、4.3 章で実例として挙げたような会話場間を移動する際や会話場が統合する場面でも記録できると考える。

5.2 3D 空間上での会話場可視化

2 つ目のシステムは 3D モデルで再現された仮想空間上に会話場の情報を可視化するシステムである。図 10 はシステムのイメージで、図 10 中の赤く発光物体が会話場を示すオブジェクトである。また、このオブジェクトの光の度合いによって会話場の規模を表現する。また、図 10 中の右側のように 3D モデルの建物が透過することで、建物内の会話場の位置を見渡すことができる。以上によって、本来構造上見えるはずのない会話場の存在や、規模、相対的な位置を知ることができる。本システムによって期待される効果としては建物内におけるコミュニケーションへの補助や機会創出の促しができると考える。例えば、建物の構造を知らない来訪者が、参加予定の講演や発表などの会場位置を大まかに把握したり、会話場の規模を把握することで、人が混雑している発表会場を避けて別の発表会場へ行くなどといった気づきの促しができると考える。また、会話場の発生が予想されるような状況だけではなく、突発的に発生した会話場や把握していなかった講演などのイベントに対する気づきを促すことができると考える。

6. まとめと今後の展望

本論文では、音環境比較による会話場検出についてと検出手法をベースとした応用システムの提案について述べた。今後の展望としては、ライフログ映像作成としての活用を考えている。理由としては、従来手法ではなく音環境比較でしか検出できないようなコンテキストが得られると考えているからである。音環境でしか得られないコンテキストの具体例としては、発表会や懇親会などの司会進行を聴いている際などが挙げられる。司会進行を聴いている際には同じ体験を共有していると考えるため、映像としては部屋全体を映すことが好ましいと考える。一方で、司会を聴いておらず、隣の人と会話をしているような状況下では共有していないと考えられる。このような状況を従来手法で捉えることは難しい。よって、音環境比較による手法を用いたライフログ要約システムは有用であると考える。今後は、以上の例のようなより直感に近いライフログ要約に取り組んでいきたい。

参考文献

[1] E. T. Hall. The Hidden Dimension. Doubleday & Company, Inc., 1966.
[2] R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock, “Meme tags and community mirrors: moving from conferences to collaboration,” in Proceedings of the 1998 ACM conference on Computer supported cooperative work, 1998, pp. 159168.
[3] T. Choudhury. Sensing and Modeling Human Networks. Doctoral thesis, Massachusetts Institute of Technology, September 2003.
[4] 森脇 紀彦, 佐藤 信夫, 脇坂 義博, 辻 聡美, 大久保 教父, 矢野 和男. 組織活動可視化システム「ビジネス顕微鏡」. The institute of electronics 2007-HCS-44(39), 2007-09, 一般社団法人電子情報処理学会.
[5] T. M. T. Do and D. Gatica-Perez, “Contextual grouping: discovering real-life interaction types from longitudinal bluetooth data,” in 2011 IEEE 12th International Conference on Mobile Data Management, 2011, vol. 1, pp. 256265.
[6] J. Rekimoto, T. Miyaki, and T. Ishizawa. LifeTag: WiFi-based Continuous Location Logging for Life Pattern Analysis. 3rd International Symposium on Location- and Context-Awareness (LOCA2007), pages pp.3549, 2007.
[7] Y. Sumi, J. Ito, and T. Nishida. PhotoChat: communication support system based on sharing photos and notes. CHI 2008 Extended Abstracts, pages pp.32373242, April 2008.
[8] B. Thiel, K. Kloch, and P. Lukowicz, “Sound-based proximity detection with mobile phones,” 2012, pp. 14.
[9] T. Nakakura, Y. Sumi, and T. Nishida, “Neary: conversation field detection based on similarity of auditory situation,” Special Section on emerging Technologies of Ubiquitous Computing Systems, No.6 June 2011, pp.1164-1172.
[10] 佐藤 弘之, 岩元 啓, 鈴木 誠, 森川 博之. CoHear:環境音を利用した近隣モバイル端末推定手法. 情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム] 2015-UBI-45(39), 1-6, 2015-02-23, 一般社団法人情報処理学会.
[11] W.-T. Tan, M. Baker, B. Lee, and R. Samadani, “The sound of silence,” in Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, 2013, p. 19.

© 2017 Information Processing Society of Japan