

音環境比較による会話場の検出と実世界接地

遠山 魁^{1,a)} 角 康之¹

概要: 会話グループの情報、つまり誰とどの程度の人数で話していたかなどの情報はライフログやその他の色々な方面での活用が期待される。コミュニケーション空間を検出する研究は人の立ち位置や距離によって得る方法が多く提唱されてきたが、相手や状況によって人同士の距離が異なるため距離によらない検出が必要であると考えられる。そこで音環境比較による検出方法を用いて、会話場の大まかな位置と会話空間の広がりを検出する実験を行った。

Detecting and Real-World Grounding of Conversation Group based on Situated Sound Similarity

KAI TOYAMA^{1,a)} YASUYUKI SUMI¹

Abstract: This paper proposes detecting conversation fields based on similarity of auditory situation among users. The distance between people, according to Edward Hall, differ greatly on situation.. Therefore we conducted a preliminary experiment. Experimental result shows method can distinguish groups of conversation in actual environment and track expanse of conversation field.

1. はじめに

本研究の目的は、展示会や博物館などといった不特定多数の人が行き交うオープンスペースにおける会話グループの発生・消滅といったダイナミクスを検出することである。

コミュニケーション空間を検出する研究としては人の立ち位置や距離などを手がかりに得る手法が多く提唱されてきた。具体的にはRFIDや赤外線タグ[1, 2, 3]などを利用することで近接関係を測定しグループを検出する方法やWiFi電波強度などを利用した測定[4, 5, 6]などである。しかし展示会などのオープンスペースでは部屋の大きさや人数、密度といったものが不特定である。また、Edward Hall[7]によるとスピーチなどといった公衆に向けられたコミュニケーション空間は公衆距離と呼ばれており、普段の会話における人同士の距離とは大きく異なる。このように相手や状況によって距離は大きく変わるため物理的距離から会話グループを得る従来の手法では困難である。そこで身体の向きや物理的距離によらない検出方法が求められる

と考えられる。

本研究では、近隣している人同士は同じような環境音下にいることが考えられるため、環境音を比較することで会話グループを検出する手法を用いる。会話グループ同士が密着しているような状況においても正しく判別できると考えられる。また、場内に向けられたスピーチなどの大きな音に対してはひとつの大きな会話グループとして検出されることも予想される。

2. 関連研究

2.1 先行研究

環境音による会話グループを検出する研究として中蔵らのNearyがある[8]。中蔵らは同じ会話グループは似たような環境音にさらされているという仮定の下、環境音を比較することで会話グループを検出する方法を提唱した。具体的にはモバイル端末を用いてユーザーの環境音計測、計測された音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毎のスペクトルをベクトルと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を超えた場合、会話グループとみ

¹ はこだて未来大学
Future University Hakodate
^{a)} k-toyama@sumilab.org

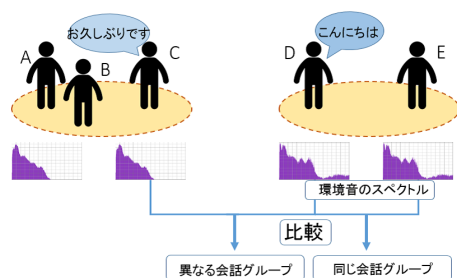


図 1 アルゴリズムのアイデア

Fig. 1 Basic idea of algorithm

なし検出している。本研究は中蔵らの Neary のアルゴリズムを参考に会話場の実空間への対応付けと会話場の空間サイズの推定に取り組む。

2.2 環境音比較による近隣推定

本研究と同じく環境音を用いた研究としては佐藤らの CoHear が挙げられる [9]。CoHear は Neary や本研究同様に声や物音などといった非定常音、エアコンなどの場所特有の定常音に着目することで近隣推定を行う研究である。また、物理的距離などによる近隣推定に比べて誤検出が少ないことや、スケーラビリティや省電力、プライバシーなどといった観点からも環境音による推定の利点を挙げている。

3. 音環境による検出手法

提案する手法は、図 1 に示すように近隣する人は同じ環境音にさらされているという仮定の下、環境音を比較することで会話グループであるかどうかを調べるというものである。具体的にはユーザーすべての組み合わせの環境音を比較し、得られた類似度が閾値を超えていたらそのペアは会話グループであると判定する。

4. 実環境における会話場検出の実験

展示会、博物館などといったオープンスペースにおいては不特定多数の人が行き来し、会話グループが至る所で生成・消滅を繰り返すことが予想される。そこで、予備検討として実環境においても会話グループの検出が正しく行うことができるかを確認した。2015 年 9 月 7 日から 11 日にかけて行われた Ubicomp で実験を行った。人数は 5 人とし、各々が持つ携帯端末のマイクで環境音を計測した。また、携帯端末は NEXUS6 を用いた。携帯端末で声を含む環境音を録音した後、音声データをオフライン上で高速フーリエ変換し、コサイン類似度を求めた。なお、会話グループであるか判別するための閾値は中蔵らの先行研究に倣い 0.775 とした。

得られた類似度から会話グループのダイナミクスを確認するためにばねモデルを作成した。図 2 に記すとおり、ば

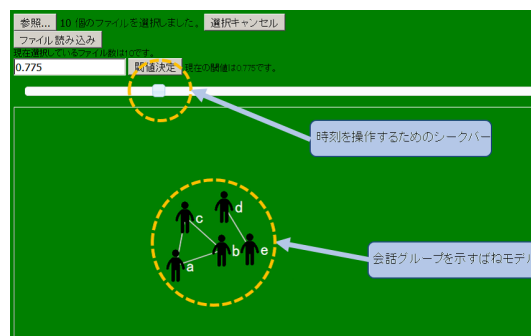


図 2 検出された会話場と参加メンバーの可視化

Fig. 2 Force-Directed Graph for visualizing conversation groups and conversation field

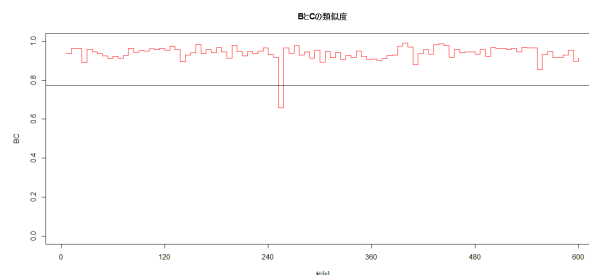


図 3 ユーザー B、C の音環境の比較

Fig. 3 Comparison between userB and userC of frequency spectrum

ねモデルはシークバーで時間を操作できるようにし、どの時間帯に誰が会話グループとみなされたかを振り返りやすくした。以下に音環境による検出が有意であることを示す事例をいくつか抜粋し紹介する。

4.1 シーン 1

シーン 1 は 9 月 9 日の 9 時 30 分ごろに計測されたもので、開会式が行われている。ユーザー A、B、C、D、E は録音開始時に一度集まり、その後は各々の好きなように行動した。ユーザー B と C は始終共に行動しており、計測から開会式が始まるまでの約 7 分間会話をし続けていた。また、ユーザー E のみは録音開始 6 分頃から他の 4 人とは別の部屋へ移動し、開会式のスピーチを聞いていた。

図 3 はユーザー B と C の音環境を比較したグラフである。グラフが示すとおりユーザー B と C は録音開始時から終了時まで閾値を大きく超える類似度を示しており、ユーザー B と C が同じ会話場にいたことを示している。

図 4 はユーザー E と他のすべてのユーザーとの類似度を比較したものである。図に記したとおり 6 分頃以降からのユーザーもユーザー E に対するの類似度はすべて閾値を下回っていた。また、開会式のスピーチや拍手などといった音も他のユーザーとは大きく異なっていたため、全く別の部屋にいたと考えられる。

以上のことから、人が行き来したり多数会話グループが

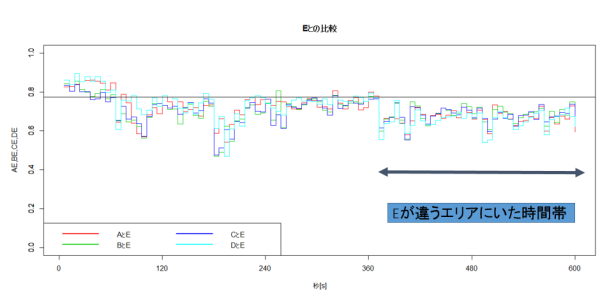


図 4 ユーザー E と他のユーザーの音環境の比較

Fig. 4 Comparison between userE and another of frequency spectrum

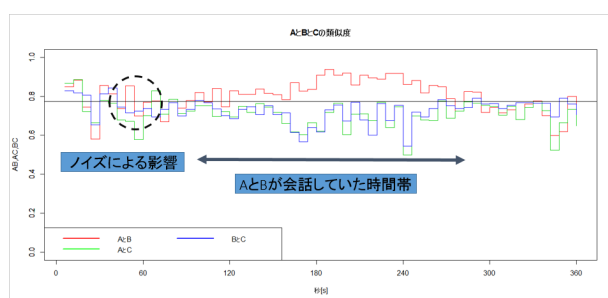


図 5 ユーザー A、B、C の音環境の比較

Fig. 5 Comparison between userA, userB and userC of frequency spectrum

発生するような会場内であっても会話をしていたユーザー B と C の会話グループを正しく検出することができた。一方で異なるエリアに移動したユーザー E が他のユーザーとグルーピングされないことも確認できた。

4.2 シーン 2

シーン 2 は 9 月 11 日の 14 時 30 分ごろに計測されたものである。ユーザー A、B、C が録音開始時に集まった後、C のみ別行動をし、A と B は録音開始後から 5 分間ほど共に行動した後散会した。

図 5 はユーザー A、B、C の音環境を比較したグラフである。グラフ上の赤線は A と B の比較、緑線は A と C、青線は B と C の比較を示している。A と B の類似度を示す赤線が閾値を大きく超え、一方で A と C、B と C の類似度は閾値を下回っている。つまり A と B は同じ会話場にいたが、C は違うエリアにいたということが読み取れる。

結果としては 3 人の大まかな行動をグラフから見て取ることができた。また、実環境においても会話しているユーザー同士の声の特徴となり会話グループを検出できていることがわかる。また、携帯端末をポケットにしまう際にノイズが発生し判別が不安定になっている箇所があった。

4.3 距離に依存しない検出

図 6 は 9 月 9 日の 19 時 30 分ごろの A と B の音環境を比較したグラフである。A と B は物理的な距離は非常に

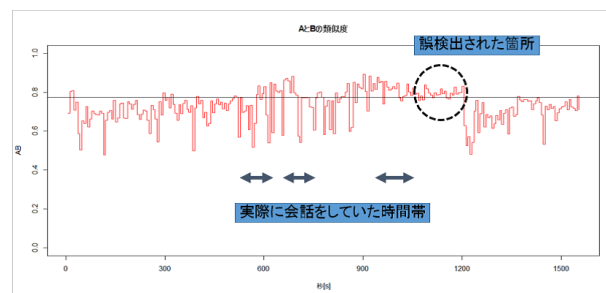


図 6 ユーザー A、B の音環境の比較

Fig. 6 Comparison between userA, userB and userC of frequency spectrum

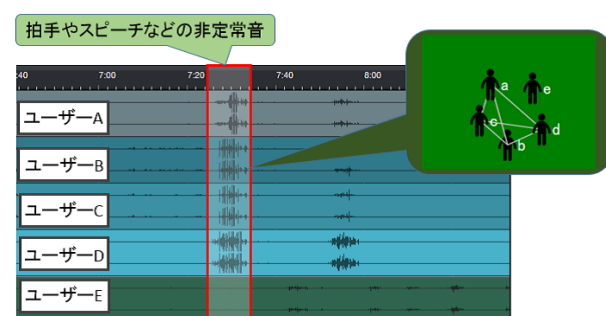


図 7 拍手やスピーチなどの非定常音

Fig. 7 Non-stationary sound by handclap and speech

近くにいるがお互いに別の会話グループに属していた。また、会話の合間に A と B は軽く会話を交わしていた。具体的には録音開始から 10 分後前後や 11 分 40 秒後前後、16 分後前後あたりである。

グラフからは会話グループが生成されたり消滅したりを繰り返していることが読み取ることができ、物理的な距離が近くても正しく判別できていることがわかった。一方で、18 分から 20 分後の間は実際には会話をしていなかったにもかかわらず会話グループとして検出されている。2 人は実際に会話を行っていなかったが同じ環境音下にさらされていたためであると考えられる。

4.4 エリア全体が会話場と検出される例

次に、エリア全体が会話場と検出される例を挙げる。図 7 は拍手やスピーチなどの音が起きた場合の検出状況を記している。拍手やスピーチなどといった部屋全体に向けられた環境音にさらされている場合、ユーザーは同一の会話グループとみなされた。

4.5 実験まとめ

結果としては実環境においても誰と、どれほどの時間、会話あるいは一緒に行動していたかなどの大まかな行動を見て取ることができた。また、物理的な距離が近くても各々のマイクに計測される音の違いによって会話グループを判別できていた。その一方で場内全体に向けられたアナ

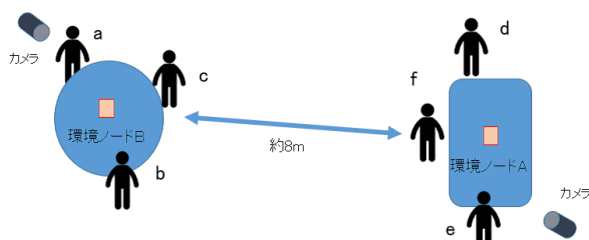


図 8 会話場検出の実験環境

Fig. 8 Experiment setting

ウンスやスピーチによってひとつの会話グループとしてみなされることも確認できた。

5. 会話グループの位置や会話空間の大きさを検出する実験

会話グループの検出だけではなく大まかな位置や会話空間の大きさに着目し、それらを推定するために普段研究活動を行っているオープンスペースで実験を行った。

図 8 に実験を行った環境について記す。端末は計 8 台用意し、全て NEXUS5 で揃えた。そのうち 6 台は実験協力者 6 人が持ち歩き、2 台の端末を机の上に固定した (以下、環境ノードと呼ぶ)。環境ノードを用いた意図としては次の通りである。

- 環境内の特定の位置に端末を設置することで会話グループの大まかな位置を確認したい。
- 環境ノード同士が結びつくことで大きな会話空間が形成されることを確認したい。

実験中には会話シーンが大きく変化した場面が 3 回あった。

シーン 1 実験開始時はそれぞれの環境ノードを中心に 3 人 1 組の 2 つの会話グループが発生していた。

シーン 2 実験開始時から 50 秒後に実験協力者 a が環境ノード A 付近にいる別の会話グループへ声をかけた。

シーン 3 70 秒後には環境ノード A 付近にいた会話グループが呼びかけに応じて環境ノード B にいる会話グループに加わり一つの会話グループとなった。実験終了時まで 6 人で会話をしていた。

5.1 会話グループのバインド

図 9 では環境ノード A 付近で会話をしている実験協力者 d, e, f と環境ノード B 付近で会話をしている実験協力者 a, b, c によって形成された二つの会話グループを検出したことを示している。環境ノードを用いることで会話グループが空間内のどの辺りに位置しているかを確認することができた。

5.2 会話場の広がり

シーン 2 では環境ノード B 付近で会話をしていた実験

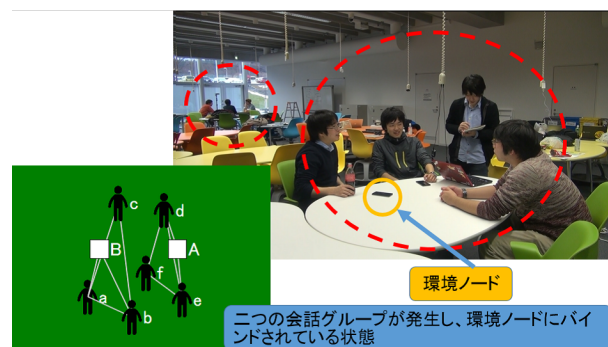


図 9 会話グループと環境ノードにバインディングされている様子

Fig. 9 Binding conversation groups

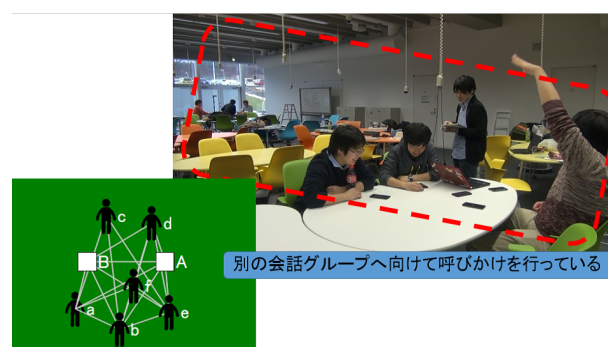


図 10 会話場が広がっている様子

Fig. 10 Expanse of conversation field

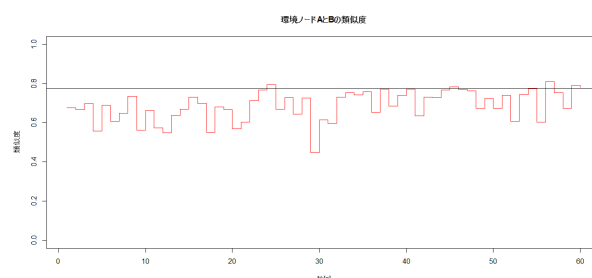


図 11 環境ノード A と B の音環境の類似度

Fig. 11 Comparison between nodeA and nodeB of frequency spectrum

協力者 a が別の会話グループに向かって呼びかけをしていた (図 10)。離れた位置に存在していた会話グループ、環境ノード同士が結びついているのがばねモデルから確認できた。呼びかけによって一時的に大きな会話グループが形成されていることがわかる。

5.3 会話の盛り上がりによる会話空間の広がり

シーン 3 では呼びかけに応じて環境ノード A 付近にいた 3 人の実験協力者が環境ノード A に集まり会話をしていた。

環境ノード A と B 音環境の類似度を示したグラフを図 11 に示す。実験協力者 a による呼びかけだけではなく、環境ノード A と B が同じ会話グループと判断されている箇

所が確認できる。シーン3では一つの大きな会話グループとなり多人数で会話が行われ盛り上がっていたことから、離れた位置にある環境ノード同士を巻き込むほどの大きな会話空間が形成されたと考えられる。

5.4 実験まとめ

環境ノードを設置することで会話グループの大まかな位置を確認することができた。また、呼びかけなどといった意図的に空間全体に向けられた音だけではなく、会話が盛り上がることによって会話空間が広がるといったことも確認できた。

6. 今後の課題

今回の実験では実環境における会話グループの検出や会話グループの大まかな位置、会話空間の広がりを検出することができた。今後はそれらの会話状況を同じオープンスペースに存在するロボットに理解させることで適切な社会的インタラクションを可能にさせるといったことを考えている。今回はあらかじめ録音した音声データを用いて予備検討を行ったが、ロボットに応用するにあたってリアルタイムで会話グループの検出を行う必要がある。そのため今後は、ユーザー各々の環境音を取得し、周波数特性を得るためのスマートフォンのアプリケーションとサーバー上で会話場の判定と会話状況を判断するシステムを実装していく。

参考文献

- [1] 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 CSCW'98*, pages 159168. ACM, 1998.
- [2] T. Choudhury. Sensing and Modeling Human Networks. Doctoral thesis, Massachusetts Institute of Technology, September 2003.
- [3] D. Wyatt, T. Choudhury, J. Bilmes, and J. Kitts. Towards Automated Social Analysis of Situated Speech Data. *Proceedings of Ubicomp 2008*, Spetember 2008.
- [4] J. Hong, G. Borriello, J. Landay, D. McDonald, B. Schilit, and D. Tygar. Privacy and Security in the Location-enhanced World Wide Web. In *Proceedings of Ubicomp 2003*, October 2003.
- [5] 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.
- [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] E. T. Hall. *The Hidden Dimension*. Doubleday & Company, Inc., 1966.
- [8] Toshiya Nakakura, Yasuyuki Sumi, Toyoaki Nishida. Neary: conversation field detection based on similarity of auditory situation. *HotMobile '09 Proceedings of the 10th workshop on Mobile Computing Systems and Ap-*

plications, Article No. 14, 2009.

- [9] 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.