

会話の選択的聴取を支援する音声チャットシステム

神保周児¹ 亀井清華² 藤田 聡³

広島大学 大学院工学研究科 情報工学専攻

1. はじめに

バンケットやパーティなどの多くの人が集まる場では、様々な話題に関する会話が並行して進められる。各参加者は、聞こえてくる会話の中から興味のある話題を聞き分け、話者の近くに移動するなどしてその会話に自然に参加する。我々は、そのような話題の聞き分けと会話への参加を効果的に支援する音声チャットシステムを Android 端末上に実装し、その基本性能の評価を行った。実装システムは、各話者から送られてくる音声ストリームを興味の近さや物理的な距離によって重み付けする部分と、重み付けされた音声ストリームを自然な形で重ね合わせる部分とからなる。この手法を既存の音声チャットシステムである Skype などに組み込むことで、たとえば数十人規模で実施されるクラス会を音声チャットで実現することが可能となる。

2. 関連研究

大橋ら[1]は、情報共有型のリアルタイムコミュニケーションシステムを開発した。彼らが提唱するコミュニケーション空間である「場」には複数の話題が存在しており、それらの話題に対して 3D サウンドシステムを使って距離と方向性を付与することで、現実世界のような複数の話題が並存する環境が実現される。異なる話題を識別する際には、注目する音源を聞き取りやすい位置にまでユーザが移動し、視覚情報を用いて音源定位をおこなうというアプローチがとられている。本稿で用いる手法はそれとは異なり、音量の大小のみによって近さを判断するという方法がとられている。そのような初歩的な手法によってもある程度の会話の識別が可能であることが、4節の評価結果から示される。

3. 提案手法

提案手法の基本方針は以下の通りである。

A voice chat system supporting selective hearing of conversation

1 Shuji JINBO, Hiroshima University

2 Sayaka KAMEI, Hiroshima University

3 Satoshi FUJITA, Hiroshima University



図1 興味ベクトルの入力画面

1. 物理的な近さを音量に反映させること
2. 興味の近さを音量に反映させること

3.1 距離の取得

ユーザ間の物理距離は iBeacon[2]を用いて取得する。具体的には、ID の発信範囲を Far(半径 10m)に設定した iBeacon デバイスを対象となる空間内にあらかじめ規則的に配置しておき、端末が受信した iBeacon の ID によってその端末の現在の座標値を推定する。位置推定手法の詳細については[3]を参照されたい。以下の議論では、各端末はそれぞれ自分の現在位置を取得しているものと仮定する。また音声ストリームの送信者と受信者の間の(座標空間上の)距離を D と記すことにする。 D は音声聞こえる限界地点までの距離が 1 となるように対数的に正規化された値であり、 $0 \leq D \leq 1$ を満たしている。

各ユーザの興味は、あらかじめ設定された複数の項目に対する五段階評価により取得する。項目数と項目の中身は、使用目的によって変更可能である。図1の例では、科学、映画、漫画・ゲーム、政治の四項目が示されている。五段階評価によって得られた興味ベクトルの間の距離を興味の近さと定義する。以下では、音声ストリームの送信者と受信者の興味の近さを S と記す。前述の D と同様、 S も $0 \leq S \leq 1$ を満たすように正規化された値である。

なお実装システムでは、各ユーザの座標値と興味ベクトルの値は、更新されるたびにシステムに参加しているすべてのユーザに通知される。

またユーザ間の距離は、重みパラメータ γ をもちいて $(1-\gamma)D+\gamma S$ のように定義され、各話者の音量の強さは、上述の距離の増加に伴って線形に減少するよう設定されている。

3.2 複数音声の合成と再生

音声ストリームは byte 配列として送信される。複数の端末から送られてくる byte 配列 X を受信した端末は、送信者ごとに計算される音量を受信した配列 X に乗じ、その結果を サイクリックキューである buffer 配列 Y に書き込む。音声の再生は、配列 Y から音声を順次取り出して再生 API に書き込むことで実現する。ここで音声の合成と再生が同一の buffer 配列 Y 上でおこなわれるため、配列上で合成がなされる位置と再生がなされる位置を少し離しておく必要があることに注意しよう。実際、それらの位置が近すぎると再生に書き込みが追いつかなくなり、音声の途切れが発生してしまう場合がある。

4. 評価

実験方法の概要は以下の通りである。今回の実験では、興味ベクトルに関する評価をおこなった。被験者数は 20 名で、実験は 4 人一組でおこなう。まず 4 名の被験者にジャンルに少しずつ重なりのある別々の台本を渡す。実験の 1 ラウンドでは 1 名が聞き役になり、聞き役はラウンドごとにローテーションし合計 4 ラウンド実行する。各話者の興味ベクトルは台本ごとにあらかじめ固定するが、聞き役の興味ベクトルは被験者の興味に基づいて自由に設定してもらう。各ラウンドで聞き役は以下の点を評価する。

- 各話者の音量の大小（大・中・小の三段階）
- 各話者がそれぞれ何の話をしてたか

これにより、興味の近さが知覚される音量にどのように反映されたかという点と、複数の話題の中から自分の興味にあった話題がちゃんと聞き取れたかどうか確かめられる。

4.1 音量の反映

はじめに話者との興味の近さが聞き役の知覚する音量の違いに意図通りに反映されているかを評価した。結果を図 2 に示す。この結果から、興味をもっとも近い話者の音量がもっとも大きいことは 100% の精度で知覚できたが、それ以外の話者の音量の知覚は、60% の精度でなされたことがわかる。

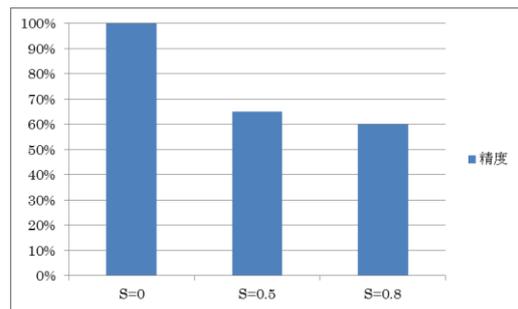


図 2 興味の近さによる音量の反映。

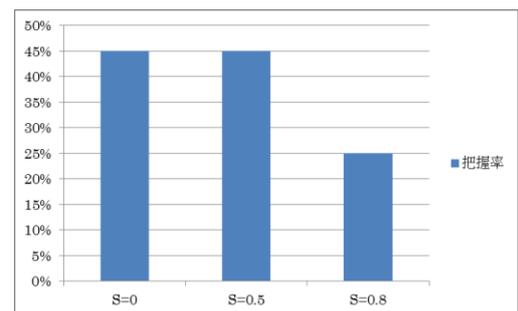


図 3 興味の近さによる会話内容の把握。

4.2 会話内容の把握

次に複数の会話の内容がどの程度正確に把握されるのかを評価した。聞き役には会話の内容をあらかじめ用意した選択肢の中から選択する形で答えてもらった。結果を図 3 に示す。この結果から、同時に再生される三つの会話のうち、半数弱の会話の内容が把握できたことがわかる。ただし興味のない会話の把握率はそれ以外の会話の約半分であり、興味のある会話がうまくハイライトされていることがわかる。

5. おわりに

本稿では、会話の選択的聴取を支援する音声チャットシステムについて述べた。実装したシステムを用いた実験の結果、興味の違いが音量の違いに反映され、興味のある会話がハイライトされることが確認された。今後の課題として、物理距離を音量に反映させることの効果を確認することと、より詳細な評価方法を確立させることがあげられる。

参考文献

- [1] 大橋, 広淵, 河合, 藤川, 砂原: 視覚情報により強化された 3D サウンド場による共有型複数音声チャットシステムの設計と実装, 情処研報, DPS, 227-232, 2006.
- [2] iBeacon, <https://developer.apple.com/ibeacon/>
- [3] 神崎僚太: iBeacon モジュールを利用した 2 次元位置座標推定の考察, 第 10 回情報科学ワークショップ, 2014.