

「ざわざわ」とした音環境の再現を目指した 音声の同時聴取能力に関する研究

池田輝政[†] 遠藤正隆[‡] 中嶋裕一[‡] 松井瑠偉人[‡] 菱田隆彰[†]
愛知工業大学[†] 株式会社リオ[‡]

1. はじめに

コロナ禍を端緒に本格的なテレワークを導入する企業が増加し、社内における人間関係の構築や維持を、他者と実際に対面することなく行う必要が出てきた。そのような状況下で人間関係が希薄化する一つの要因として、インフォーマルコミュニケーションの減少が挙げられる。従来なら食堂や休憩室などで発生していたであろう雑談などのコミュニケーションは、テレワーク環境下では発生しづらい。遠隔会議システムの普及により多人数間での音声や動画を用いたコミュニケーションが容易になったが、会議を想定しているが故に時間や参加者を限定して使用する形式となる。したがって、偶発的に起こることが期待されるインフォーマルコミュニケーションの場として最適とは言えない。テレワーク環境下でもインフォーマルコミュニケーションを喚起するような新しいシステムを提供する必要がある。

2. 「ざわざわ」とした音環境の再現

コミュニケーションの場となる「空間」の提供に着目したシステムの一例として、小川ら[1]によって提唱された Commidor がある。これは共有空間である「廊下」を仮想的に移動するイベントを発生させ、そこで他者との偶発的なエンカウントを誘発するという試みである。

一方、我々はコミュニケーションの場における「音環境」に着目する。現実の食堂や休憩室では、少人数による会話が同時多発的に発生している場合が多い。つまり、自分が会話している対象の音声以外に、他者による会話の音声と同時に聞こえている状態となる。我々はこのような所謂「ざわざわ」とした音環境を再現することが、インフォーマルコミュニケーションを

喚起する場において重要なのではないかと考える。また、辻村ら[2]の研究において、適切な音量の会話雑音が会話のしやすさに好影響を与えることが指摘されている。このような場では往々にして、噂話に聞き耳を立てる、偶然聞こえたワードにより話題が変わる、など他者からもたらされる情報を意識的、無意識的に取得する行為が行われている。これらも一つのインフォーマルコミュニケーションと捉えるならば、その環境を再現することには意義がある。

人間はこのような「ざわざわ」した音環境下でも、特定の音に意識を集中して他の音を見捨てるような聴取の仕方が可能である。これを「カクテルパーティ効果[3]」という。カクテルパーティ効果には様々な要素が関与するが、特に音源の方向による影響が大きい。したがって、音声の定位を適切に調節することで、特定の音声を聞きやすくしたり、逆に無視しやすくしたりすることが出来る可能性がある。

本研究では、「ざわざわ」とした音環境を提供出来るオンラインコミュニケーションシステムの実現を目指す。本稿ではその前段階として、複数の音声を同時聴取する際に定位や音量が聞き取りやすさに与える影響を調査し、その結果について検討する。

3. 実験の概要

実験の基本的な流れは、まず被験者にターゲットとなる人物の音声を記憶してもらう。次に質問文を提示し、それに対する応答文を読み上げる音声を複数同時に再生する。被験者にはその中からターゲットの応答内容を聞き取ってもらうが、聞き取れたと思えるまでは何度でも聞き直しが出来る。最後に応答文のリストを提示して回答してもらう。この回答の正誤、及び聞き直し回数を、音声の同時聴取における聞き取りやすさの判断材料とする。

今回は1つの質問文につき4つの応答文を用意した。応答用の音声は「VOICEVOX[4]*」上で男性（低音、高音）、女性（低音、高音）の4種類の

A study on simultaneous listening of multiple voices for reproducing "buzzing" sound environment

[†] Terumasa Ikeda, [‡] Masataka Endo, [‡] Yuichi Nakashima,

[‡] Ruito Matsui, [†] Takaaki Hishida

[†] Aichi Institute of Technology

[‡] RIO CORPORATION

音声を設定し、各応答文を読ませた WAV ファイルを生成した。応答する音声の数と種類、各音声の音量と定位の組み合わせを「シーン」と定義し、これを複数用意することで様々なシチュエーションでの聞き取りやすさを調査する。

同じターゲットの音声を聞き続けることで記憶が強化され、出現順が後のシーンほど聞き取りやすくなる可能性がある。その対策として、シーンの出現順はランダムになるようにした。また、各シーンで使用する質問文、応答文の割り当てもランダムにして、文の内容による聞き取りやすさの偏りが出にくくなるよう配慮した。

4. 実験結果と考察

実験はテレワーク環境下を想定してイヤホンによる聴取で行うこととし、我々が所属する研究室のメンバーを被験者として2回実施した。初回は20名、2回目は19名が参加し、合計23シーンの結果を取得した。これらの結果から様々な知見を得たが、本稿では特徴的だった結果について述べる。

ターゲットの定位による聞き取りやすさの変化を調べるべく、図1で示した5つのシーンを用意した。定位は左から中心を通過して右までの割合を-1.0から1.0までの数値として表現する。定位が等間隔になるように4声を並べ、シーンごとにターゲットの位置を変更した。その結果が表1である。

ターゲットが中央に配置されたシーンCの平均再生数が最も少なく誤答も無かったのに対して、ターゲットが中央から外れるシーンでは誤答する被験者が現れた。また、ターゲットが完全に片側に振られたシーンAとEでは平均再生数が顕著に増えている。イヤホンでの聴取では反響音が無いため、この場合の音声は片耳からしか聞こえてこなくなる。それが聞き取りやすさに大きく影響を与えたものと推察される。

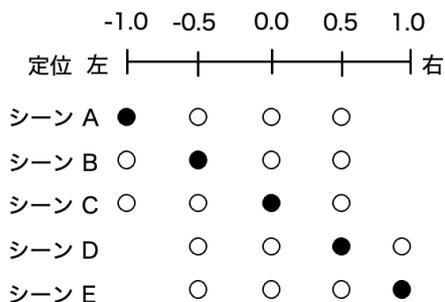


図1 各シーンの音声定位とターゲット（黒丸）

* VOICEVOX: 玄野武宏, VOICEVOX: 青山龍星, VOICEVOX: 波音リツ, VOICEVOX: 四国めたん

表1 ターゲットの定位と聞き取りやすさの関係

| シーン | 平均再生数 (回) | 誤答率 (%) |
|-----|-----------|---------|
| A | 2.747 | 10.526 |
| B | 1.947 | 21.053 |
| C | 1.421 | 0.000 |
| D | 1.842 | 10.526 |
| E | 2.737 | 21.053 |

これらの結果から、音声は複数同時再生されている状態でも、ターゲットの定位を中央に設定することでカクテルパーティ効果を補強出来ることがわかる。また、非ターゲットの音声を左右どちらかに完全に振ってしまうことで、無視しやすくなる可能性がある。

5. まとめ

テレワーク環境下におけるインフォーマルコミュニケーションの喚起を目的に「ざわざわ」とした音環境を提供出来る新たなオンラインコミュニケーションシステムの実現を目指し、音声の同時聴取能力についての調査と結果の検討を行った。音量や定位がカクテルパーティ効果に及ぼす影響について確認することが出来た。特に、完全に片側に定位を振ることが聞き取りやすさに大きな影響を与えることがわかったのは収穫であった。

今後の展望として、この結果を踏まえて音声の定位や音量をリアルタイムにコントロールして「ざわざわ」とした音環境を提供するシステムの製作に取り掛かる。このシステムが実用に足るものになれば、テレワーク環境下におけるQOLの向上にも寄与出来る。

参考文献

[1] 小川和也, 高島健太郎, 西本一志: Commidor: テレワーク状況におけるインフォーマルコミュニケーションを誘発する仮想廊下, 情報処理学会研究報告. グループウェアとネットワークサービス, 2021-GN-113 (9), pp. 1-7 (2021)

[2] 辻村壮平, 秋田剛, 小島隆矢, 佐野奈緒子: 複数人による知識創造活動を行う会議に及ぼす室内音環境の影響, 日本建築学会環境系論文集, Vol. 80, No. 711, pp. 397-405 (2015)

[3] 大串健吾: 音響聴覚心理学, pp. 256-259 (2019)

[4] “VOICEVOX”, <https://voicevox.hiroshiba.jp> (参照 2022-12-22)