

音声認識文に基づく複数効果音の動的立体構成システムの提案

情 17-0051 氏平 聖力

指導教員 米澤 朋子

1. はじめに

近年、音声認識や立体音響が、エンターテインメントやスマートフォン、パソコンのアプリなどで、幅広く活用されている。特に立体音響は、テーマパークのアトラクションなど、娯楽中心に使用されており、バイノーラル録音やサラウンド方式などで臨場感を高め、まるでその場にいるかのような体験を提供する[1]。

本研究では、音声認識と立体音響を組み合わせ、絵本や小説などの文を音読した際の声が音声認識で認識し、その文に合った複数の効果音を立体音響として聴かせるシステムを提案する。立体音響を使用することで、架空の世界に対する臨場感を高め、視覚障害者にも楽しめるものになると考えた。ステレオカメラから取得した画像をもとに推定した距離に応じ、立体音響出力するシステム[2]は、視聴覚変換を利用して音像風景を提示するものである。音像風景（サウンドスケープ[3]）を作り出す際に、文章に応じた立体的な風景を感じさせるため、本研究では複数の効果音の定位を文章内容から取得し、立体音響にすることを旨とする。

2. 提案システム

本システムのハードウェアは、立体音響を聴くためのイヤホンまたはヘッドホン、立体音響を生成するためのPC、音声を入力するためのマイクから構成され、ソフトウェアは、Visual Studio 上でC言語を使い、その中でOpenAL（立体音響生成ライブラリ）を使用した。音声認識にはSAPI5を使用し、形態素解析にはMecabを使用した。発せられた文を音声認識し、その分を形態素解析して効果音データベースの中からその文に応じた効果音を検索し、出力する。出力される効果音は、文の中に位置が入っている場合、その位置に効果音が割り当てられ、立体音響となって出力される。フロー図を図1に示す。

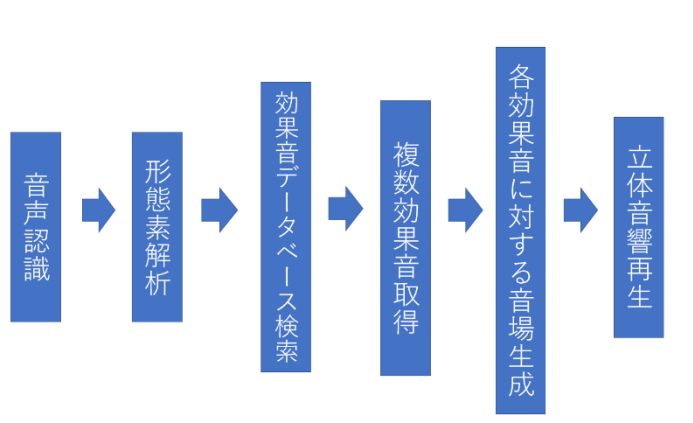


図 1：システム フロー図

3. 検証

指定した文章をユーザが入力し、その文章に応じた複数の効果音を 18-23 歳の男女 25 名の実験参加者に聴かせシステムの効果を検証した。1) 入力インタフェース要因（音声、テキスト）および効果音数要因（1, 3, 5）の 2 要因被験者内実験計画を実施し、臨場感の有無と操作性の良し悪し、システムとの対話性を評価した。その結果、音声入力の方が操作性が向上し、効果音が複数の場合は音が聴きにくく疲労感を感じるという結果であった。臨場感や対話性に関しては有意差が見られなかった。2) 効果音の時系列上配置要因（直列：音が順番に流れる、並列：同時に流れる）および立体音響要因（各効果音の音場の有無）の 2 要因被験者内計画を実施し、音の聴きやすさ、臨場感の有無、システムへの没入感を評価した。その結果、並列より直列の方が、音が聴きやすさ、臨場感が向上した。また立体音響の方が、没入感、臨場感が向上した。また、臨場感、没入感についての交互作用において、特に時系列上配置要因に関係なく、立体音響であった場合は、臨場感が向上し、立体音響であるかに関係なく、時系列上配置要因が直列の場合、没入感が向上した。

4. おわりに

本研究は、絵本、小説などの文章を音読するユーザの声を音声認識で認識し、その文に応じた効果音を立体音響として読者に聴かせることで、物語の臨場感を向上させるシステムを提案した。検証の結果、入力インタフェースと効果音の数によって、操作性、音の聴きやすさが向上した。また、音を順々に提示することや立体音響によって、音の聴きやすさ、没入感、臨場感の向上が確認された。

今後の展望として、絵本や小説などに限らず、視覚障害者に向けての臨場感会話支援システムの展開や、文章の内容を VR で可視化するなど、さらに楽しめるエンターテインメントへの拡張も検討していきたい。

参考文献

- [1] 浜田. バイノーラル音場再生系について, 日本音響学会誌, Vol.48, No.4, pp.250-257, 1992
- [2] 鈴木, 濱村, 菅原, 加藤, 大倉. 立体音を用いた視覚障害者用水族館鑑賞システムの開発, 人間工学, Vol.55, No., pp.1C3-3, 2019
- [3] 鳥越. サウンドスケープとはなにか, 環境技術, Vol.19, No.7, pp.409-411, 1990