

令和 4 年 6 月 2 日現在

機関番号：33919

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K12193

研究課題名(和文)自由聴点音響信号処理にもとづく音響クロスリアリティに関する研究

研究課題名(英文)A study of acoustic cross-reality based on selective listening point audio

研究代表者

西野 隆典(NISHINO, Takanori)

名城大学・都市情報学部・教授

研究者番号：40329769

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本課題では、ユーザが所望する音環境や、ユーザに聞いて欲しい音環境の立体音響再生を主題とした、音響技術によるクロスリアリティの実現に向けた検討を行った。快適さの観点での音の調査と加工の検討では、不要と感じられる音についての大規模調査を実施するとともに、自己の音声を違和感なく聴取するための加工技術を提案した。また、人間と同程度に音源定位を行う音源定位モデルの検討では、学習と評価で異なる音響信号を用いた検討を行い、性能評価を行うとともに課題整理を行った。

研究成果の学術的意義や社会的意義

本課題は、立体音響を活用した音響情報提示の基礎検討を行ったものである。ユーザに違和感を抱かせないような音響提示を行うための検討からは、世代間の差を明らかにするとともに、既存の音声処理技術を用いても有効な音声加工が可能であることを確認した。これらの結果から、違和感が少ない音響提示が可能であることが示唆された。また、深層学習応用から、人間と同様の音源定位を行うために必要なモデルや学習方法についての課題について整理することができた。本課題の成果は、仮想現実技術やロボット技術において活用できると考える。

研究成果の概要(英文)：This study investigated a realization of acoustic cross reality such as virtual reality, augmented reality, and mixed reality, with the subject of three-dimensional acoustic reproduction of the sound environment desired by the user and provided to the user. From the viewpoint of providing a comfortable sound environment, a large-scale survey of discomfort environmental sounds was conducted, and a method of converting one's recorded voice to one's own heard voice was also proposed. In addition, a sound source localization model based on deep learning was examined. The goal of this sound localization model will achieve its performance to the same degree as humans. The experiments using different acoustic signals for training and evaluation were conducted and revealed its performance and some problems.

研究分野：音響工学

キーワード：音響クロスリアリティ 立体音響 頭部伝達関数 深層学習応用 快音・不快音

1. 研究開始当初の背景

立体音響技術は、任意の音響空間の再現にとどまらず、仮想空間におけるコミュニケーションへの応用や、実環境との融合による新たな情報提示技術として期待されている。音響空間の再現や、情報提示技術として用いるための課題として、どのような音情報をユーザに提示するか、また生成した空間情報がどのような精度でユーザに提示できるかが挙げられる。どのような音情報を提供するかについては、ユーザがどのような音に関心を持っているか、ユーザの視線情報を活用する試みがなされている[1]。また、精度を測るために被験者実験が行われるが、被験者の負担軽減も必要であると考えられる。

2. 研究の目的

本研究の目的は、ユーザが所望する音環境や、ユーザに聞いて欲しい音環境を立体音響再生するシステム構築のために、音源分離技術により得られた分離信号をユーザの嗜好に合わせて統合する方法と、統合された立体音響信号の質を評価する手法を検討することである。ユーザの嗜好に合わせた音を提示するためには、ユーザの行動や嗜好の情報をどのように取得し、システムに反映させるかが課題となる。また、システムを効率よく評価するためには、人間の知覚と同等の性能を有する音像定位評価システムが必要であると考えられる。

そこで本研究では、どのような音を提示するかという課題では、聴取時の違和感を減らすことを主題とし、ユーザの行動や嗜好の調査と、ユーザにとって違和感が少ない音声加工方法の検討を行った。また、音源定位評価モデルについては、深層学習によるモデルを用い、学習データによる精度の違いについて検討を行なった。

3. 研究の方法

本研究は、違和感が少ない音の提示に関する検討と、音源定位モデルに関する検討を行った。

(1) 違和感が少ない音の提示に関する検討

違和感が少ない音の提示に関する検討では、環境音に対するユーザの嗜好の調査、ならびに、自己の音声を聴取した際の違和感の低減法の検討を行った。環境音に対するユーザの嗜好の調査は、インターネットを利用したアンケート調査により実施した。また、自己音声聴取時の違和感の低減のため、既存の声質変換技術[2]を活用することで解決を図った。

(2) 音源定位モデルに関する検討

音源定位モデルの検討では、深層学習を用いることで実現を図った。特に残響時間条件と、音源信号条件に着目し、学習と評価で異なる条件において、深層学習モデルの性能がどのように変化するか評価した。

4. 研究成果

(1) 違和感が少ない音の提示に関する検討

身の回りで発生する不快な環境音について、どの程度の不快感があるかについて調査を行った[3]。大学生 18 名に対して、不快であると感じている環境音を回答させ、その中から回答頻度が高かった環境音 13 種類について、インターネットを利用しアンケート調査を実施した。この環境音に対する不快感の調査結果から、不快さに世代間があることが確認できたが、個々の環境音については傾向がみられなかった。

加えて、この調査から、音響の仮想現実技術で必要となるイヤホンやヘッドホン利用について、日に 1 時間以上利用するとの回答が、20 代では 8 割程度であるのに対し、60 代以上では、2 割程度であった。回答に顕著な差が見られたことから、一つの技術で全世代を対象とすることの困難さが明らかとなった。このことは、検討するシステムの利用者層を細分化し、それぞれの層に合わせた検討を進めていく必要があると言える。

また、自己音声聴取時の違和感の低減の検討[4]からは、頭部や耳介による音声の変形と、骨伝導音などの体内伝導音を模擬するために、声質変換ツールを用いることの有効性を示すとともに、声質変換ツールのパラメータの探索範囲を限定することができた。

(2) 音源定位モデルに関する検討[5]

音源定位モデルに関する検討では、4 層の畳込み層と 2 層の全結合層から成るネットワークを用い、音源方向を水平面上の 72 方向とするクラス分類問題として音源方向推定実験を行った。実験では、ダミーヘッドマイクロホンを用いて測定した両耳室内インパルス応答(Binaural Room Impulse Response: BRIR)を用いた。なお、この BRIR は可変残響室において、残響時間を 151、179、198、224、238、261、303、417 ms として測定されたものである[6]。

音源信号として、RWC 研究用音楽データベース[7]に含まれるクラシック音楽(No. 4、592 s)、ポピュラー音楽(No. 55、女声ボーカル、249 s)と、講義音声(男声、1,616 s)、ホワイトノイズ(900 s)の 4 種類を用いた。各音源信号に残響時間が異なる 8 種類の BRIR を畳込み、1 s (48,000 点)ごとのスペクトログラムを求め、学習と評価に用いた。なお、左耳信号のスペクト

ログラムは画像の R チャンネル、右耳信号のスペクトログラムは画像の G チャンネルに収め、画像の B チャンネルは全要素を 0 とした。

実験は、学習と評価に用いる音源信号の条件(音源条件)として、1) 学習、評価とも音源信号としてホワイトノイズを使用、2) 学習は、クラシック音楽、ポピュラー音楽、講義音声とし、評価はホワイトノイズを使用、の 2 条件で行った。また、それぞれの音源の条件において、学習と評価に用いる BRIR の残響時間の組み合わせを 4 条件(残響時間条件)用いた。

音源方向推定の結果の例を図 1 に示す。図 1 は、学習には残響時間が短いデータ (151 ~ 261 ms) を用い、評価には残響時間が最も長いデータ (417 ms) を用いた。図の横軸は正解音源方向、縦軸は推定音源方向であり、推定結果が対角線上にある場合が正解となる。また、一点鎖線上に推定結果がある場合は、前後誤りとなる。なお、推定結果の頻度を濃淡で表示した。図より、学習と評価に用いる音源信号が異なる場合においても、前後誤りが生じていないこと、両耳方向 (90°、270°) での誤差が小さいことなど、人と同じような推定結果とはなっていないことが確認できる。検討したモデルでは、人よりも高度な推定が可能となっており、人と同じような誤りを生ずるようなモデルとなるためには、人間の聴覚特性モデル[8]の導入や、正解方向の幅を許容したマルチラベル分類の検討、簡略化された頭部伝達関数を用いた検討などが考えられる。

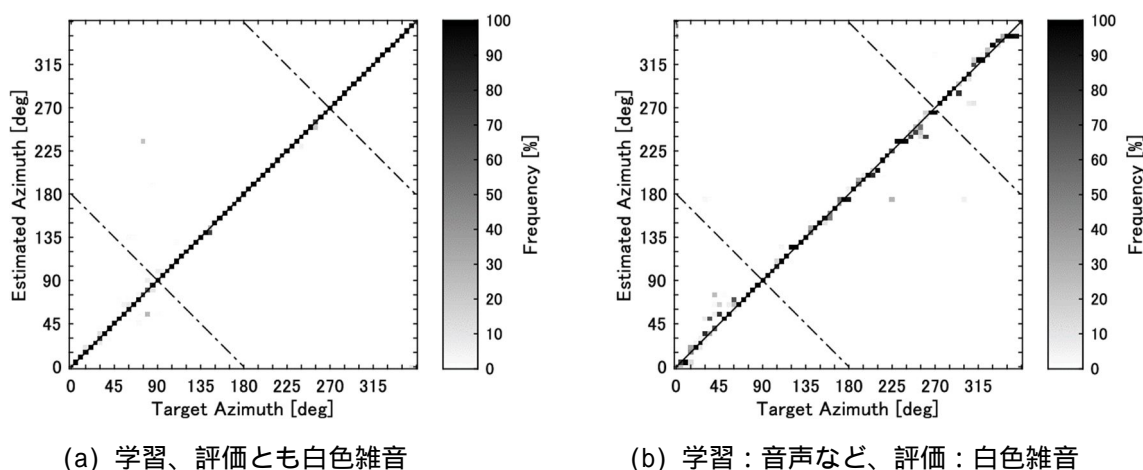


図 1 深層学習による音源定位の結果の例 (いずれも、学習には残響時間が短いデータ、評価には残響時間が最も長いデータを用いた)

< 引用文献 >

- [1] Facebook Lab, Enhanced hearing, <https://www.facebook.com/TechAtFacebook/videos/678868379502878/> (2022年6月2日閲覧)
- [2] M.Morise, F.Yokomori, K.Ozawa, WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, IEICE transactions on information and systems, E99-D(7), pp.1877-1884, 2016.
- [3] 西野隆典、野田要、選択型傾聴システムにおける主観評価にもとづく不要音判定の検討、日本音響学会 2020 年春季研究発表会、2020 年 3 月
- [4] 野々山大樹、鈴木千文、西野隆典、声質変換手法を用いた自己聴取音声の生成と評価、電気学会論文誌 C, 141(12)、pp.1267-1268、2021.
- [5] 西野隆典、畳込みニューラルネットワークを用いた残響時間が異なる両耳信号による音源方向推定の検討、日本音響学会 2021 年春季研究発表会、2021 年 3 月
- [6] 斉藤文訓、西野隆典、伊藤克亘、武田一哉、虚像法を用いた両耳室内インパルス応答の推定、日本音響学会 2004 年秋季研究発表会、2004 年 9 月
- [7] 後藤真孝、橋口博樹、西村拓一、岡隆一、RWC 研究用音楽データベース：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース、情報処理学会論文誌、45(3)、pp.728-738、2004
- [8] The Auditory Modeling Toolbox, <http://amtoolbox.org/> (2022年6月2日閲覧)

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 野々山 大樹, 鈴木 千文, 西野 隆典	4. 巻 141
2. 論文標題 声質変換手法を用いた自己聴取音声の生成と評価	5. 発行年 2021年
3. 雑誌名 電気学会論文誌C	6. 最初と最後の頁 1267 ~ 1268
掲載論文のDOI (デジタルオブジェクト識別子) 10.1541/ieejieiss.141.1267	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 Hibino Sho, Suzuki Chifumi, Nishino Takanori	4. 巻 42
2. 論文標題 Classification of singing insect sounds with convolutional neural network	5. 発行年 2021年
3. 雑誌名 Acoustical Science and Technology	6. 最初と最後の頁 354 ~ 356
掲載論文のDOI (デジタルオブジェクト識別子) 10.1250/ast.42.354	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 西野隆典	4. 巻 14
2. 論文標題 音響クロスリアリティの実現に向けて	5. 発行年 2021年
3. 雑誌名 電子情報通信学会Fundamentals Review	6. 最初と最後の頁 279 286
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/essfr.14.4_279	査読の有無 無
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

〔学会発表〕 計14件（うち招待講演 0件/うち国際学会 3件）

1. 発表者名 西脇正悟, 西野隆典
2. 発表標題 発話に伴う筋電位分析にもとづく身体疲労が歌唱におよぼす影響の評価
3. 学会等名 日本音響学会2022年春季研究発表会
4. 発表年 2022年

1. 発表者名 西脇正悟, 西野隆典
2. 発表標題 筋電計測による発声時の負荷の評価
3. 学会等名 令和3年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2021年

1. 発表者名 西野隆典
2. 発表標題 畳込みニューラルネットワークを用いた残響時間が異なる両耳信号による音源方向推定の検討
3. 学会等名 日本音響学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 日比野翔, 鈴木千文, 西野隆典
2. 発表標題 畳込みニューラルネットワークを用いた昆虫の鳴き声の判別
3. 学会等名 日本音響学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 Chiaki Nakagaito, Takanori Nishino, and Kazuya Takeda
2. 発表標題 Generation of Origami Folding Animations from 3D Point Cloud Using Latent Space Interpolation
3. 学会等名 SIGGRAPH ASIA 2020 (Poster) (国際学会)
4. 発表年 2020年

1. 発表者名 野々山大樹, 鈴木千文, 西野隆典
2. 発表標題 自己聴取音声の生成で生じる特徴量の変化
3. 学会等名 日本音響学会2020年秋季研究発表会
4. 発表年 2020年

1. 発表者名 小林慎司, 西野隆典
2. 発表標題 畳込みニューラルネットワークを用いた交通環境音の識別の検討
3. 学会等名 令和2年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2020年

1. 発表者名 野々山大樹, 鈴木千文, 西野隆典
2. 発表標題 音声変換により生成した自己音声の評価
3. 学会等名 令和2年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2020年

1. 発表者名 中垣内千晶, 西野隆典, 武田一哉
2. 発表標題 深層学習による三次元折り紙形状推定における視点位置の影響
3. 学会等名 第19回情報科学技術フォーラム(FIT2020)
4. 発表年 2020年

1. 発表者名 野々山大樹, 西野隆典
2. 発表標題 声質変換手法を用いた自己音声生成の検討
3. 学会等名 情報処理学会研究報告[音学シンポジウム2020 (第127回 音楽情報科学研究会)]
4. 発表年 2020年

1. 発表者名 Yusuke Sakai, Toshimitsu Watanabe, Yoshio Ishiguro, Takanori Nishino, and Kazuya Takeda
2. 発表標題 Effects on the Perception of Speed and Normality When Virtual Reality Scenes
3. 学会等名 2019 Automotive User Interfaces (国際学会)
4. 発表年 2019年

1. 発表者名 榮井 優介, 石黒 祥生, 西野 隆典, 武田 一哉
2. 発表標題 FollowSelect: 準備動作が必要な機器の利用に適した経路追従型メニュー選択手法
3. 学会等名 インタラクション2020
4. 発表年 2020年

1. 発表者名 西野隆典, 野田要
2. 発表標題 選択型傾聴システムにおける主観評価にもとづく不要音判定の検討
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------