7ZK-04

音響特徴量を用いた「ながら聴き」音声のレコメンド手法に関する有効性の検証

足立 潤治[†1,a)]　河瀬 彰宏[†1]

概要：これまで「ながら聴き」場面における音響特徴量に基づく音声トラックの推薦システムの有効性は，検証されて来なかった．本研究では，「ながら聴き」場面において，音響特徴量を用いた音声トラックの推薦の有効性を明らかにすることを目的とする．日常的に音声配信サービスを利用する 10 名を対象に，「ながら聴き」場面を演出し，MFCCs と Earth Mover's distance による音声トラックの推薦結果と，既存のサービスの推薦結果を交互に聴取・評価させる実験を行った．その結果，音響特徴量を用いた推薦の有効性を明らかにした．

1. はじめに

2019 年以降，Podcast のような音声メディアの聴取者は増加している．Radiotalk 社が自社サービスの利用者に対して行った調査結果によれば，ユーザは，運転中や料理中といったスマートフォンから両手が離れている状態である「ながら聴き」場面での使用例が多いことを明らかにした[*1]．しかし，調査の結果は公開されておらず，その有効性は証明されていない．また，音響特徴量を用いた楽曲の推薦に関する研究は行われている一方で，Podcast などの音声メディアや，人の話し声の特徴量を用いた推薦手法に関する研究はほとんど実施されてこなかった．

Logan(2004) は，Mel Frequency Cepstral Coefficients(MFCCs) に基づく 19 次元の音響特徴量を楽曲群から抽出し，Rubner et al.(2000)[4] が考案した Earth Mover's distance(EMD) を用いた楽曲推薦システムを提案し，その有効性を明らかにした [3]．Domingues et al.(2013) は，協調フィルタリングと音響特徴を用いた推薦手法について，両者の欠点を補うハイブリッド型の推薦システムを提案した．ハイブリッド型の推薦システムを構築する際に，MFCCs を含む音色に関する 4 種類の音響特徴量を使用した．そして，構築したハイブリッド型の推薦システムを，協調フィルタリング型と，コンテンツベース型の推薦システムと比較した．その結果，ハイブリッド型の推薦システムは，従来の推薦システムよりも，総合的に優れた音楽推薦を行うことを明らかにした [1]．滝澤ら (2009) は，「ながら聴き」場面に適した音楽の自動選曲を行うシステムを提案し，そのシステムがながら聴きに適しているかの判定を行った．その結果，作業状況に応じた自動選曲が有効である可能性が見られた [5]．

このように，音楽の音響特徴量を用いた推薦手法に関する研究は行われており，音楽のながら聴きにおける有効性の検証も行われている．しかし，Podcast などの音声メディアや，人の話し声に特化した推薦手法は提案されてこなかった．以上より，本研究の目的は，音声配信サービスのながら聴きにおいて，声の音響特徴量を用いた推薦システムの有効性を明らかにすることである．この推薦システムの有効性が示されることにより，様々なプラットフォーム上において，音声配信サービスの聴取ユーザが好みの声をもつ配信者にたどり着く可能性を広げることに寄与すると考えられる．

2. 実験方法

本研究では，実験に使用する 2 種類の推薦システムを構築し，それらを用いた 2 段階の運用実験を実施した．実験に使用した Podcast トラックは，2021 年 11 月 30 日時点で Spotify 上に配信されていた日本語の音源を用いた．運用実験では，使用するトラックの内容による評価の偏りが懸念されるため，本研究では，各 Podcast のジャンルのページにおいて特集されていた全 Podcast を用いた．Spotify 上では，14 ジャンルが存在するが，運用実験では，「オリジナル&独占」ジャンルにおける全 Podcast が他カテゴリにも登場する問題があること，「教育」ジャンルには，日

†1 現在，同志社大学文化情報学部
Presently with Faculty of Culture and Information Science, Doshisha University, Kyotanabe-shi, Kyoto 610–0394, Japan
a) adachi.junji@dh.doshisha.ac.jp

*1 https://prtimes.jp/main/html/rd/p/000000001.000043103.html[2022 年 1 月 7 日訪問]

本語以外の言語による Podcast が多く含まれることから，「オリジナル&独占」および「教育」を省く 12 ジャンルを対象に用いた．12 ジャンルにおいて特集された全 Podcast より 10 エピソードづつを収集し，エピソード数が 10 に満たない Podcast のトラックの収集は実施しなかった．

本研究では，2 つの推薦システムを構築した．1 つは，Logan et al.(2001)[2] および Logan(2004)[3] に基づく，音響特徴量を用いた推薦システムである．このシステムでは，各トラックから音響特徴量として 19 次元の MFCCs を算出し，EMD によってトラック間の音響特徴量の類似度を算出することで，類似した音響特徴量を持つトラックの推薦を実施した．もう 1 つの推薦システムは，音響特徴量を用いない，Spotify アプリ上の Podcast 推薦サービスを用いた．運用実験の第 1 段階では，実験協力者 10 名に対して，次の手順に従って聴取実験で用いる推薦トラックの算出を行った：(1)Spotify に登録させ，2 週間の Podcast 聴取を行わせた．その際，好みのトラックを発見した場合，そのトラックを記録してもらった．(2)2 週間の経過後に，実験協力者からトラック群の記録を提出してもらった．(3) 提出された記録を上述の 2 種類の推薦システムの入力として用いて，各システムからトラックを推薦した．

第 2 段階では，第 1 段階と同じ実験協力者 10 名に対して，次のルールに基づく聴取実験を行った：(i) 家庭用ゲーム機 Nintendo Switch の専用ソフト「リングフィットアドベンチャー」をプレイさせながら，第 1 段階で算出された推薦トラックの聴取を行った．その際に，2 つの推薦システムから算出された上位の推薦トラック（音響特徴量を用いた推薦結果と用いていない推薦結果）を交互に再生させた．(ii) ながら聴き中の実験協力者は，自身の嗜好に合わないトラックが再生されていると判断した場合は，トラックをスキップすることが認めた．実験協力者によるトラックの聴取傾向（スキップの有無）に基づき，2 つの推薦システムの推薦結果と，実験協力者の嗜好の関係性について，正答率 $= ((TT+FF)-(TF+FT))/(T+F)$ を算出した．ただし，TT は，音響特徴量を用いて推薦され，実験協力者が再生したトラック数；FF は，音響特徴量を用いずに推薦され，実験協力者がスキップしたトラック数；TF は，音響特徴量を用いて推薦され，実験協力者がスキップしたトラック数；FT は，音響特徴量を用いずに推薦され，実験協力者が再生したトラック数；T は，再生が完了されたトラック数；F は，スキップされたトラック数である．

3. 結果と考察

図 1 は，Podcast の推薦トラックと同一のトラックが構築システムから推薦された割合を上位 1 位，5 位，10 位，20 位に分けて集計した結果である．図 1 の 2 つの折れ線は，本研究において音響特徴量を用いて構築した推薦システムと，Logan(2004)[3] の推薦精度の比較であり，推薦システムが適切に機能していることを確認した上で聴取実験を実施した．

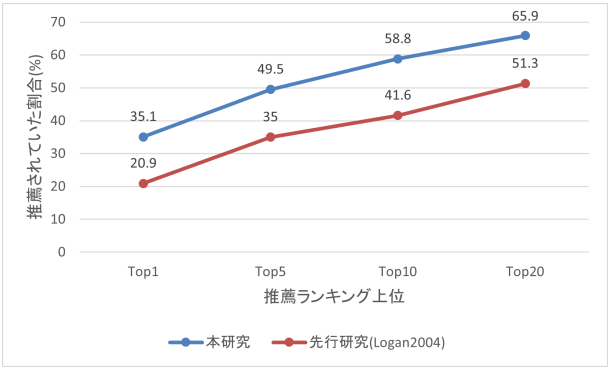

図 1 本研究と Logan(2004)[3] の推薦精度の比較

聴取実験の結果，10 名のうち 3 名は，非負の正答率が得られ，最大値は，0.55 であった．また，2 名の実験協力者について，音響特徴量を用いて推薦されたトラックが，聴取者の嗜好と相反する結果（正答率 < 0）であることが確認された．以上の実験結果より，音響特徴量を用いた推薦システムが，音声配信のながら聴き場面において有効に行われていることを明らかにした．

4. 結論

本研究では，ながら聴き場面において，音響特徴量を用いた音声トラックの推薦の有効性を検証した．聴取実験では，Podcast の聴取経験の有無を考慮せず実験協力者を選出した．そのため，運用実験の第 1 段階において，日常的に Podcast を利用しない実験協力者から得られた記録が少ない問題があった．今後の課題として，サンプルサイズを増加する際に，日常的に Podcast を聴取する実験協力者を含めることによって，より精緻な結果が得られると考えられる．

参考文献

[1] Domingues, M.A., Gouyon, F., Jorge, A.M., Leal, J.P., Vinagre, J., Lemos, L., and Sordo, M.: Combining usage and content in an online recommendation system for music in the long tail, *International Journal of Multimedia Information Retrieval*, **2**(1), pp.3–13 (2013).

[2] Logan, B., Salmon, A.: A Music Similarity Function Based on Signal Analysis, *International Conference on Multimedia and Expo 2001 (ICME)*, pp.22–25, (2001).

[3] Logan, B.: Music recommendation from song sets, *International Conference on Music Information Retrieval 2004*, pp.425–428, (2004).

[4] Rubner, Y., Tomasi, C., and Guibas, L.J.: The earth mover's distance as a metric for image retrieval, *International journal of computer vision*, **40**(2), pp.99–121, (2000).

[5] 滝澤勇介・西本一志：作業状況に基づく 「ながら聴き」用自動選曲プレーヤ" LISWO", 情報処理学会研究報告, ヒューマンコンピュータインタラクション・特集：新領域創造インタラクション, **2009**(28), pp.115–122, (2009).