

i-vectorに基づく賑わい音の推定方式における学習データの増加に伴う識別性能に関する検討 Study on the recognition performance of i-vector based hubbub sound estimation with the increase of training data

呉 セン陽

Wu Zhenyang

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本研究で行う賑わい音の推定は、賑わっている状況という音場面の推定である。賑わっている状況とは、商業施設や観光地の周辺や催し物などに多くの人が集まり、活気のある状況であり、このような状況で発生する音を本研究では「賑わい音」と定義する。本研究では i-vector に基づく話者識別手法を応用した賑わい音の推定方式を提案する。提案方式における学習データを变化させて性能評価実験を行った。

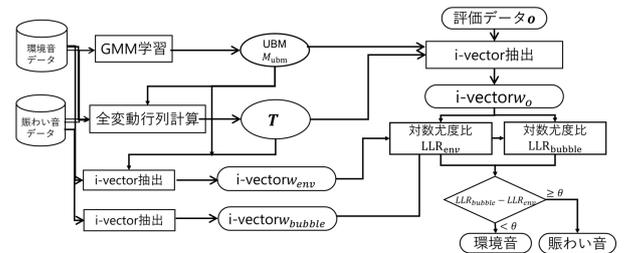


図 1: 提案方式

1 はじめに

人間の生活を取り巻く環境音は、多種多様な音源やその場の状況や雰囲気など、多くの情報を含んでいる。収録した環境音を解析し、その中に含まれている情報を得るために、環境音を分析する研究が行われている。本研究で行う賑わい音の推定は、賑わっている状況という音場面の推定である。賑わっている状況とは、商業施設や観光地の周辺や催し物などに多くの人が集まり、活気のある状況であり、このような状況で発生する音を本研究では「賑わい音」と定義する。環境音から賑わい音を検出することにより、街の賑わっている場所を検出することができると考えられる。

i-vector は話者認識において高い性能を達成している発話表現の一つで、発話ごとに算出した混合ガウス分布 (GMM) に因子分析を適用することで得られる。i-vector を用いた識別方式は、話者識別の分野においても高い性能を示しているため、そこで、i-vector に基づく話者識別手法を応用した賑わい音の推定方式を提案する。

本研究では、提案方式の有効性を示すために、観光地で収録された環境音を用いて賑わい音の推定実験を行った。提案方式の使用により、i-vector の賑わい音識別器システムを構築することは可能であると考えられる。

2 提案方式

図 1 は提案方式である。i-vector に基づく話者識別手法では、入力音声から i-vector を求め、入力から求めた i-vector と登録された i-vector の類似度を計算することにより話者識別を行う技術である。また、類似度を求める際には PLDA (Probabilistic Linear Discriminant

Analysis) によって話者識別に寄与しない情報を低減するアプローチが用いられる。このとき、話者内変動と呼ばれる同一話者の音声の発話の違いやチャンネル歪みなどの音響的な違いによる影響を補正できる。これにより話者内変動に頑健な話者識別システムを構築できる。そこで、二つの話者識別するように環境音と賑わい音の高い識別性能が得られる事が期待できる。GMM スーパーベクトルは、時系列データである発話をベクトル空間上の一点として表現するものである。i-vector も、この GMM スーパーベクトルを基礎としている。

3 実験

3.1 実験条件

音響モデルの特徴量には MFCC を用いた。分析帯域幅は 0 Hz から 16 kHz とした。抽出時のフレームサイズは 25 msec、フレームシフトは 10 msec、分析窓の窓関数はハミング窓とした。学習には HTK3.4.1 を用いる [2]。

本研究では、38, 59 次元の MFCC を用いる。特徴量次元数は 38, 59、混合数は 128, 256, 512 として構築する。T のランクは 100, 400, 600 とした。

3.2 実験データ

学習に用いるデータは文献 [1][3] で環境音と賑わい音を収録したデータである。

評価データでは 2017 年 9 月 23 日に、催し物のない休日の収録された。追加されたデータでは 2017 年 10 月に、秋祭りの収録された。[3] 追加されたデータは三段階に分けられる。各段階は 345 個約 4.6 % に増加した。学習データには賑わいのデータが従来の 3.5 % から 17.3 % に増加した。

表 1: 学習データ

	環境音収録	賑わい音収録
日程	2014/6/25 ~ 2015/1/31	2016/1/10 ~ 7/16
場所・行事	住宅街 駅の近辺	成人式 センター試験
	商店街	入学式/花火大会
データ数	7,499	263

3.3 評価尺度

本研究の評価指標として F 尺度を用いる。最大値 1 に近いほど高評価になる。再現率と適合率の調和平均によって求められる。また、ROC 曲線を描いてその特性を用いる。曲線より下側の面積が広いほど性能が高くなる。縦軸は TPR(=Recall), 横軸は FPR としてプロットする。FPR は賑わっていないデータのうち、識別器により誤って賑わいのデータとして検出されたデータの割合である。TPR は賑わいのデータのうち、識別器により正しく賑わいのデータとして検出できたデータの割合である。

3.4 実験

今回は二つの実験を行った。一つ目は最良の実験条件を得られるために実験条件の変化させた実験を行った。二つ目は最良の実験条件で学習データを追加させて実験を行った。

3.4.1 両方とも追加した結果

表 2 の左側では環境音データと賑わい音データ両方とも追加した結果である。表 2 より、i-vector に基づく賑わい音の推定方式について、より詳細に分析パラメータを変更したところ、F 尺度の最大値 0.7927 であった。また、わずかではあるが、特徴量次元数 MFCC や全変動行列 TvRank を増やすことで識別性能は上がる傾向にあることが示唆された。本研究の実験条件においては、計算量の少ない特徴量次元数 MFCC59、混合数 256mix、Tvrank 600 が最良であると考えられる。

3.4.2 賑わい音データだけ追加した結果

表 3、図 2 より、今回は賑わい音データ比率を 8.1%、12.7%、17.3% と変化させ、音響モデルを作成した。追加しない結果 F 尺度は 0.7893 から 0.7923 まで上がった。賑わいデータを増加して識別性能はわずかに改善された。最良の実験条件の場合は F 尺度の最大値は全体として上がる。また、学習データ(賑わい音)の増加の量と関係が少ないことが示唆された。

4 まとめ

本報告では i-vector に基づく賑わい音の推定方式における学習データの増加に伴う識別性能を確認した。一つ目の実験では、i-vector に基づく賑わい音の推定方

表 2: Fmeasure の最大値

MFCC38	TvRank100	TvRank400	TvRank600
128mix	0.7500/0.7496	0.7692/0.7632	0.7853/0.7686
256mix	0.7468/0.7515	0.7623/0.7695	0.7789/0.7833
512mix	0.7539/0.7524	0.7733/0.7711	0.7815/0.7759
MFCC59	TvRank100	TvRank400	TvRank600
128mix	0.7637/0.7631	0.7716/0.7720	0.7866/0.7899
256mix	0.7581/0.7589	0.7762/0.7772	0.7927/0.7893
512mix	0.7612/0.7640	0.7764/0.7776	0.7831/0.7863

表 3: MFCC59 次元, Mix256, TvRank600 の条件下の実験結果

賑わいのデータ比率	Fmeasure
(1)3.5 %	0.7893
(2)8.1 %	0.7923
(3)12.7 %	0.7898
(4)17.3 %	0.7919

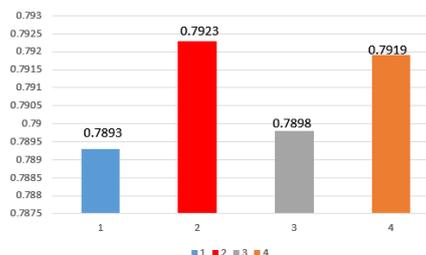


図 2: F 尺度の対比

式について、より詳細に分析パラメータを変更したところ、F 尺度の最大値 0.7927 であった。実験結果により、最良の実験条件を得られた。また、賑わい音のデータだけ追加した結果により、F 尺度の最大値はわずかに改善された。ただし、学習データ(賑わい音)の増加の量と関係が少ないことが示唆された。

今後は、i-vector と他の特徴量を組み合わせて認識手法を提案しており、認識精度が改善することが期待している。

参考文献

- [1] T.Tanaka, *et al.*, "A Classification Method For Crowded Situation Using Environmental Sounds Based On GMM-UBM," 5th Joint Meeting, Acoustical Society of America and Acoustical Society of Japan, pp. 3110, Nov. 2016.
- [2] The HTK Book, "<http://htk.eng.cam.ac.uk/>"
- [3] 朝田興平, *et al.*, "クラウドソーシングによる賑わい音識別方式のフィールド実験評価," 2018 年日本音響学会春季研究発表会, pp. 79-82, March. 2018.