

声の主観的評価と音声特徴量の相関分析

寺下 逸生テオ, 日向寺 拓海, 川勝 真喜 (東京電機大)

背景・目的

背景

近年, 自動音声が普及しており, 様々な場所で使用されているが, 聞き取りづらさや聞き心地の無さが見受けられる

目的

音声の聞き取りやすさや好みに影響する要因を明らかにし, 音声技術の向上に貢献することを目指した

既存研究^[1]

WeissとBurkhardtの研究^[1]

声の好感度に影響を与える要因について検討を行った
話者の音声を評価した結果

男性の声: 基本周波数の低さ
女性の声: スペクトル分散の大きさ
共通: 速い発話

また 「温かい」「柔らかい」「響く」 声が好感が得られる

本研究では,

被検者数 20→100 使用音声 ドイツ語→日本語
好みに関わる評価項目の推定の際に用いる音声数 2→8

この条件と独自の
実験手順で行った

主観評価実験・分析

実験目的

声の好みに寄与する特徴量の推定

被検者

100人(男性90人 女性10人)年齢 18～54才

使用音声

高道らによる公開データセットである,
JVS (Japanese versatile speech) corpus ^[2]
同一の文章であり, 標準的と思われる男女 4 名ずつの音声を実験者が選んだ

実験手順

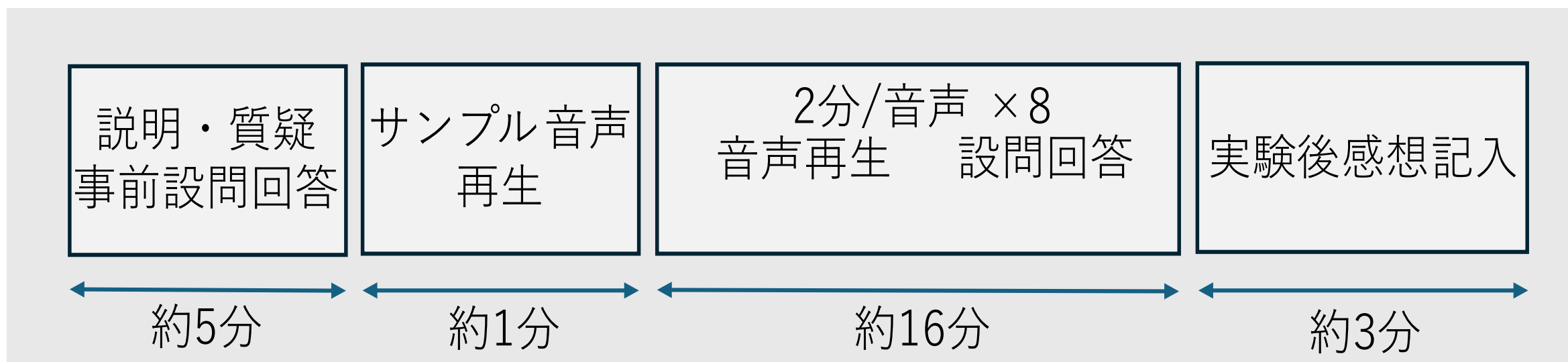


図1 実験手順

→ 8つの音声は実験毎にランダムに再生
評価方法

5段階評価による主観的評価

しゃべり方に関する項目

- ・明瞭であるか (滑舌)
- ・速度
- ・好みのリズムか
- ・好みのイントネーションか

声に関する項目

- ・この声は好みか
- ・温かさ
- ・柔らかさ
- ・心地よさ
- ・リラックスの度合い
- ・かすれ度合
- ・高さ
- ・明るさ
- ・力強さ
- ・響き
- ・ボリューム

→合計15項目

その他の項目

その他

- ・年齢
- ・性別
- ・直近2年以上の音楽歴

特徴量

特徴量の変数を「_」で分け, 以下のように設定

前半部分

- ・silent(sl)_time: 無音秒数
- ・active_time: 発話秒数
- ・db: 音の大きさ (デシベル)
- ・mfcc_i: MFCCの第i次元 (全13次元)
- ・freq: 基本周波数

後半部分

- ・mean: フレームごとの平均
- ・std: フレームごとの標準偏差

(例)

freq_mean
基本周波数の平均

mfcc_10_std
MFCC第10次元の標準偏差

MFCCとは

人間の聴覚特性を模倣した音声特徴量で, 音声認識や話者識別によく使われる。

結果

全被検者のアンケート結果の
平均値と各特徴量の相関係数を
計算した

結果.1

MFCC第3次元とMFCC第10次元の
平均値が「この声は好みか」に対し,
大きな相関係数値を示した

結果.2

「温かさ」, 「心地よさ」,
「リラックスの度合い」, 「響き」,
「好みのイントネーションか」につ
いて「この声は好みか」と同様の
大きな相関をもつ特徴量を確認できた

グループ化

100人分のアンケート
結果から好みの傾向で分
けるためにk-means(k=3)
でクラスタリング。
その後, PCAで圧縮し
クラスタの可視化した

表1 集計結果の全体と特徴量の相関の上位2位

	1st	2nd
この声は好みか	mfcc_3_mean(0.767)	mfcc_10_mean(0.693)
温かさ	mfcc_13_std(0.814)	mfcc_3_mean(0.684)
心地よさ	mfcc_10_mean(0.825)	mfcc_11_mean(-0.781)
リラックスの度合い	mfcc_3_mean(0.754)	mfcc_10_mean(0.714)
明るさ	mfcc_2_mean(-0.831)	freq_std(0.816)
響き	mfcc_10_mean(0.874)	mfcc_3_mean(0.855)
好みのことばの間	mfcc_1_std(-0.917)	mfcc_3_std(-0.679)
好みのイントネーションか	mfcc_3_mean(0.69)	mfcc_4_mean(-0.654)
高さ	freq_mean(0.974)	mfcc_5_mean(-0.963)
速度	sl_time_std(-0.986)	active_time(-0.933)

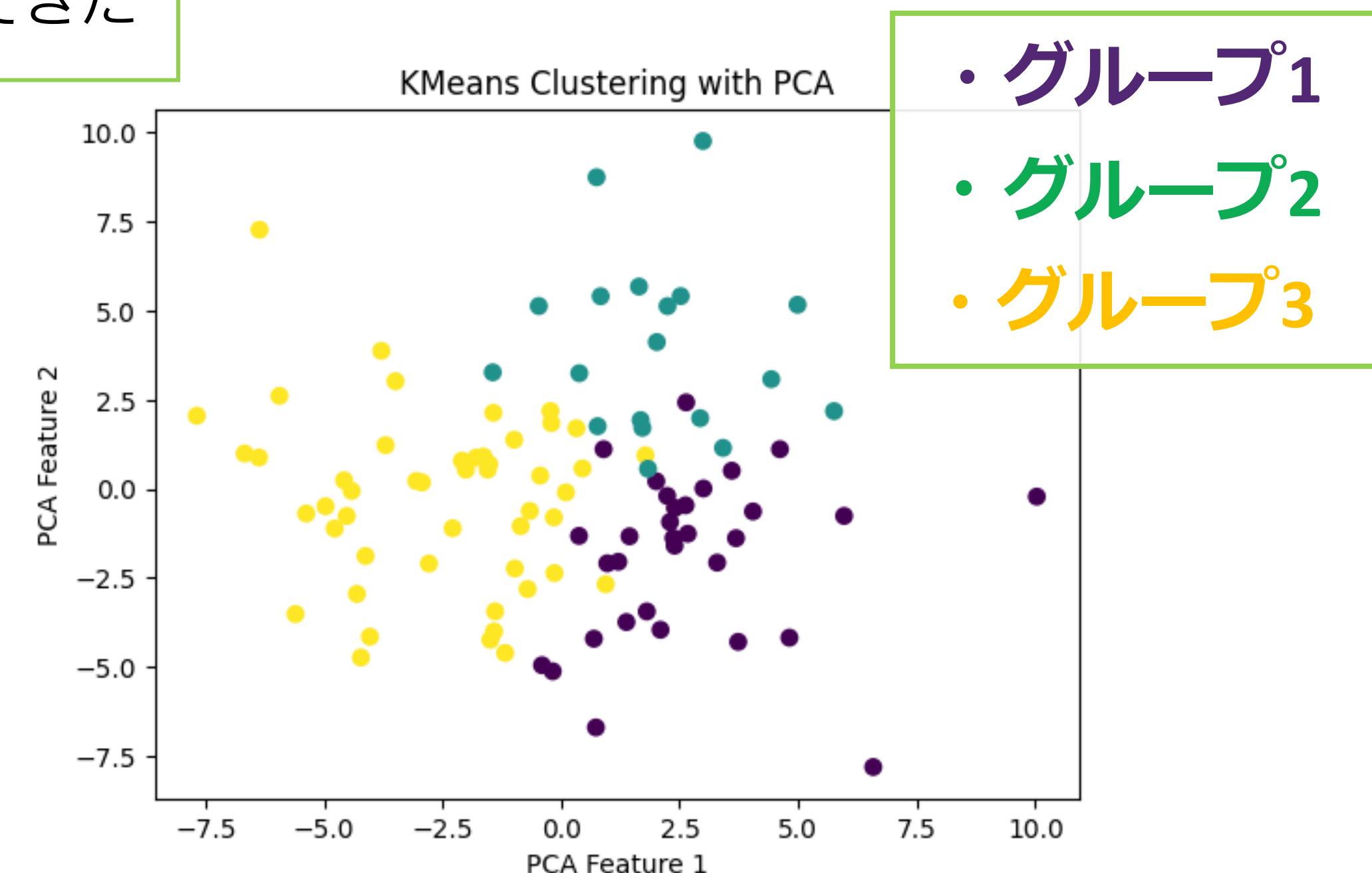


図2 第一成分と第二成分の散布図

表2 グループ1の集計結果と
特徴量の相関の上位2位

	1st	2nd
この声は好みか	mfcc_10_mean(0.847)	mfcc_3_mean(0.838)
温かさ	mfcc_10_mean(0.765)	mfcc_6_std(-0.75)
心地よさ	mfcc_10_mean(0.927)	mfcc_3_mean(0.821)
リラックスの度合い	mfcc_10_mean(0.839)	mfcc_3_mean(0.835)
明るさ	mfcc_2_mean(-0.817)	mfcc_11_mean(-0.789)
響き	mfcc_10_mean(0.917)	mfcc_3_mean(0.889)
好みのことばの間	mfcc_10_mean(0.767)	mfcc_3_mean(0.644)
好みのイントネーションか	mfcc_3_mean(0.87)	mfcc_10_mean(0.748)
高さ	freq_mean(0.971)	mfcc_5_mean(-0.949)
速度	sl_time_std(-0.976)	active_time(-0.957)

周波数特性型

表3 グループ2の集計結果と
特徴量の相関の上位2位

	1st	2nd
mfcc_6_mean(0.78)	silent_time(0.726)	
silent_time(0.864)	mfcc_6_std(-0.728)	
silent_time(0.818)	active_time(0.719)	
active_time(0.951)	silent_time(0.925)	
mfcc_1_std(0.804)	mfcc_3_std(0.668)	
mfcc_3_mean(0.791)	mfcc_12_mean(0.688)	
mfcc_10_mean(0.803)	mfcc_3_mean(0.8)	
silent_time(0.657)	mfcc_3_mean(0.648)	
freq_mean(0.968)	mfcc_5_mean(-0.964)	
sl_time_std(-0.95)	active_time(-0.868)	

言葉の間型

表4 グループ3の集計結果と
特徴量の相関の上位2位

	1st	2nd
freq_std(0.847)	mfcc_4_mean(-0.793)	
mfcc_11_mean(-0.802)	mfcc_12_std(-0.796)	
freq_std(0.895)	mfcc_11_mean(-0.838)	
mfcc_11_mean(-0.795)	mfcc_12_std(-0.727)	
freq_std(0.878)	mfcc_2_mean(-0.854)	
mfcc_9_std(0.637)	mfcc_13_std(0.598)	
mfcc_4_mean(-0.905)	freq_std(0.851)	
mfcc_2_mean(-0.823)	freq_std(0.819)	
freq_mean(0.968)	mfcc_5_mean(-0.96)	
sl_time_std(-0.983)	active_time(-0.922)	

基本周波数の変動型

まとめ

主観評価実験, 相関分析により, 声の好みに関係する特徴量を明らかにした。

さらに, k-meansによるクラスタリングにより, グループごとに好みに寄与する特徴量が大きく異なることを示した。

今後の課題

音声の追加

- ▶ 本研究は8つ
→ 評価に偏りが生まれている可能性あり

被検者ごとに集計結果を標準化

- ▶ 五段階評価は被検者ごとに,
尺度の差が生まれていた可能性あり

[1] W.Benjamin, et. al. "Voice attributes affecting. Likability perception.", Interspeech, 2010.

[2] S. Takamichi, et. al. "JVS corpus: free Japanese multi-speaker voice corpus, " arXiv preprint, 1908.06248, Aug. 2019.