

研究背景

研究背景

- TTS(Text to Speech)による音声支援が発展している
- 選択のために全ての試聴をすることはユーザの負担になる

研究目標

- 可視化を用い、TTSモデルを効率よく選択する
- 個人がなりたい印象に関連付けて検索する

本研究の手順

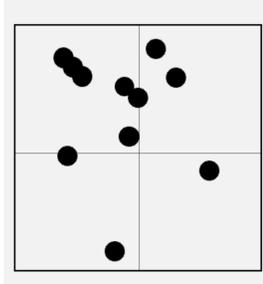
1 TTSモデルの特徴量抽出

VOICEVOX COEIROINK CoeFont



基本周波数、フォルマント構造、スペクトル傾斜 etc...

2 対話型推薦手法



事前に構成した特徴空間より、対話的にTTSモデル候補を提示する。

3 印象評価に基づく選択



印象評価をもとに絞り込み、最終的な選択を行う。

関連技術

対話型推薦手法

MusiCube：特徴量空間における対話型進化計算を用いた楽曲提示インタフェース(齊藤,2014)

- 楽曲特徴量から楽曲の雰囲気や推定
- 対話型遺伝アルゴリズムを用いてユーザの嗜好を提示



Fig4: MusiCubeのGUI画面

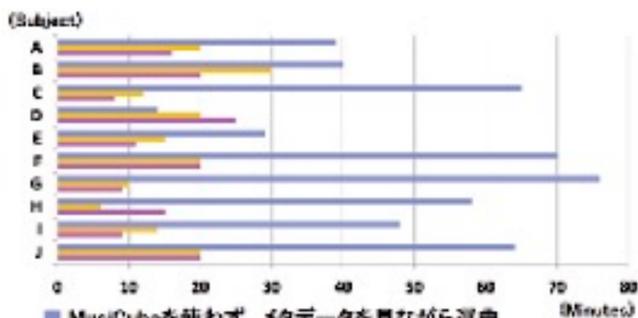


Fig5: 手法ごとのプレイリスト作成時間の比較

POINT

- ✓ 早期に満足度の高い推薦結果が得られる
- ✓ 嗜好に起因する特徴量を理解することができる
- ✓ 選択動作が直感的である

音声の印象評価

女性の声を例にした音響特徴量と印象評価の関係性の調査(菅原,2014)

- 事前のアンケート評価で印象分類評価
- 音響特徴量との関係を調査

POINT

- ✓ 音声の印象評価に対し、基本周波数とフォルマント構造の相関が観察された

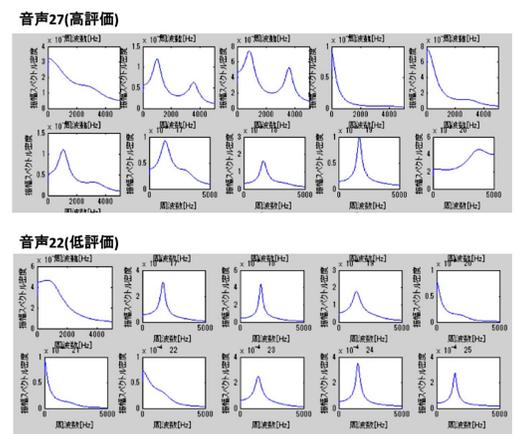


Fig1: 評価とフォルマント構造の違い

音声の聞き取りやすさに関わる性別による音響特徴の違い(小林,倉片,2021)

- 女性の方が母音空間が広い
- 男性の方が各子音のCOGに

POINT

- ✓ 女性は母音、男性は子音が聞き取りやすい

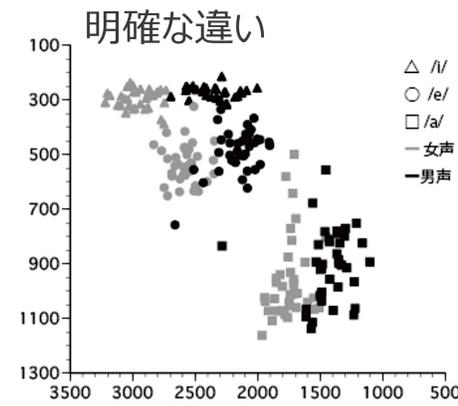


Fig2: 音声の性別によるフォルマント空間の違い

	母音	女声	男声
破裂音	/t/	1829.3(±154.4)	1641.5(±157.3)
摩擦音	/s/	5817.4(276.5)	5807.9(±223.2)
	/z/	3925.3(±677.7)	3320.1(±536.5)

()内は95%信頼区間を示す

Fig3: 各子音部でのCOG[Hz]

※COG: 破裂音の時間長と周波数の重心