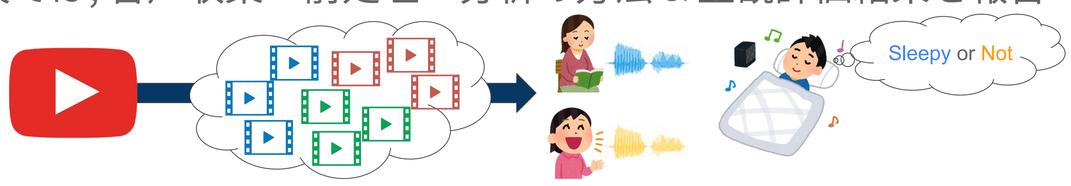
音学シンポジウム2023@電通大 ポスターセッション1 (05)

睡眠を誘発する音声刺激の生成に向けた分析と評価

◎熊田 順一[1], 齋藤 佑樹[1], 高道 慎之介[1], 渡邊 亞椰[1], 丹治 尚子[1], 長野 瑞生[2], 井島 勇祐[2], 猿渡 洋[1]
(1: 東京大学, 2: NTT)

概要:「聴いていると眠くなる声とは何か？」を調査する研究

- **睡眠障害: 世界的に深刻な社会問題の一つ**
  - Covid-19 の流行に伴い, その深刻さも増大する傾向[1]
  - 一般的な臨床介入法: 認知行動療法や睡眠導入剤の使用
    - いずれも高価もしくは有害な副作用の可能性あり[2]
- **関連研究: 音楽療法に有効な音響特徴量の調査と生成法**
  - スペクトル重心が低く, 滑らかに変動し, リズム音の強調が弱い音楽が有効である可能性を示唆[3]
  - 敵対的生成ネットワークに基づく音楽特徴量変換[4]
  - 「どのような**音声**が眠気を引き起こすか?」は未調査…
- **本研究:「眠くなる声」の合成に向けた基礎検討 (音声特徴量分析)**
  - 利用者が手軽に聴取できる音声コンテンツ生成への応用に向けて
  - YouTube の音声コンテンツを対象に, 睡眠導入音声と通常音声を比較
  - 本発表では, 音声収集・前処理・分析の方法 & 主観評価結果を報告

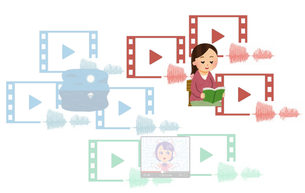


- **結果: 音声の基本周波数 (F0) & energy の重要性を示唆**
  - 低めで音量が小さく, 音高・音量のばらつきが小さい音声が望ましい?

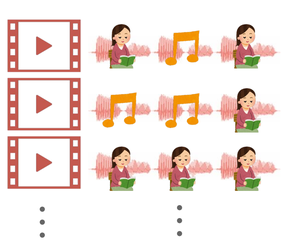
YouTube からの睡眠導入音声コンテンツ取得

全体的な処理フロー

- **JTubeSpeech[5]を参考に構築**
  1. 検索キーワードの作成: 朗読, 絵本読み聞かせ, ラジオ配信をカバー
     【睡眠導入】, 【睡眠朗読】, 【寝落ちラジオ】, etc.
  2. YouTube から動画取得: 得られた動画IDをリスト化
  3. 分析対象話者の選定:
     (1) 自動字幕付き
     (2) 背景音が最小限
     (3) 睡眠導入以外の発話有

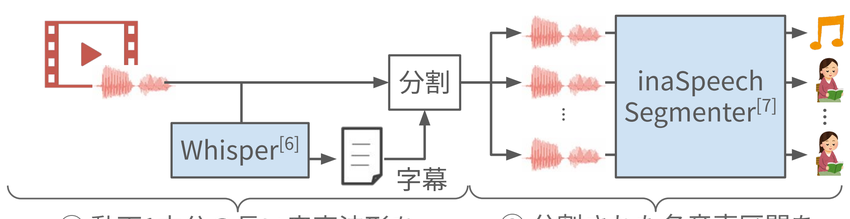


  4. 音声区間分割: 事前学習済みモデルを活用した2段階での分割



音声データとその前処理

- **分析対象データ: 女性プロ話者1名 (7h)**
  - **睡眠導入**動画: “あのときの王子くん”朗読 (3h)
  - **雑談配信**動画: YouTube Live アーカイブ (4h)
- **事前学習済みモデルを用いた音声区間分割**



① 動画1本分の長い音声波形を自動字幕ベースでラフに分割
② 分割された各音声区間を { 音声, 無音, 雑音, 音楽 } に分類

本研究で収集した音声区間の量 / inaSpeechSegmenter による分類結果の割合

	区間数	区間長 平均 [s]	総時間 [h]	音声	無音	雑音	音楽
睡眠導入	2,890	2.27	1.83	0.602	**0.199**	**0.185**	**0.014**
雑談配信	4,020	2.96	3.33	0.859	0.115	0.019	0.007

考察

- **YouTube 動画の検索結果**
  - 睡眠導入**音声**以外 (ノイズ) も多く存在
    - 字幕の情報量がない動画 (例:「ご視聴ありがとうございます!」等)
    - 睡眠導入用リラクゼーション**音楽**

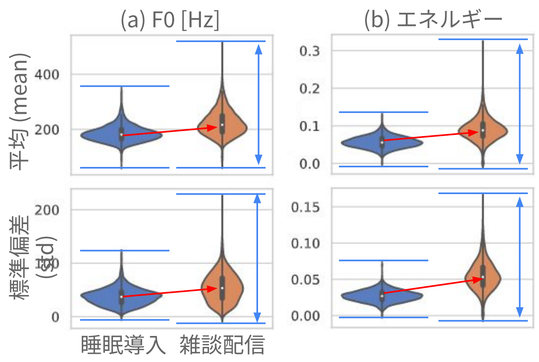
    → 字幕 alignment スコアで除去可能?
  - 音声**生成**タスク向けデータは少なめ?
    - BGMあり, 低品質収録, 字幕信頼性低

    → ダークデータ音声合成[8]の活用?
- **音声区間分割結果**
  - 睡眠導入: 非音声区間の割合が高
    - 無音: 朗読特有の間の取り方に起因?
    - 雑音: 瞬間的に発話が生じる無音区間
    - 音楽: 冒頭・終了で入るBGM (編集)
  - 雑談配信: 音声区間の数と長さが大
    - 自発音声特有の発話密度の高さに起因?
    - フィラーや言い淀みなども多く含む

睡眠導入音声の分析 & 音声印象に関する主観評価

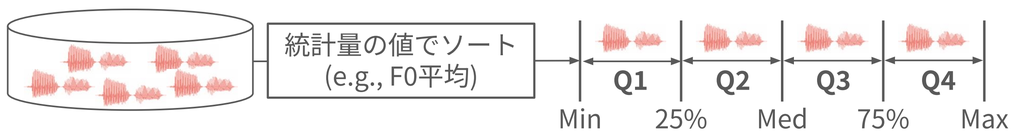
音声分析

- **韻律特徴量**
  - F0: WORLD ボコーダ[9]で抽出
  - Energy: librosa.feature.rms
- **統計量の可視化 (右図)**
  - 概ね, 睡眠導入動画の音声が mean, std ともに小さい



主観評価条件

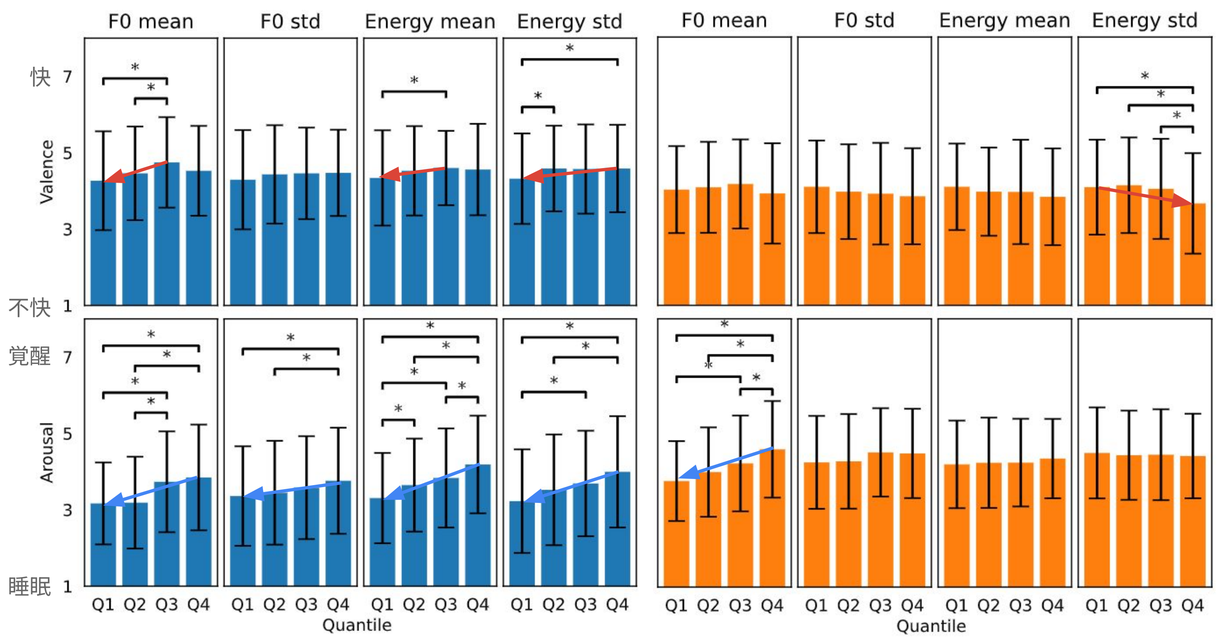
- **音声刺激の用意 (睡眠導入/雑談配信で共通)**
  - 準備: データを韻律特徴量の四分位数で分割
    - 四分位数の計算時には 音声区間長が 3.5[s] 以下のデータを除外
    - グループの数: { F0, energy } x { mean, std } x { Q1, Q2, Q3, Q4 } = 16
      (韻律特徴量 / 統計量 / 四分位グループ)



  - 前処理: WORLD ボコーダによる分析再合成処理を適用
    - 信号処理を用いた音声加工を想定 (サンプリング周波数: 48 [kHz])
- **評価基準: valence (不快-快) & arousal (睡眠-覚醒)**
  - 音声刺激を聴いた後の**被験者自身**の気分・心理状態を7段階で回答
- **被験者数: Lancers で募集した240名 (30名/評価)**
  - (分割基準統計量, 発話スタイル) のグループごとに評価を実施
  - 各グループから無作為に抽出された20サンプルを評価

主観評価結果

- **評価スコアの barplots (* は $p < 0.05$ で有意差あり)**
  - Valence (不快-快) に関する評価結果 (上2つ)
    - **ボコーダ分析再合成での劣化**が不快であると評価される傾向あり?
  - Arousal (睡眠-覚醒) に関する評価結果 (下2つ)
    - 睡眠導入音声: 韻律特徴統計量の値が小さくなると“睡眠”側にシフト
    - 雑談配信音声: F0 mean が小さくなると睡眠誘発効果あり?



- **本発表のまとめ**
  - 睡眠導入音声: **低めの声で音量を小さく, 音高・音量のばらつきが小さくなるように意図して発話**される傾向あり
  - 通常発話でも, F0 mean が低い音声は睡眠導入効果あり? → 要調査
- **今後の展望: テキスト音声合成 / 声質変換タスクでの調査**

Reference

[1] E. Altena et al., Journal of Sleep Research, vol. 29, no. 4. 2020.
[2] L. Degenhardt et al., Drug and Alcohol Dependence, vol. 67, no. 1, 2002.
[3] G. T. Dickson et al., Musicae Scientiae, vol. 26, no. 3, 2020.
[4] J. Yang et al., Proc. ICASSP, 2022.
[5] S. Takamichi et al., arXiv:2112.09323, 2021.
[6] A. Radford et al., Proc. ICLR, 2023.
[7] D. Doukhan et al., Proc. ICASSP, 2018.
[8] K. Seki et al., Proc. ICASSP 2023.
[9] M. Morise et al., Speech Communication, vol. 84, 2016.