



# 避難呼びかけ音声の持つ緊急性の分析と音声合成への適用の検討

☆ 原田 そら (木更津高専), 中田 亘 (東大・工), 高道 慎之介, 齋藤 佑樹 (東大院・情報理工), 齋藤 康之 (木更津高専), 猿渡 洋 (東大院・情報理工)



## 1. 概要

### 呼びかけ音声の重要性と研究課題

- 災害時は早期の避難行動が重要
- 事例研究 [1] から、避難行動喚起には避難呼びかけの有意性が示唆
- 呼びかけ効果を高めるには伝達媒体の1つ「避難呼びかけ音声」の適切な設計も有効

### 避難呼びかけ音声の先行研究

- 緊急性の高い音声は避難行動を促す[2]
- F0, 話速と緊急性の印象との関連性が示唆[3]
- 屋外の行政防災無線や合成音声の活用に関連し、音声のピッチ加工は可聴性に影響なし[4]

### 本研究の目的

- 効果的な呼びかけ音声, および音声合成は実現可能か?

1. 音声にある緊急性の要因について, 緊急性の異なる音声群を比較
2. 音声群間の音響特徴量差を利用し, 音声の緊急性を操作
3. 他話者音声にて学習済みのニューラルTTSへ緊急性操作機能を付与



## 2. 方法論

### 緊急性に寄与する音声特徴量の解明

- 3群の音声を準備  
平静時音声, 加工対象音声, 緊急時音声

- 訓練を受けた同一話者による緊急性の異なる音声の評価実験

#### → 緊急性に影響する音響特徴量の検討

- 音声のF0, 話速を操作し, 各種統計量の計算による分析

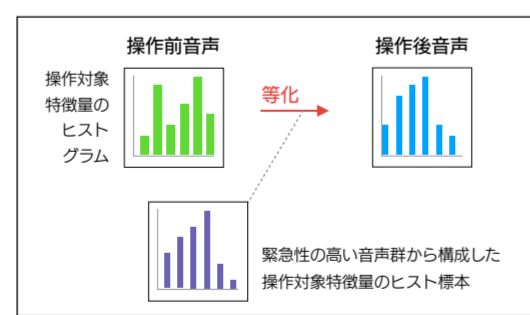
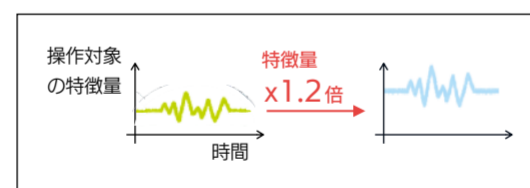
#### → 緊急性付与の変換規則を検討

- ① 平均シフト:  
音声群間のF0平均, モーラ時間長平均を用いた平均値の等化操作
- ② ヒストグラム等化(F0のみ):  
緊急時音声のF0累積分布関数を用いて操作前音声のF0分布を操作

音声群	音声の特徴
平静時音声	緊急性が低く, 災害と無関係の内容を発話している平静音声
加工対象音声	災害情報を提供し, 必要に応じて行動するよう(例えば, 備えや警戒)呼びかける平静音声
緊急時音声	緊急性が高く, 災害を避けるための即座な行動を求める音声

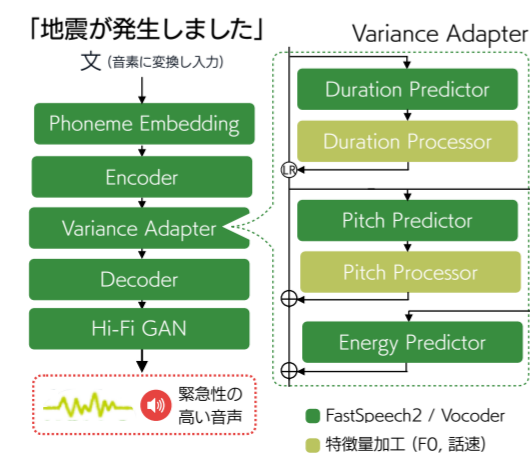
  

音声群	音声の緊急性	内容の緊急性
平静時音声	平静	平静
加工対象音声	平静	緊急
緊急時音声	緊急	緊急



### 学習済みニューラルTTSに対する緊急性操作

- 変換規則を学習済みニューラルTTSに適用
- FastSpeech2 [5] がもつ中間表現としての韻律特徴量を利用して操作



## 3. 実験的評価

### 評価手法

- 先行研究[3, 6]をもとに「信頼性」「聞き取りやすさ」「緊急性」を評価
- 5段階MOS評価による聴取実験
- 「信頼性」と「聞き取りやすさ」は呼びかけ音声としてのふさわしさの指標

### 利用する音声

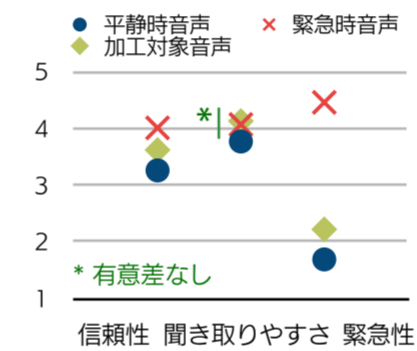
- NHKの同一アナウンサーによる読み上げ音声を収集

音声群	収集した音声
平静時音声	災害とは無関係なニュースを伝える読み上げ音声, 32個 (7分52秒)
加工対象音	台風情報と留意すべき点を伝えるニュースの読み上げ音声, 13個 (1分57秒)
緊急時音声	津波警報の発令と, 即時避難を求める臨時ニュースの音声, 97個 (5分9秒)

### 実験1: 各群音声における避難呼びかけ音声としての適性の評価

- 異なる緊急性の要素を持つ3音声群を評価
- 要素の違いから緊急性認知はどのように変化?

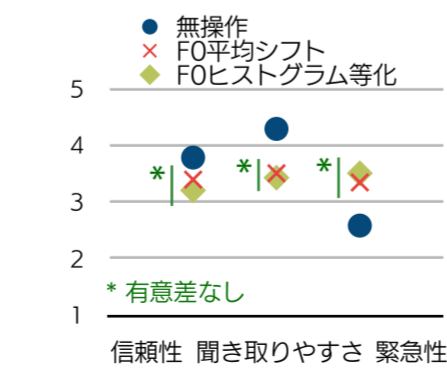
→ 緊急時音声は緊急性が高く, 加工対象音声と平静時音声は相対的に低いことが示唆



### 実験2-1: F0, 話速それぞれの操作による緊急性への影響差異

- 緊急性の異なる音声群間には音響特徴量差がある
- 特徴量差をもとに, 緊急性が低い加工対象音声のF0, 話速を操作すると, 緊急性はどのように変化?

→ 各操作による緊急性の向上が示唆



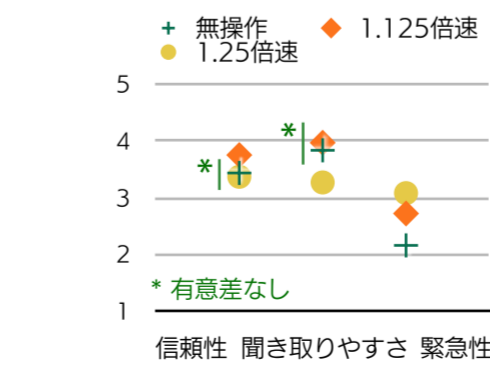
**F0操作: 操作により緊急性が向上**  
手法間による有意差なし  
ただし, 先行研究[7]では類似実験で有意差あり

音声群	話速 (モーラ時間長)			
	平均(s)	分散	歪度	尖度
平静時音声群	0.119	3.084E-03	1.710	0.119
緊急時音声群	0.106	1.940E-03	0.627	1.011

音声群	F0			
	平均(Hz)	分散	歪度	尖度
平静時音声群	128.69	1207.56	0.245	-0.842
緊急時音声群	189.59	1572.80	-0.271	-0.337

※平均シフト倍率: F0 は, 表の音声群間倍率 (1.47倍), 話速は先行研究 [3] を参考に設定 (1.125倍, 1.25倍)

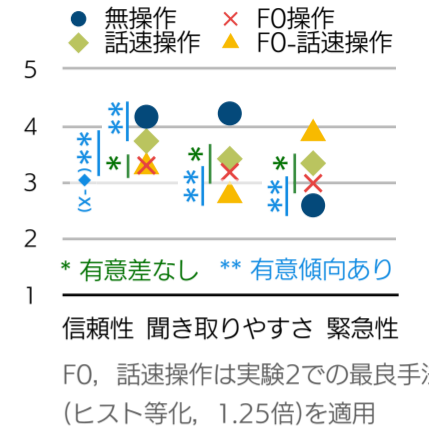


**話速操作: 倍速に応じて緊急性向上**  
1.125倍…緊急性とふさわしさが両立  
1.25倍…緊急性が最も改善

### 実験2-2: F0, 話速を組合せた操作による緊急性への影響差異

- F0操作, 話速操作, 両操作を行った音声のうち, どの操作が最良?

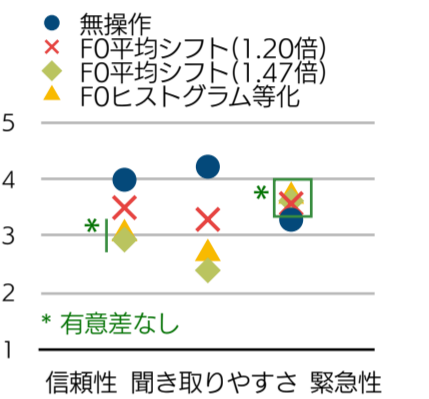
→ F0-話速操作(F0, 話速を操作)した音声の緊急性が最も高い



### 実験3: 学習済みニューラルTTSに対する緊急性操作の適用

- アナウンサーと同性の別話者音声で学習した, 学習済みニューラルTTSにF0-話速操作を実装 (話速は1.25倍固定)
- 緊急性の操作は, 学習済みニューラルTTSにも有効?

→ 操作手法によらず, 緊急性が向上



## 4. まとめ

1. 緊急性の異なる音声群間には, 音声特徴量の差があり, F0操作と話速操作は緊急性の付与に有効 [実験2-1]
2. F0-話速操作が, 緊急性付与において最も有効 [実験2-2]
3. 学習済みニューラルTTSに対しても, 同手法での操作は有効 [実験3]

### 今後の展望

- 学習済みニューラルTTSでの操作手法改善
- 行動評価等に踏み込んだ実験デザインの検討

### 参考文献

[1] 柿本 竜治, 吉田 護, “状況認識を考慮した令和2年7月豪雨時の避難行動意思決定モデルの推定,” 土木学会論文集 D3 (土木計画学), vol. 78, no. 2, pp. 45-57, 2022.

[2] 小林 まおり, 赤木 正人, “避難呼びかけ音声の心理的評価,” 日本音響学会誌, vol. 74, no. 12, pp. 633-640, 2018.

[3] Ofuji Kenta and Ogasawara Naomi, “Verbal disaster warnings and perceived intelligibility, reliability, and urgency: The effects of voice gender, fundamental frequency, and speaking rate,” Acoustical Science and Technology, vol. 39, no. 2, pp. 56-65, 2018.

[4] 伊藤 憲三, 田村 幸子, “防災行政無線に音声合成を用いるための最適制御法に関する検討,” 岩手県立大学地域政策研究センター地域協働研究研究成果報告集 1 【平成24年度教員提案型/地域提案型・前期】, pp. 32-33, 2013.

[5] Yi Ren et al., “FastSpeech 2: Fast and high-quality end-to-end text to speech,” vol. arXiv 2006.04558, 2020.

[6] 坂本 湧暉, “避難誘導音声における緊急感の操作を行うための音響特徴に関する研究,” 北陸先端科学技術大学院大学 2020年度修士論文, 2021.

[7] 小林 まおり 他, “音声の緊急感に影響する音響特徴の検討,” 信学技報, vol. 118, no. 149, pp. 79-84, 2018.