音声入力に基づく
合成音声の韻律制御システム

秋田大学　情報データ科学部　情報データ科学科
准教授　中島 佐和子

2026年3月3日

音声合成技術のトレンド：高音質化から可制御性へ

合成音声の質は大きく向上した。現在は、多様な音声属性を含む音声表現を、いかに柔軟かつ直感的に制御し、所望の音声表現を生成するかが研究課題の一つ。

Controllable TTS

- 明示的制御（Explicit control）
- 潜在表現の制御（Latent space control）
- プロンプトベース制御／自然言語による制御
  （Prompt-based / Natural-language control）

Xie, Tianxin, et al. "Towards controllable speech synthesis in the era of large language models: A systematic survey."
Proc. of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025.

従来技術とその課題

●プロンプトベース制御／自然言語による制御

自然言語による直感的な操作が可能で、
生成タスクや試行錯誤には非常に有効。

しかし、生成結果がモデル状態や更新に依存するため、
同一の話し方を安定して再現したり、演出判断を設計資産として蓄積・評価する点において難しい側面がある。

従来技術とその課題

● 明示的制御

既に実用化されているものに、単語やフレーズの読み上げ方を微調整するインタフェース等があるが、

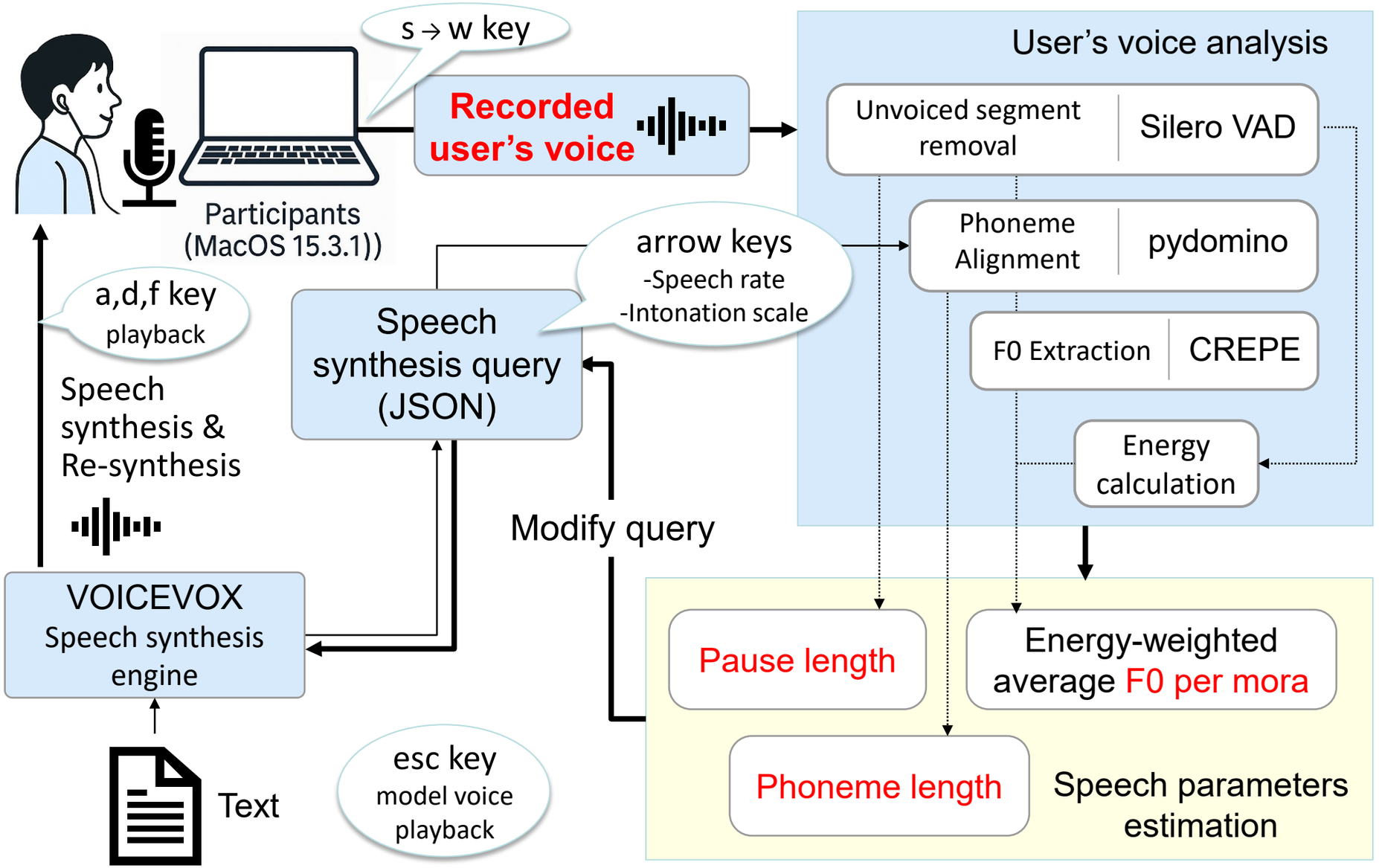
- モーラ毎の基本周波数、音素長、ポーズ長などを視覚的に表示し、それらのパラメタの数値をカーソル操作やキーボード入力により変更する方式

✓容易に操作できる直感的なインタフェースとは言い難い

✓視覚障害者にとってはアクセシビリティが極めて低い

新技術説明会
New Technology Presentation Meetings!

新技術による合成音声の韻律制御手法

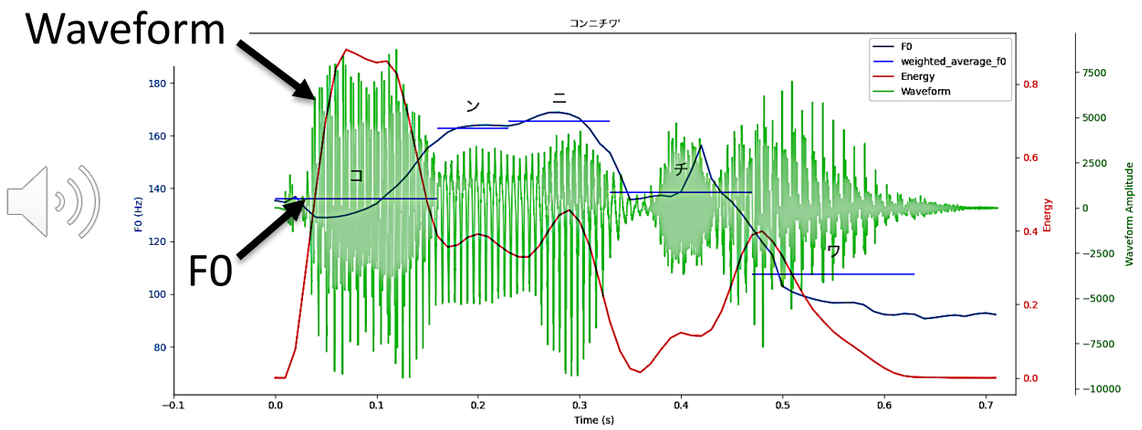


S. Nakajima et al. "Empowering People with Visual Disabilities in Audio Description Production: Development and Preliminary Evaluation of a User-Centred Synthesised Voice Adjustment Tool." *Proc. of the International Conference of the Association for the Advancement of Assistive Technology in Europe (AAATE), pp 399–404,* 2025.

新技術の特徴･従来技術との比較

- 利用者が音声合成エンジンにより合成された音声を聞いた上で、所望の韻律で音声合成するテキストを読み上げることで、合成音声の話者性を維持したまま読み上げ方を直感的に調整できるシステムを開発。
- ユーザは意図した読み上げ方を数値的に解釈する必要がなく、人間本来の発話動作に即して直感的に表現することができる。これにより、各パラメタの調整が容易になるだけでなく、数値的な表現が特に難しいパラメタ間の相互作用の調整も容易になる。特に、調整時間については大幅に削減できる。
- 人間の演出判断の保存・再利用・検証に発展可能。

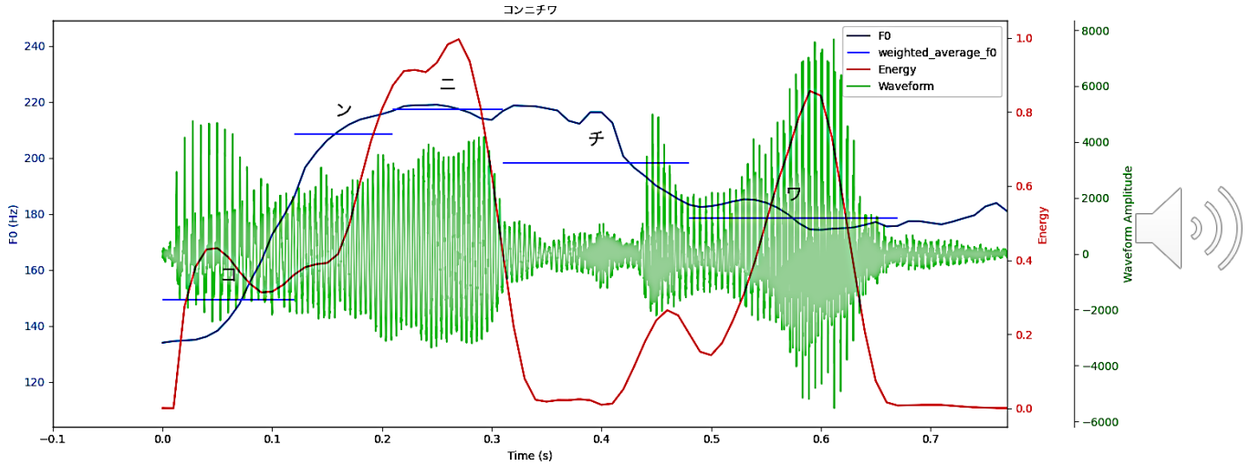
新技術説明会
New Technology Presentation Meetings!



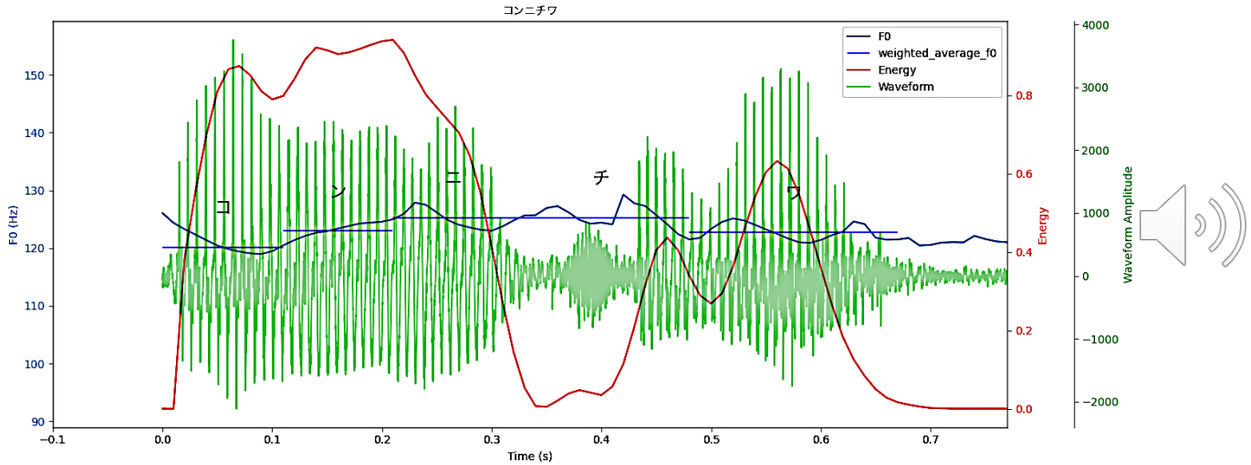
新技術によりトーン調整した合成音声

Neutral tone in Engine-Original

Cheerful tone



Somber tone

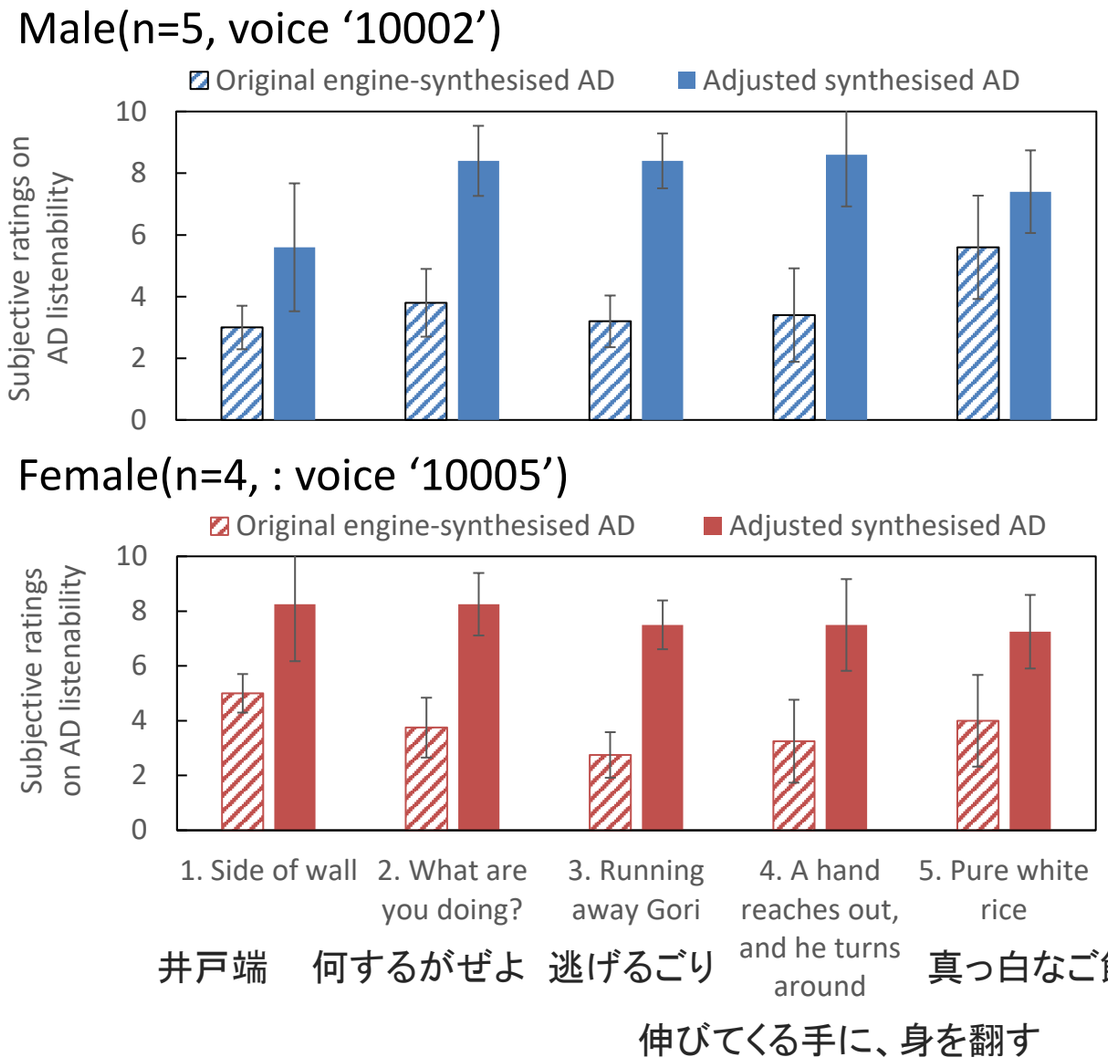


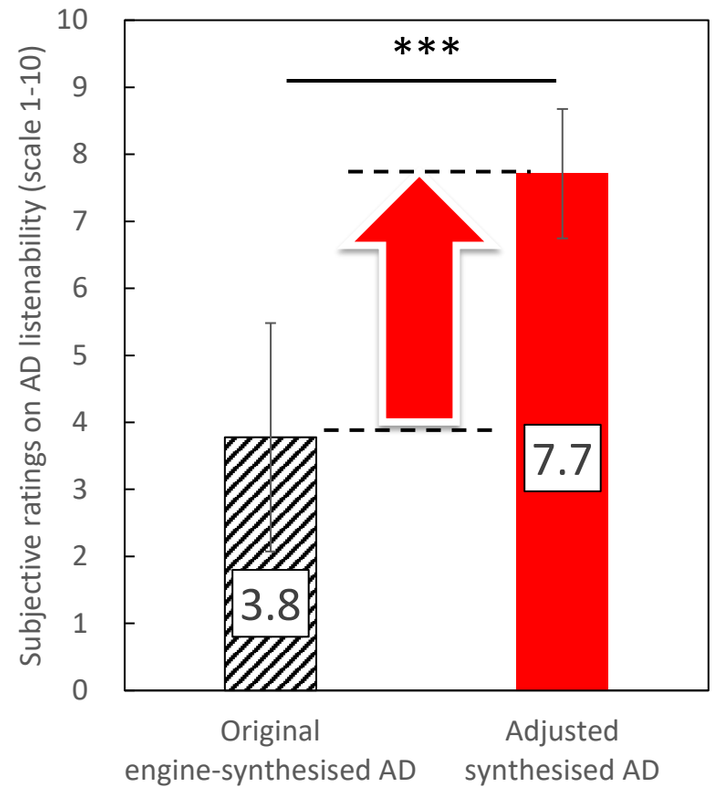
S. Nakajima et al., 2025.

Synthesised using VOICEVOX

本実験は，秋田大学手形地区における人を対象とした研究に関する倫理規定第12条第1項に基づいた倫理審査のうえ，秋田大学倫理審査委員会の承認を得て実施された．

新技術による合成音声の韻律調整結果





合成音声の聞き心地は約2倍向上

S. Nakajima et al., 2025.

本実験は，秋田大学手形地区における人を対象とした研究に関する倫理規定第12条第1項に基づいた倫理審査のうえ，秋田大学倫理審査委員会の承認を得て実施された.

想定される用途1

● 音声表現の高度化

直感的な韻律コントロール、話者性の保持と表現力の強化

● メディア制作

動画・広告・教育コンテンツ、ゲーム・インタラクティブ音声、
朗読・出版など

● アクセシビリティ & 公共分野

音声ガイド・読み上げ、公共アナウンスなど

● その他の応用

研究用途、バーチャルアシスタントなど

想定される用途2

人間の演出判断（間・抑揚・強調）を韻律として
「残す・引き継ぐ・再利用する」ためのツール

日本のコンテンツ産業では、

- 人間が演じる音声 → 「作品」「表現」「文化」
- 合成音声が使われる音声

→ 「量が多い」「コストが厳しい」「でも演出は必要」

- 「合成音声の領域に、人間の演出を“持ち込む”」
  - 演出を自動化しない
  - 演出を“共有可能な知識”にする
  - 日本語・日本文化の「演出の作法」を壊さないAIなど

実用化に向けた課題

調整精度と実行安定性

- 調整精度の向上
  - ✓ 指定した韻律（間・抑揚・強調）が合成音声上で意図どおり再現される必要がある。
  - ✓ エンジンや条件が変わっても結果が大きくブレないことが求められる。
- 微調整との併用
  - ✓ 演出判断を完全自動化しない。
  - ✓ 人間による微調整・修正を前提とした設計が必要。
  - ✓ 「調整 → 確認 → 微調整」というワークフローの確立

実用化に向けた課題

人間の演出判断を“使える形”で保存・利用する

1. 人間の演出判断を、どう正解データとして扱うか

   演出判断を、

   ① 自動化せず

   ② 人間の判断として残しつつ

   ③ ツールに取り込む仕組み

2. 演出意図と韻律操作の対応関係
3. 制作フローにどう組み込むか
4. 品質評価をどう定義するか

企業への期待

- 指定した韻律を合成音声上で意図どおりかつ安定して発声させるためには、音声合成基盤・レンダリング制御などに関する企業の技術・知見が不可欠である。
- 本研究では、人間の演出判断を正解データとして活用し、韻律指定と音声出力の対応関係について、企業と共同で検証・改善を行うことを希望する。

企業への貢献、PRポイント

- 本技術は、人間の演出判断を韻律として保存・適用できるため、合成音声でも日本語らしい話し方と演出意図を維持することが可能。企業が保有する音声合成技術による各種サービスにおける品質の向上や優位性に貢献できる。
- 導入検討時は、自然性・理解度・疲労などの評価実験から、科学的根拠に基づく有効性検証を支援することが可能。
- 本格導入に向けて、韻律設計・評価方法・運用フローに関する技術支援や知見共有を行い、現場で実際に活用できる形での導入を支援。

本技術に関する知的財産権

- 発明の名称　：　音声調整装置及びプログラム
- 出願番号　：　特願2026-017457
- 出願人　：　秋田大学
- 発明者　：　中島佐和子、水戸部一孝

お問い合わせ先

秋田大学　未来研究統括機構
イノベーションオフィス

ＴＥＬ　018－889 － 2712
e-mail　staff@crc.akita-u.ac.jp