

REALTIME TRANSFORMATION OF VOICE IDENTITY AND STYLE

Tech ID: 34328 / UC Case 2026-055-0

PATENT STATUS

Patent Pending

BRIEF DESCRIPTION

Converting voice identity in real-time while maintaining perfect linguistic clarity and emotional nuance is a significant hurdle in speech synthesis. Researchers at UC Berkeley have developed a system for real-time voice style conversion that transforms a source speaker's speech to match the timbre, accent, and emotion of a target speaker. The technology utilizes a content extraction network with conformer blocks and a unique low-dimensional quantization method—using fewer than 100 levels—to preserve linguistic fidelity. By extracting continuous representations before quantization, the system maintains higher speech quality than traditional discrete methods. A diffusion-based generation network then creates a mel-spectrogram conditioned on these features and a target style embedding, which is finally converted to audio via a vocoder. The system is designed for streaming operation through the use of chunked-causal attention mechanisms, enabling near-instantaneous transformation.

SUGGESTED USES

»

Entertainment and Gaming: Allowing players or actors to adopt the voices of specific characters in real-time with full emotional expression.

»

Localization and Dubbing: Converting the voice of a foreign-language speaker to a target voice while preserving their original acting performance and accent.

»

Call Center Personalization: Harmonizing the vocal timbre of agents to a specific brand identity while maintaining their natural speech patterns.

»

Assistive Communication: Helping individuals with speech impairments or vocal cord damage communicate in their original voice or a chosen identity.

»

Privacy and Anonymization: Protecting the identity of speakers in sensitive contexts by transforming their voice to a consistent, non-identifiable target.

ADVANTAGES

»

Real-Time Streaming: Integrated chunked-causal attention allows for low-latency processing, making it suitable for live conversations and interactive media.

»

High Linguistic Fidelity: The use of continuous representations before the quantization bottleneck ensures that words and syllables remain clear and accurate.

CONTACT

Michael Cohen
mcohen@berkeley.edu
tel: 510-643-4218.



INVENTORS

» Anumanchipalli, GopalaKrishna

OTHER INFORMATION

CATEGORIZED AS

» **Communications**

» Other

» **Computer**

» Software

RELATED CASES

2026-055-0

»

Nuanced Style Transfer: Captures and replicates subtle characteristics such as specific accents and emotional states, moving beyond simple pitch shifting.

»

Efficient Modeling: The low total number of quantization levels across dimensions allows for a highly compressed yet expressive content representation.

»

Flexible Identity Control: Can adapt to a wide range of target speakers by simply extracting a style embedding from a short audio sample.

RELATED MATERIALS

ADDITIONAL TECHNOLOGIES BY THESE INVENTORS

- ▶ [Realtime Transformation Of Voice For Privacy Protection](#)
- ▶ [Methods For Dysfluent Speech Transcription And Detection](#)
- ▶ [Learning Multimodal Sim-To-Real Robot Policies With Generative Audio](#)
- ▶ [Articulatory Feedback For Phonetic Error-Based Pronunciation Training](#)



University of California, Berkeley Office of Technology Licensing

2150 Shattuck Avenue, Suite 510, Berkeley, CA 94704

Tel: 510.643.7201 | Fax: 510.642.4566

<https://ipira.berkeley.edu/> | otl-feedback@lists.berkeley.edu

© 2026, The Regents of the University of California

[Terms of use](#) | [Privacy Notice](#)