LOW-LATENCY PITCH-SHIFTING WITH STN-DECOMPOSITION

MARIA POLAK, CUMHUR ERKUT
mpolak@student.aau.dk
Aalborg University Copenhagen, Denmark

INTRODUCTION

Pitch-shifting is a widely used effect in music production, vocal correction, and sound design, allowing for pitch modifications without altering tempo. While offline pitch-shifting methods can achieve high quality, real-time solutions face challenges such as latency and audio artifacts, particularly in polyphonic signals.

KEY CONTRIBUTIONS

- **Fuzzy STN and Noise Morphing (NM) in Pitch-Shifting:** Introducing fuzzy STN decomposition [1] and Noise Morphing [2] into processing to improve pitch-shifting quality
- **Online Implementation:** A real-time implementation of the proposed algorithms, suitable for musical audio signals

METHODS

HOW DOES IT WORK?

This project incorporates STN decomposition into the pitch-shifting pipeline and applies specialized processing to each extracted component. The separated audio streams—sines, transients, and noise—are processed as follows:

- **Sines:** Processed with a Phase Vocoder derivative (Vase-Phocoder)
- **Transients:** Reapplied post-processing to maintain integrity
- **Noise:** Time-scaled by Noise Morphing algorithm and then resampled

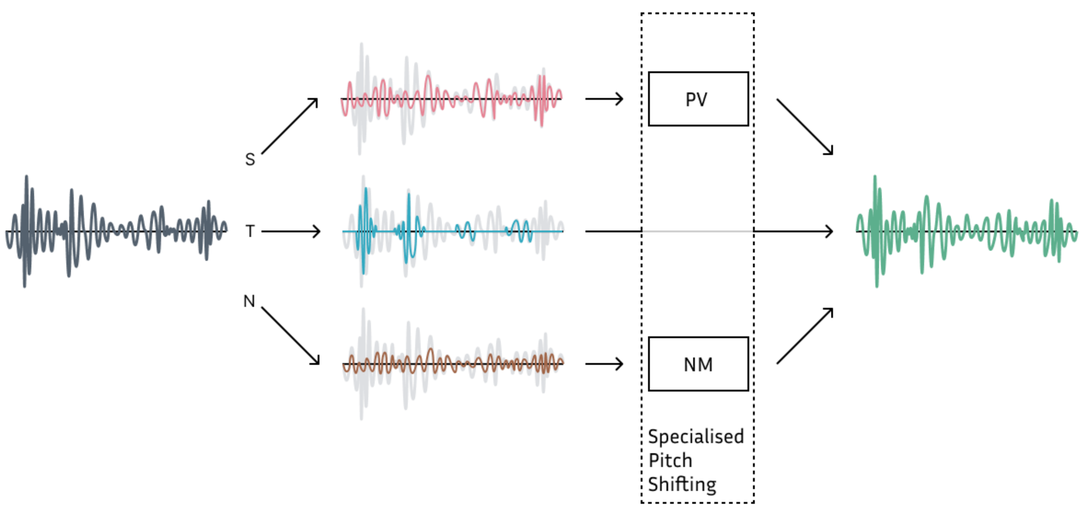

Figure 1. System overview. The signal is decomposed into Sines, Transients and Noise streams, and each stream is processed separately.

HOW FAST IS IT?

- Operates under latency threshold of *166.1ms*, aligned with commercial pitch-shifting solutions.
- Internal algorithms' adjustments for online processing:
  - **Median Filtering:** Fuzzy STN decomposition's horizontal median filters were adjusted to operate only on past frames, ensuring real-time execution without needing future data.
  - **Frames interpolation:** NM's frame interpolation was modified to generate evenly spaced intermediate frames, enabling smooth time-stretching without requiring re-indexing of original frames.

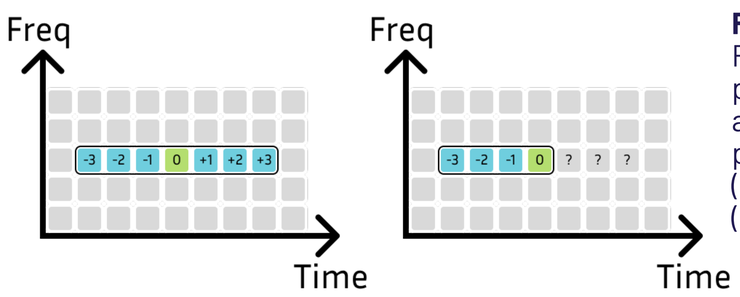

Figure 2. Horizontal Median Filter modification. In online processing no information about future signal is provided.
(Left) Original
(Right) Online processing

RESULTS

WHAT DO THE PEOPLE SAY?

- **Blind listening test** compared the proposed method (PS) with:
  - Elastique Pitch V2 (EL)
  - Waves SoundShifter (WS)
  - Signalsmith Stretch (SS)
  - Simple resampling (AN).
- Mean Opinion Score (MOS) as a **metric** on perceived pitch-shift quality

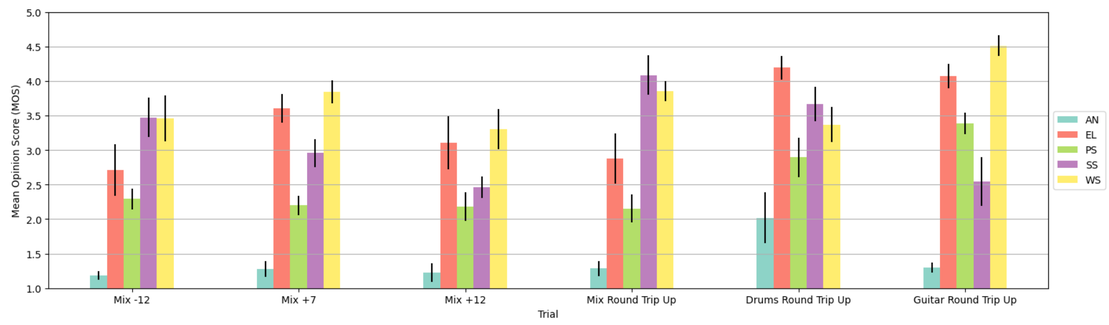

Figure 3. Listening test results - MOS with 95% confidence intervals.

Participants:

- 50% - said the stimuli sounded very similar to each other
- 38% - liked the sound of the artifacts present in some stimuli
- 25% - would put a higher rating on *"bad"* sounding conditions, if quality was not to be rated, as they *"developed a taste to it"*

INTERVIEW HIGHLIGHTS

LET'S TAKE A LOOK (SINCE WE CAN'T LISTEN)

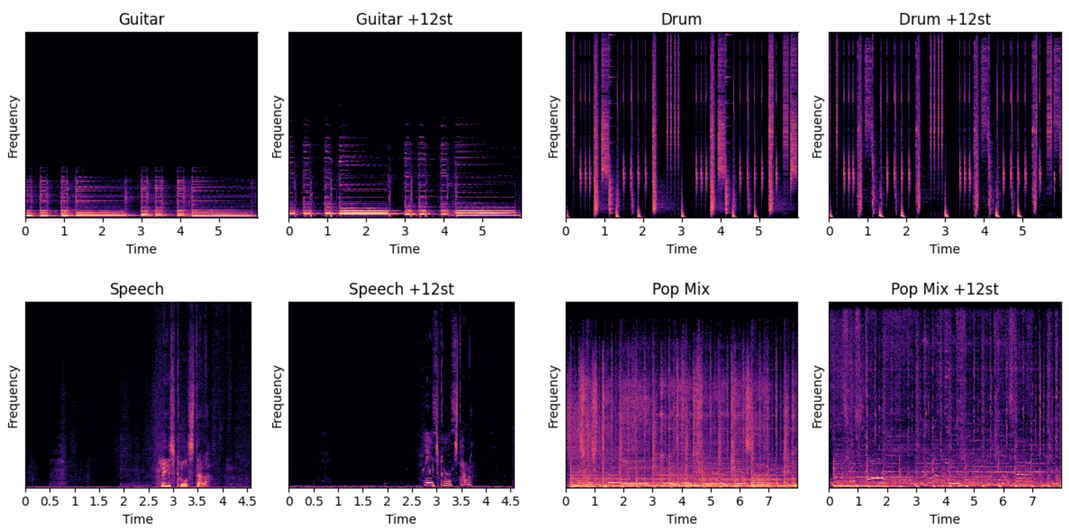

Figure 4. Signals pitch-shifted with the proposed method. From left to right - Guitar Loop, Drums Loop, Speech Signal, Pop Song

CONCLUSIONS

- System balances latency, quality, and computational efficiency.
- While it falls short of commercial state-of-the-art methods, qualitative analysis suggests potential use in emerging music trends.
- Future work will refine decomposition, optimize performance, and explore additional evaluation metrics.

REFERENCES

- [1] Fierro and Välimäki, "Enhanced Fuzzy Decomposition of Sound Into Sines, Transients, and Noise."
- [2] Moliner et al., "Noise Morphing for Audio Time Stretching."