CONFORMER-BASED ULTRASOUND-TO-SPEECH CONVERSION

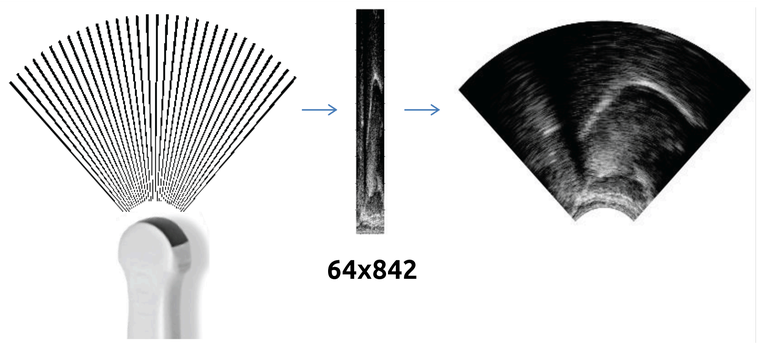
Ibrahim Ibrahimov, Csaba Zainkó, Gábor Gosztolya

1. DATASET

UltraSuite-TaL80

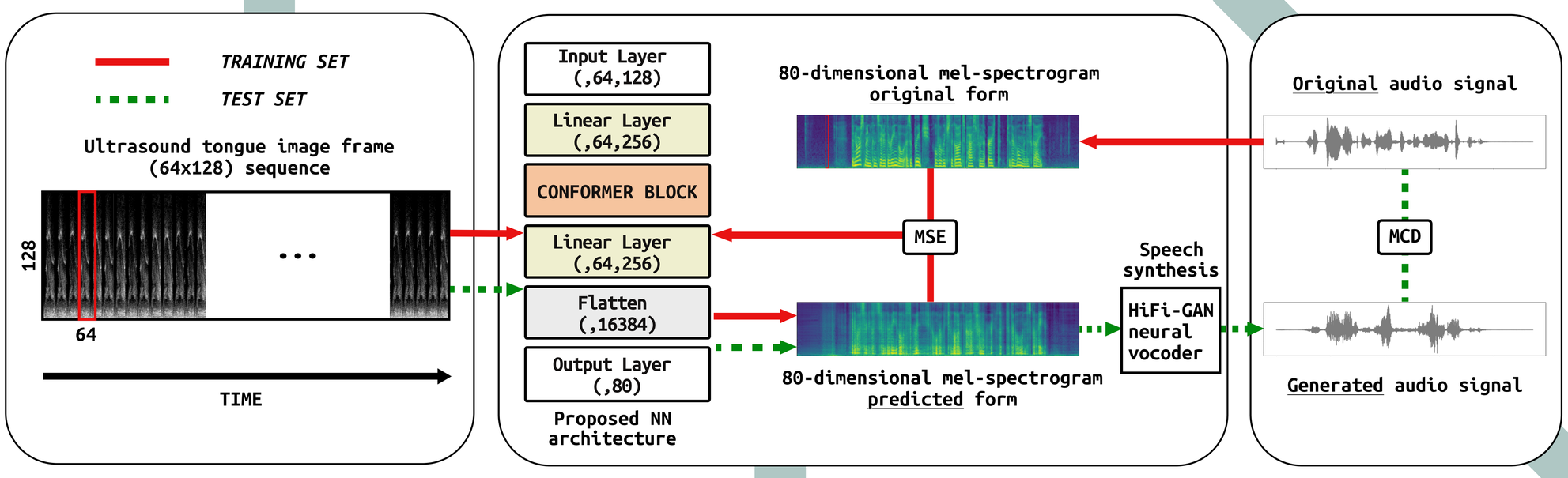
spkr ID	gender	# of samples
01fi	female	204
02fe	female	141
03mn	male	193
04me	male	190

ultrasound scanline, raw, wedge forms



data pre-processing

Ultrasound scanline (64 × 842 px) ▶ **Bicubic interpolation** ▶ Resized to 64 × 128 px ▶ **Pixel normalization [-1, 1]** ▶ READY TO GO!



3. OBJECTIVE RESULTS

↓MSE (mel-spec generator)

	Baseline	Conformer BASE	**Conformer with bi-LSTM**
01fi	**0.46**	0.51	**0.48**
02fe	0.62	0.62	**0.58**
03mn	0.39	0.46	**0.38**
04me	0.46	0.52	**0.45**

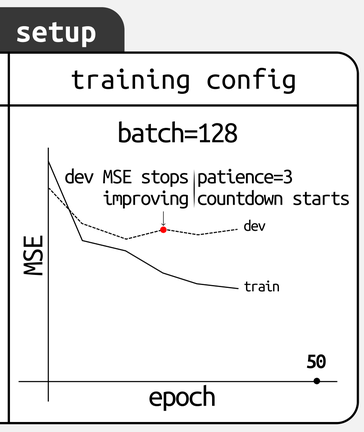
↓MCD (speech synthesis)

	Baseline	Conformer BASE	**Conformer with bi-LSTM**
01fi	**3.22**	3.51	**3.25**
02fe	**3.01**	3.12	**3.04**
03mn	**3.64**	4.13	**3.71**
04me	**3.17**	3.26	**3.25**

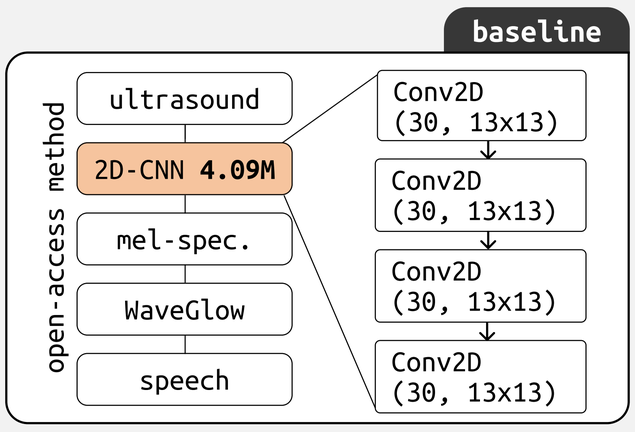
2. METHODS

experimental setup

data split	training config
005_xaud … 015_xaud — test	batch=128
90% — train	dev MSE stops improving | patience=3 countdown starts
10% — dev	MSE, epoch, 50



baseline



CONFORMER BLOCK
- feed forward module
- multi-head self attention module
- convolution module
- feed forward module

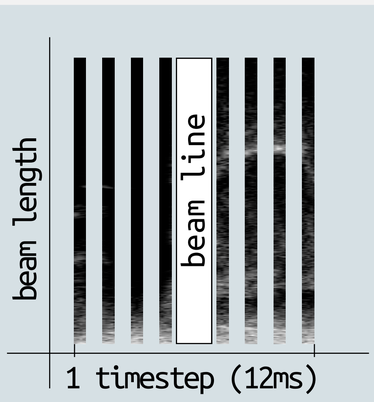
↓

bi-LSTM

↓

bi-LSTM

conformer-BASE



frame level learning

TOTAL 64 x (# of timesteps)

INPUT 64x128 frames in consecutive order

2.06M param
3 times less time than baseline

with bi-LSTM

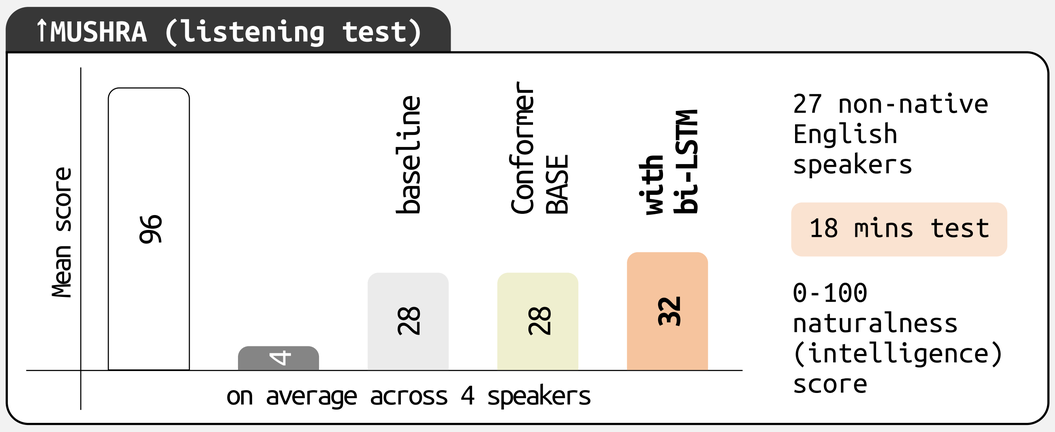
sequence level learning using LSTM structure	**5.35M** param
Inspired by J. Kwon's work on acceloremeter-to-speech synthesis	**1.5 times less time** than baseline

Neural Vocoder - Speech Synthesis

For fair comparison utilized vocoder: HiFi-GAN | VCTK -V1

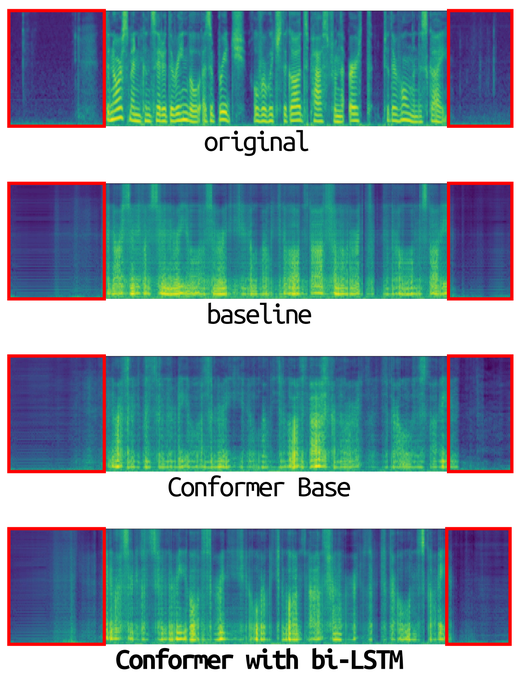
4. SUBJECTIVE RESULTS

↑MUSHRA (listening test)



Visual outro - Discussion

Visual inspection of mel-spectrograms
speaker "01fi" - sample "014_xaud"



MAIN TAKE-AWAY OF THE PAPER

CONFORMER-BASE
- 3 times faster
- as good as baseline

CONFORMER WITH BI-LSTM
- 1.5 times faster
- significantly better than baseline

BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
UNIVERSITY OF SZEGED
RGAI