This research was supported by the Swedish Research Council projects Connected (VR-2019-05003), Perception of speaker stance (VR-2020-02396), the Riksbankens Jubileumsfond project CAPTivating (P20-0298).

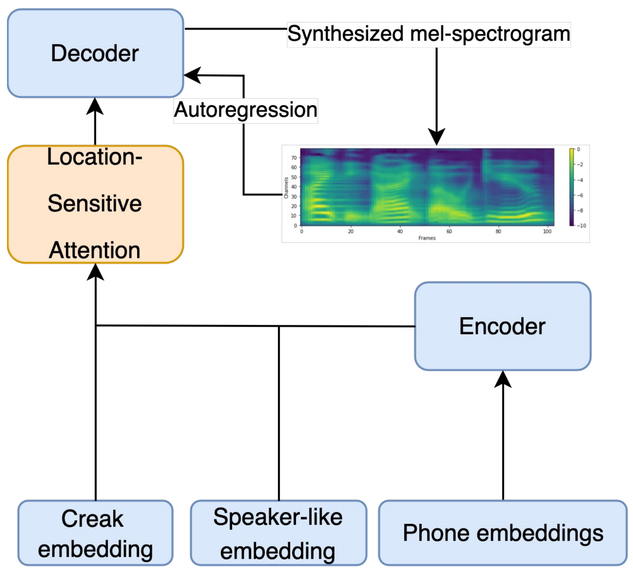
Samples

The Role of Creaky Voice in Turn Taking and the Perception of Speaker Stance: Experiments Using Controllable TTS

Harm Lameris, Éva Székely, Joakim Gustafson
Division of Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

1. **Synthesized creaky voice** for a speech perception study examining **certainty, valence, sarcasm**, and **turn finality**
2. Automatic annotation of a spontaneous speech dataset for creaky voice using CreaPy [1] and DeepFry [2]
3. Added creak conditioning to Tacotron 2 [3] to control presence and duration of creak at synthesis
4. Analysed the presence of creaky voice in the stimuli

Architecture



Properties of creaky voice

- Low F_0 (irregular)
- Articulation
  - Constricted glottis
  - Low glottal airflow
- Positional and non-positional creak

Perception of creaky voice

- Used as a hedge (uncertainty) [4]
- Negative [5]
- Sarcastic [6]
- Turn-yielding [7]

Data and annotation

- AptSpeech [4]
  - 5h40m spontaneous speech
  - 2h26m of read speech (Arctic corpus)
- DeepFry & CreaPy
- creak percentage = $\frac{creak\ duration}{total\ word\ duration}$

Creak annotation	Chosen value
DeepFry>0 & CreaPy>0	Highest value
DeepFry>0 or CreaPy>0	Highest if value>0.1 else 0
DeepFry=0 & CreaPy=0	0

Stimuli creation

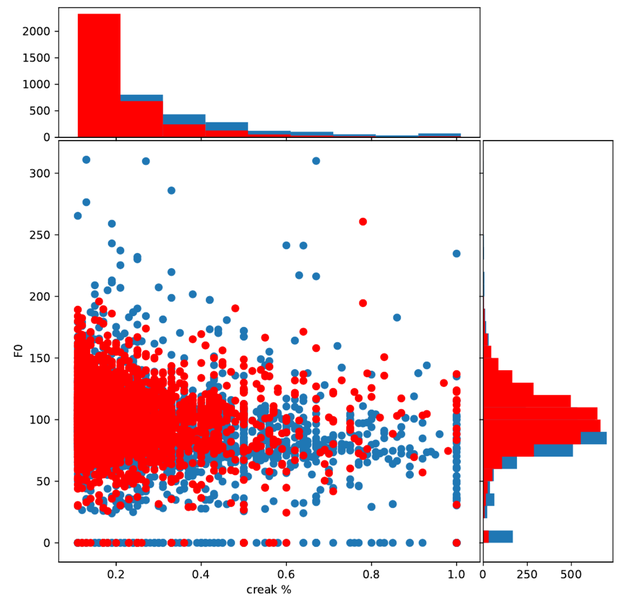
Non-positional

(1) A trip to the desert
(2) 200 ms silence
(3) *isn't endless sand and scorching sun a cool experience?*

Positional

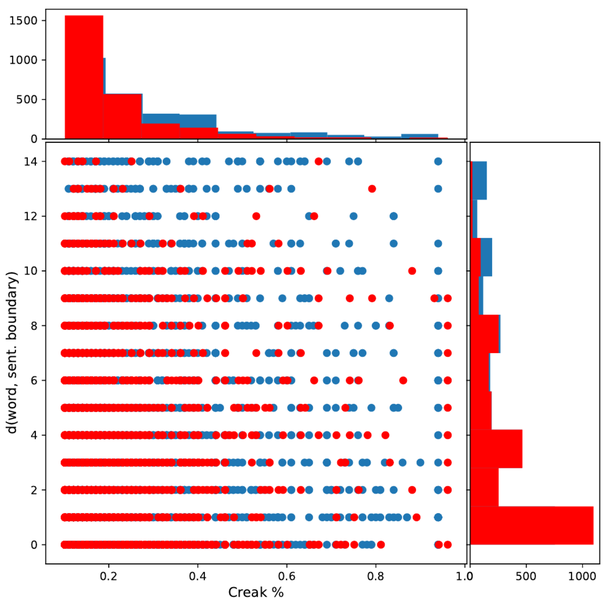
(1) A trip to the desert
(2) 200 ms silence
(3) it's a way to experience *nature's beauty*

Italics indicate creaky phonation



Creak percentage vs F_0

- Spontaneous speech is more varied
  - Creak percentage
  - F_0



Creak percentage vs sent bound

- Spontaneous speech is more varied
  - Creak percentage
  - Creak distribution

KTH Royal Institute of Technology

Objective evaluation

- Creak percentage of stimuli

Creak type	Creak percentage
No creak	0.04±0.03
Positional creak	0.09±0.06
Non-positional creak	0.13±0.08

Subjective evaluation

- 25 participants — 1-7 Likert
- Non-positional creak vs modal
- Positional creak vs. modal

- How **certain/positive/sarcastic** does the speaker sound?
- How much does the speaker sound like he is **done talking**?

Results

Non-positional creak

Question	Non-positional	Modal
Certainty	5	**5**
Valence	4	**4**
Sarcasm	4	4
Turn finality	**6**	6

Positional creak

Question	Positional	Modal
Certainty	5	5
Valence	5	4
Sarcasm	2	2
Turn finality	**6**	6

Bold indicates significantly higher rating (Wilcoxon signed-rank)

Non-positional creak

- **Less** certain
- **Less** positive
- **More** turn final

Positional creak

- **More** turn final

References

[1] Paierl, M., Röck, T., Wepner, S., Kelterer, A., & Schuppler, B. (2023). Creapy: A python-based tool for the detection of creak in conversational speech. In *Proc. ICPhS*.
[2] Chernyak, B. R., Simon, T. B., Segal, Y., Steffman, J., Chodroff, E., Cole, J. S., & Keshet, J. (2022). DeepFry: Identifying vocal fry using deep neural networks. in *Proc. Interspeech*, pages 3578–3582.
[3] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., et al. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. In *Proc. ICASSP*.
[4] Lee, S. (2015). Creaky voice as a phonational device marking parenthetical segments in talk. *Journal of Sociolinguistics*, 19(3), 275–302.
[5] Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, 40(1-2), 189–212.
[6] Fónagy, I. (1981). Emotions, voice and music. *Research aspects on singing*, 33, 51–79.
[7] Włodarczak, M., & Heldner, M. (2022). Contribution of voice quality to prediction of turn-taking events. In *Proc. Speech Prosody*, pages 485–489.
[8] Lameris, H., Wlodarczak, M., Gustafson, J., & Székely, É. (2023b). Neural speech synthesis with controllable creaky voice style. in *International Congress of Phonetic Sciences (ICPhS)*, pages 3141–3145.