Credibility Through Non-native Language Varieties in Conversational Pedagogical Agents

PhD Project Presentation

Sven Albrecht
sven.albrecht@phil.tu-chemnitz.de

TU Chemnitz

08.09.2021

HYBRID SOCIETIES

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

About me



- BA & MA in English and American Studies from TU Chemnitz
- worked as vocational school teacher in Germany
- worked as high school teacher in China
- currently working at TUC as part of the DFG funded CRC **Hybrid Societies** and the Erasmus+ project **TEFL-ePAL**

Objectives

Mission

In hybrid societies, humans and embodied digital technologies should interact as seamlessly as humans among each other.

RQ1 Which specific non-native linguistic cues of CPAs influence the learning performance of non-native human learners?

RQ2 Which specific non-native linguistic cues influence attributed credibility and acceptance of CPAs by non-native human learners?

RQ3 How much does a linguistically credible CPA influence the learning performance in non-native educational contexts?

Pilot Study

Sociolinguistic interview with Chinese PhD student at TUC

- reading passages
- wordlist
- interview questions based on Tagliamonte (2006, Appendix B) supplemented with target group specific questions

Transcription of data

- orthographic (sentences, words)
- IPA (phonemes)

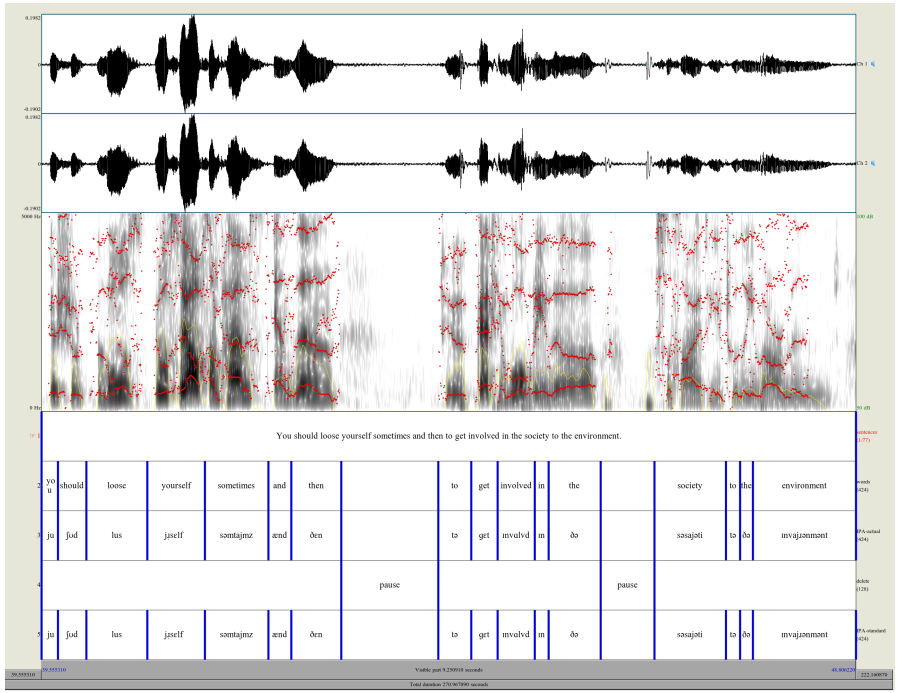


Figure 1: Transcription Screenshot

TTS System

Goal

A TTS synthesis system that can synthesize English text in different Chinese accents.

In the synthesized speech we want to control the following features:

- morphosyntactic cues e.g. syntax, grammar
- phonetic cues e.g. pronunciation of phonemes
- prosodic cues e.g. stress, intonation

Our TTS system is based on two different models and uses transfer learning.

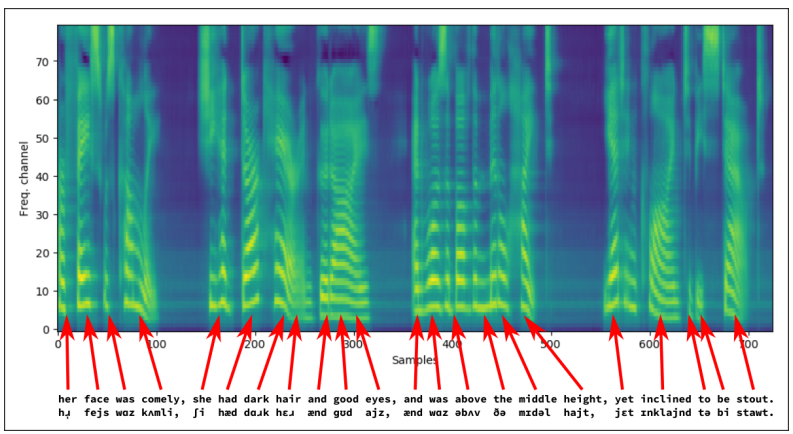


Figure 2: Example output mel-spectrogram

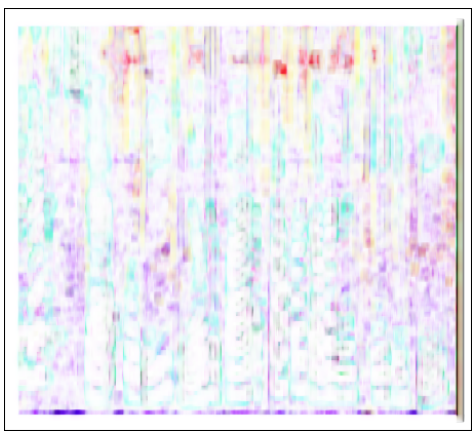


Figure 3: Difference original and synthesized audio

TTS System

Currently we are able to control:

- morphosyntactic cues with a rule based approach
- phonetic cues with a phone-based TTS (based on Tacotron2 by Shen et al., 2018)

We developed some helpful tools for speech synthesis:

- for recordings: e.g. resampling, automatic detection of silence
- for text: e.g. G2P conversion, symbol mapping

Audio Examples

`https://stefantaubert.github.io/tacotron2/`

Preparation of Data Collection

Challenge

Corpus linguistic literature (e.g. Love, Dembry, Hardie, Brezina, & McEnery, 2017) and computer science literature (e.g. Bozkurt, Ozturk, & Dutoit, 2003) propose different corpus creation criteria.

Sociolinguistic methodology has to be adapted to TTS application:

- previous TTS systems trained on reading passages
- selection of reading passages by selecting sentences according to phone and diphone coverage using greedy selection (Taubert, Sternkopf, Kahl, & Eibl, fc.)
- supplementary word list (i.e. missing phonemes/diphones)
- interview questions
  - many questions from Tagliamonte (2006) not relevant
  - add questions about high school and university life

Data Collection and Covid-19

Challenge

International travel impossible, Chinese visa application suspended indefinitely, yet data collection through field work is crucial for the project's success.

Alternative forms of data collection:

- send recording equipment to Chinese partner university and conduct online/hybrid interviews
- equipment stuck in customs for months
- include other varieties, in particular Nigerian English
- leverage social network of visiting scholars from Nigeria (in-group interviews)

Collecting data for TTS is relatively easy, collecting data for TTS and linguistic analysis is challenging.

Measuring Linguistic Credibility

Assumptions:

- some linguistic features contribute more to perceived credibility than others
- the NN TTS systems allows fine grained control of linguistic features
- higher linguistic credibility leads to better learning outcomes (c.f. Rey & Steib, 2013)

Learning outcomes will be measured in participant studies (post Covid):

- prior knowledge, interest (could be confounding variables)
- retention, i.e. remembering facts
- transfer, i.e. applying knowledge to new tasks

Note

The stimuli will be based on factual knowledge about a subject, no language learning.

Measuring Vowel Spaces

Quantification workflow (WIP)

- Forced alignment using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017)
- Automated vowel formant measurements in Praat
- Vowel plots generated in R
- Hampel filtering of outliers (Hampel, 1974)
- speaker intrinsic, vowel extrinsic, formant intrinsic normalization (Lobanov, 1971)

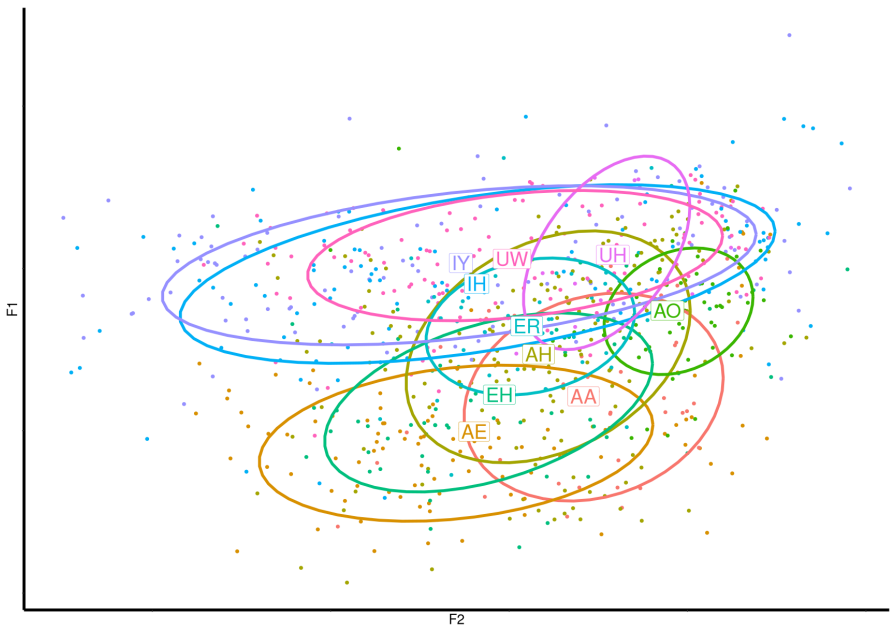


Figure 4: Vowel space plot of pilot test subject

Next Steps

Completing work packages:

- baseline psychological experiments
- running TTS system with very limited input data
- developing a linguistic TTS quality metric
- data collection
- testing the significance of intercultural factors influencing credibility and learning
- exploring methods of measuring credibility and learning performance

Planned publications:

- Current research on the linguistic features of Chinese English (in World Englishes, under review)
- Linguistic Credibility as a Key Issue in the Communication of Humans and Humanoids in a Hybrid Society
- Methodological Considerations - Comparison of online/distance vs. fieldwork/interviews

References

Bozkurt, B., Ozturk, O., & Dutoit, T. (2003). Text design for TTS speech corpus building using a modified greedy selection. In *Eighth european conference on speech communication and technology*.

Hampel, F. R. (1974, June). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*(346), 383–393. doi: 10.1080/01621459.1974.10482962

Lobanov, B. M. (1971, February). Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608. doi: 10.1121/1.1912396

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3), 319–344.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Interspeech* (Vol. 2017, pp. 498–502).

Rey, G. D., & Steib, N. (2013). The personalization effect in multimedia learning: The influence of dialect. *Computers in Human Behavior*, *29*(5), 2022–2028.

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., … others (2018). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4779–4783).

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Taubert, S., Sternkopf, J., Kahl, S., & Eibl, M. (fc.). A Comparison of Text Selection Algorithms for Sequence-to-Sequence Neural TTS. In *NeurIPS 2021*.

Credibility Through Non-native Language Varieties in Conversational Pedagogical Agents

PhD Project Presentation

Sven Albrecht
sven.albrecht@phil.tu-chemnitz.de

TU Chemnitz

08.09.2021

HYBRID SOCIETIES

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation