CRC 1410 – Project D03

DEVELOPING A SPEECH SYNTHESIS TOOLKIT FOR NON-NATIVE CREDIBLE CONVERSATIONAL PEDAGOGICAL AGENTS

Prof. Maximilian Eibl[c], Prof. Günter Daniel Rey[b] & Prof. Josef Schmied[a] | Sven Albrecht[a] Rewa Tamboli[b] Stefan Taubert[c]

a Professorship of English Language and Linguistics, Institute for English and American Studies, Faculty of Humanities CUT · b Professorship of Psychology of Learning with Digital Media, Institute for Media Research, Faculty of Humanities, CUT · c Professorship of Media Informatics, Faculty of Computer Science, CUT

GOAL

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?
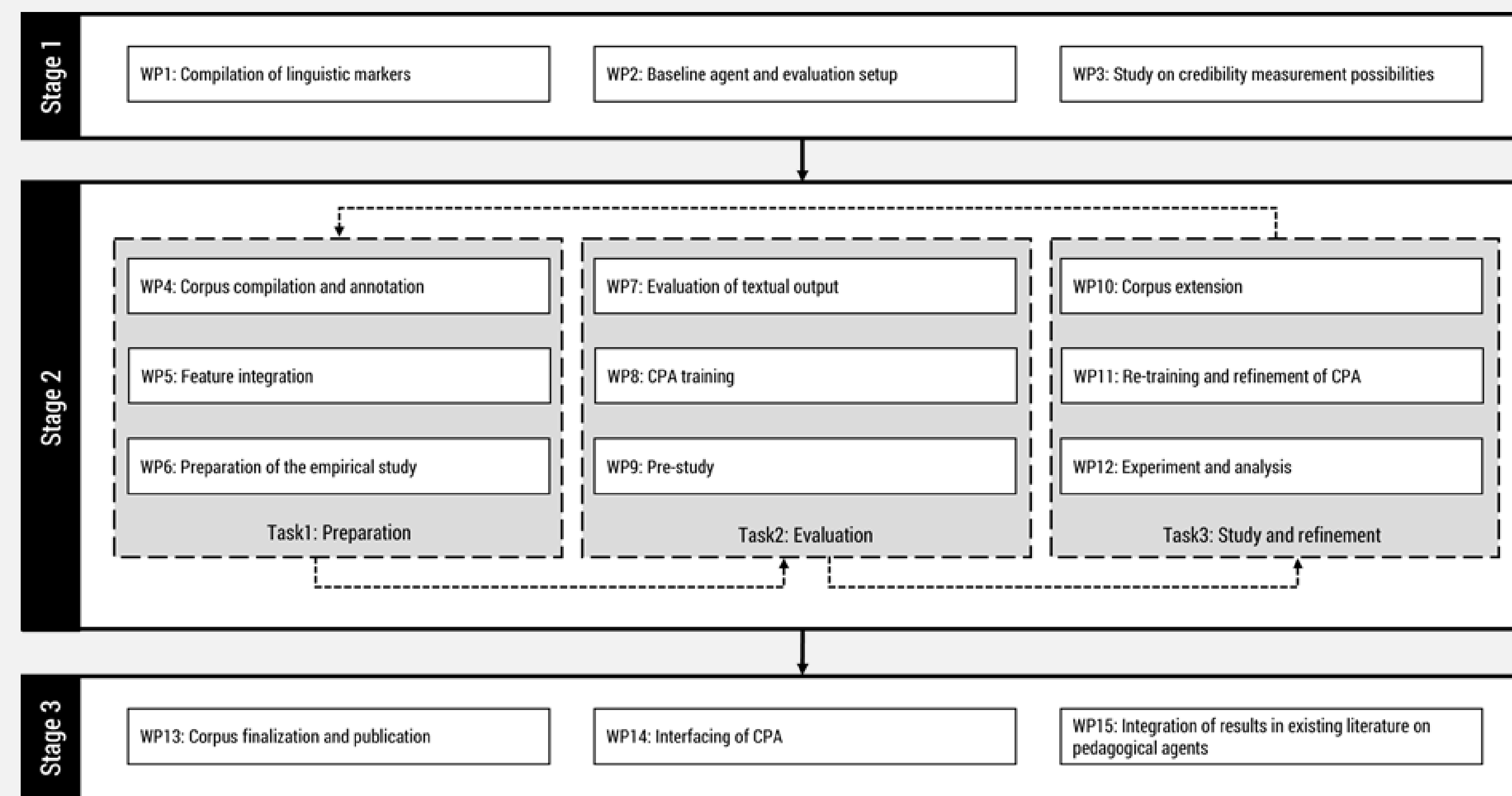

Figure 1: D03 Work Plan

PILOT

Sociolinguistic interview with Chinese PhD student at TUC:

- Reading passages
- Word list
- Interview questions based on Tagliamonte (2006, Appendix B) supplemented with target group specific questions

Transcription of data:

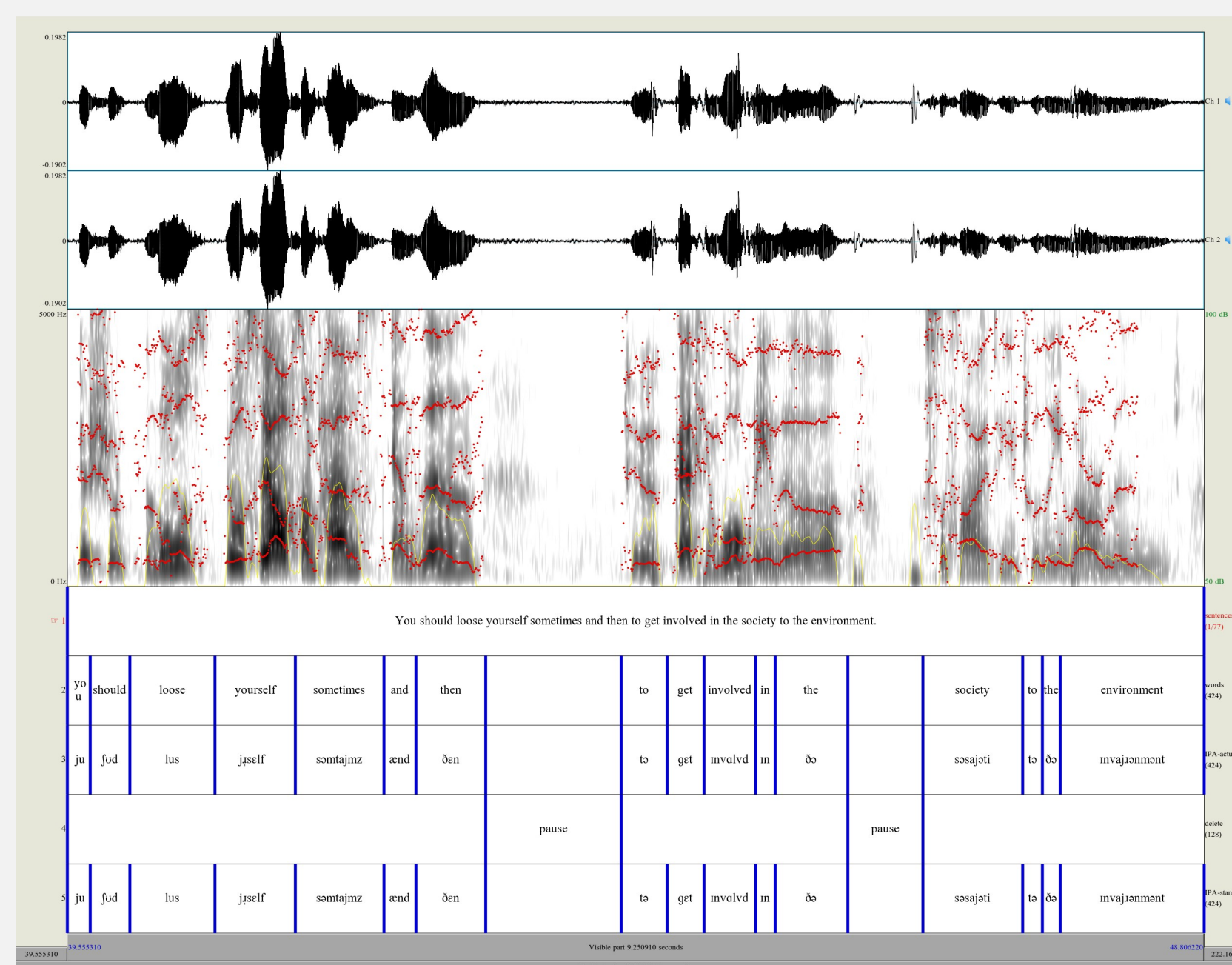- Orthographic
  · Sentences
  · Words
- IPA (phonemic/phones)

Figure 2: 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 which can be very different between accents:

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

Currently we are able to control:

- morphosyntactic cues with a rule based approach
- phonetic cues with a phone-based TTS (based on Tacotron 2 by Shen et al. 2018)

Our TTS system is based on two separate models (see Figures 3, 4, 5).

- usage of transfer learning

We developed some helpful tools for speech synthesis:

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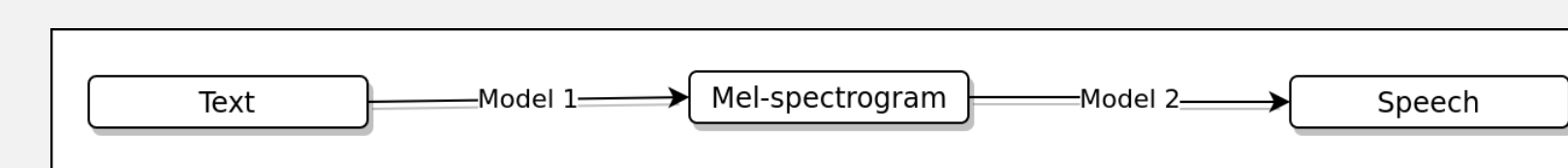- for text: e.g. G2P conversion, symbol mapping
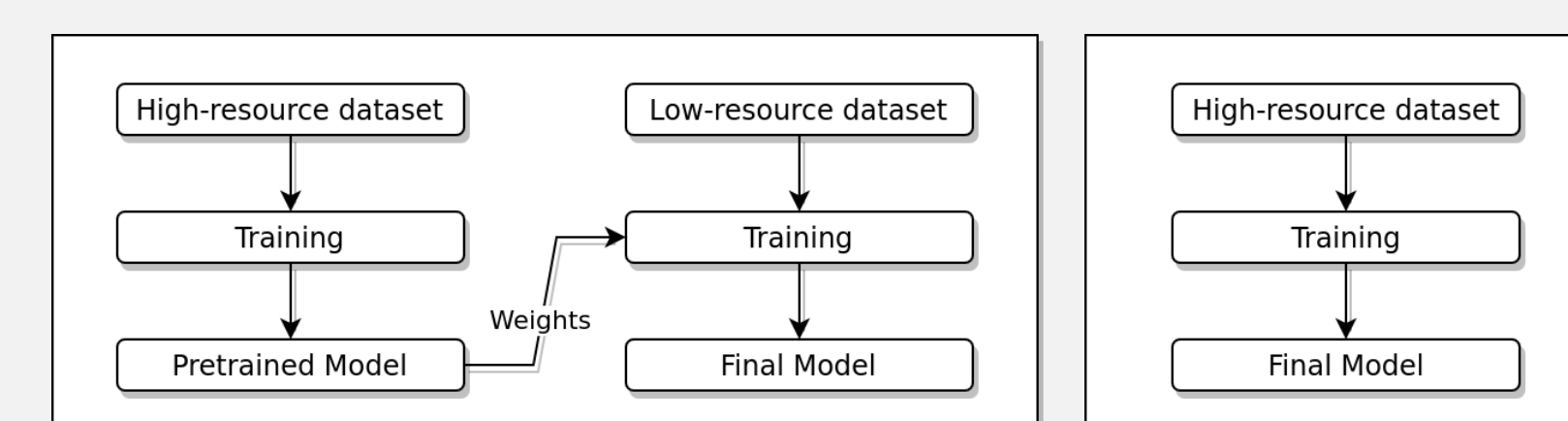

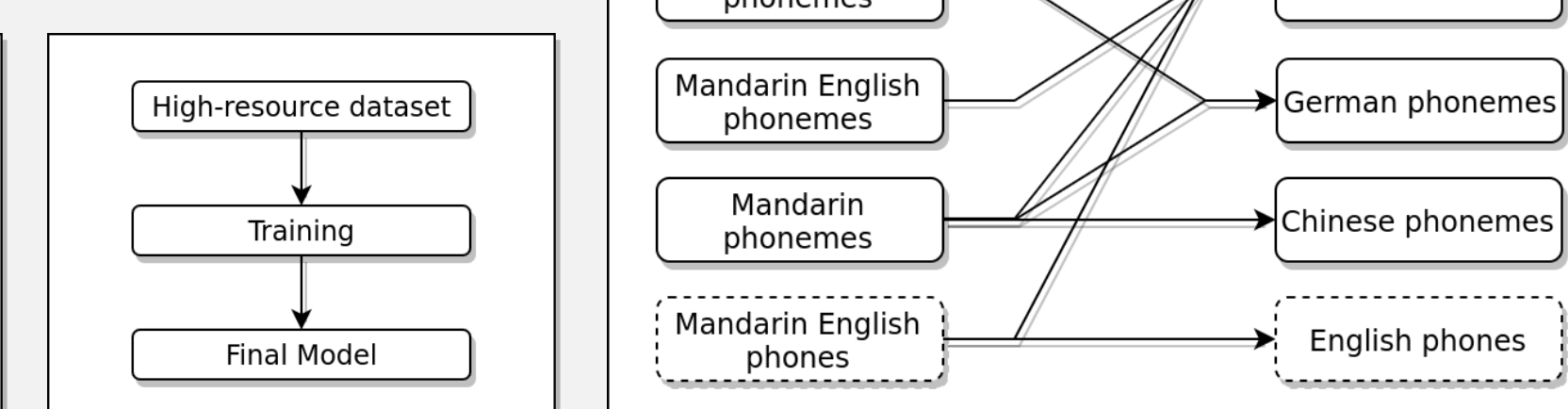Figure 3: Both models

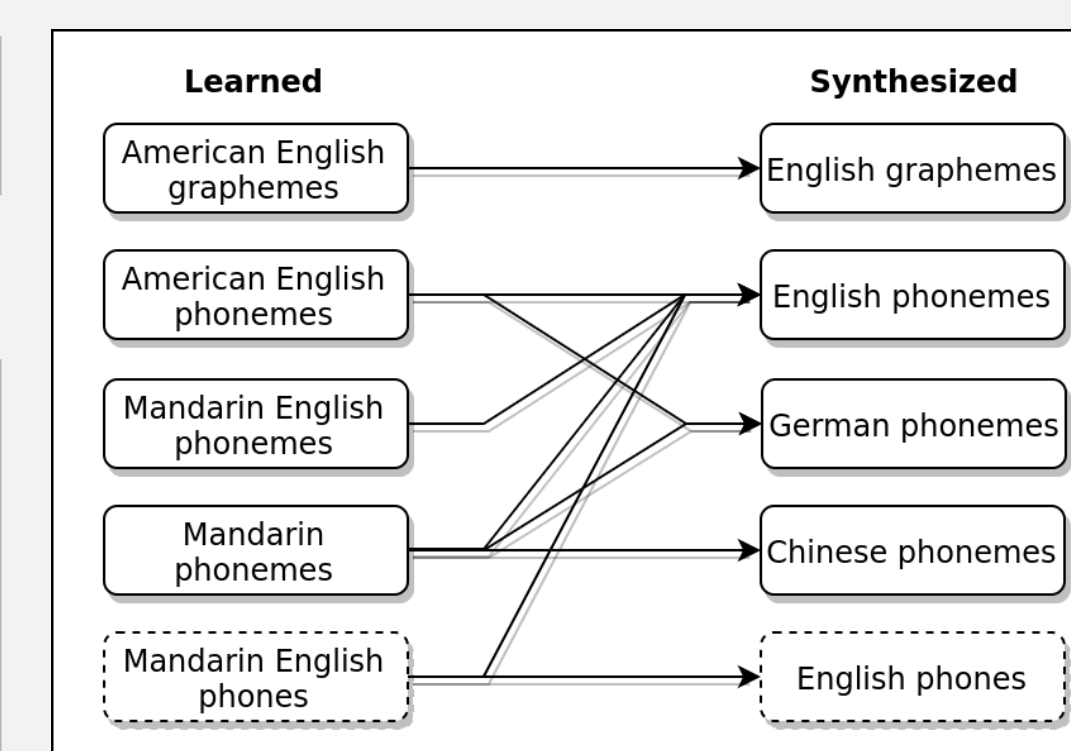Figure 4: Training first model

Figure 5: Training second model

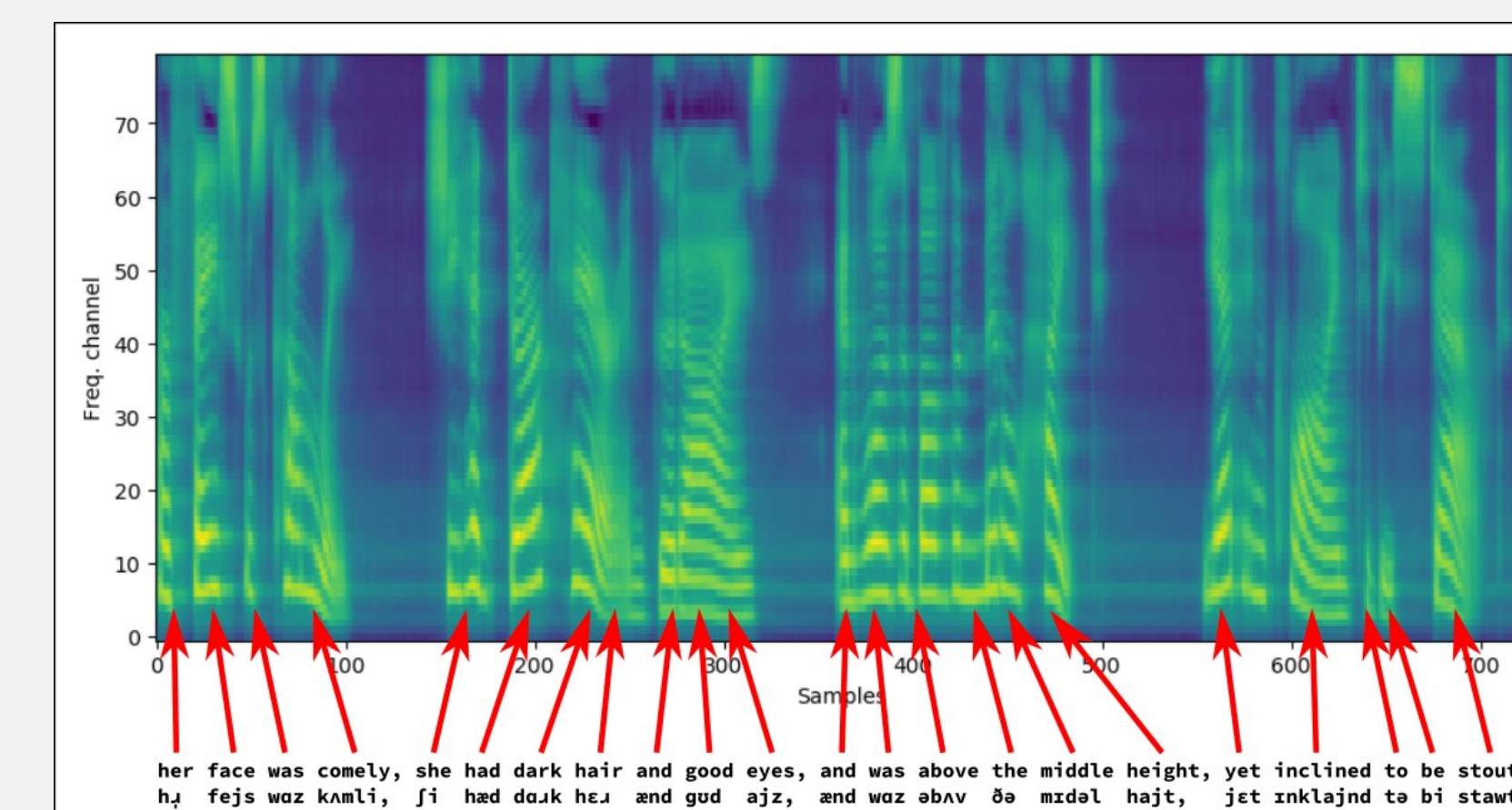Figure 6: Tried symbol representations until now

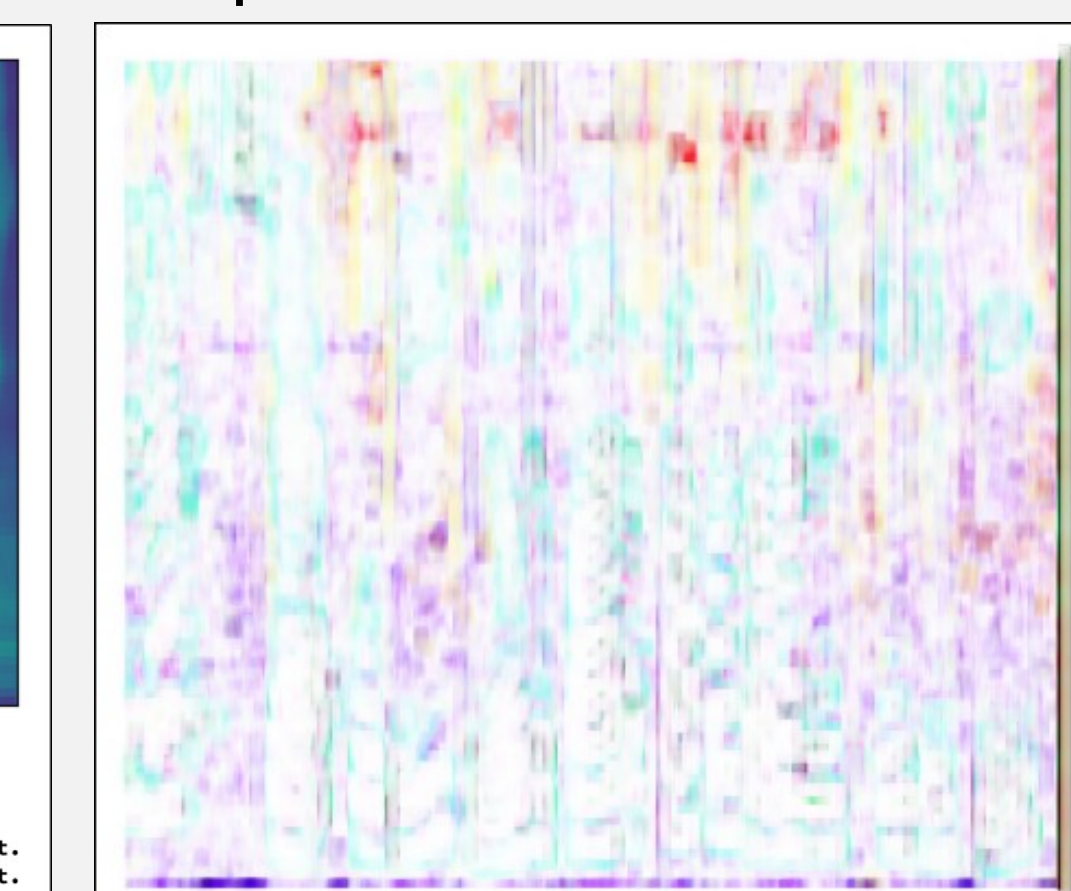Figure 7: Example output mel-spectrogram (first model)

Figure 8: Difference original and synthesized audio (second model)

CHINESE ENGLISH

- China has a very diverse linguistic landscape:
  · Mandarin/Cantonese
  · Minority languages
  · Regional dialects
- Focus on Yunnan and Guangxi
- Examples:
  · /ʌ/ pronounced as /ɑ/ (Ao & Low, 2012, p. 31)
  · /θ/ pronounced as /s/ (Ao & Low, 2012, p. 31; Deterding, 2016, p. 6)
  · Overlapping vowel length FLEECE & KIT (Deterding, 2016, p. 14)
  · Absence of reduced vowels in monosyllabic function words (Ao & Low, 2012, pp. 31–32; Deterding, 2016, p. 15)
  · Deviant word stress /kənˈsəːn/ vs. /ˈkənsəːn/ (Deterding, 2016, p. 16)
  · Subject pronoun copying 'That play, it was terrible.' (Xu, 2008, p. 7)

Figure 9: Map of China with "our" provinces highlighted

NEXT STEPS

- Baseline psychological experiments
- Running TTS system with very limited input data
- Developing a linguistic TTS quality metric
- Data collection trips (post Corona)
- Testing the significance of intercultural factors influencing credibility and learning
- Exploring methods of measuring credibility and learning performance

REFERENCES

- Tagliamonte, S. A. (2006). Analysing sociolinguistic variation. Cambridge University Press.
- Ao, R., & Low, E. L. (2012). Exploring pronunciation features of Yunnan English: The pronunciation of Yunnan speakers of English provides an interesting context for investigating Chinese English. English Today, 28(3), 27–33.
- Deterding, D. (2016). The Pronunciation of English by Students in Guangxi, South China. REFLections, 22, 1–20.
- Xu, Z. (2008). Analysis of syntactic features of Chinese English. Asian Englishes, 11(2), 4–31.
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R., Saurous, R. A., Agiomyrgiannakis, Y., & Wu, Y. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. https://arxiv.org/abs/1712.05884