Production and Perception of non-native English in China

Focus on Sociophonetic Variation by Humans and Artificial Agents

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

PhD Colloquium

12.12.2023

HYBRID SOCIETIES

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Overview

1. Introduction
   - Objectives
   - Research Questions
2. Methods
   - Data Collection
   - Analysis Methods
3. Results
   - Features of Chinese English
   - The Chinese English Vowel Space
   - Durational Variation
   - Learning with a non-native Agent
4. Conclusion
   - Summary of Results
   - Limitations
   - Outlook

Objectives

Mission

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

Project D03: *Credibility Through Non-native Language Varieties in Conversational Pedagogical Agents*

- 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?

Data Collection

Challenge

Sociolinguistic literature (e.g. Love et al., 2017) and computer science literature (e.g. Bozkurt et al., 2003) propose different corpus creation criteria.

Sociolinguistic methodology has to be adapted to TTS application:

- previous TTS systems trained on reading passages
- phone and diphone coverage in reading passages using greedy selection (Taubert et al., 2022)
- interview questions adapted from Tagliamonte (2006)

Analysis Methods

quantification workflow

- forced alignment using the Montreal Forced Aligner (McAuliffe et al., 2017) with manual corrections
- automated vowel formant measurements in Praat (see Albrecht, 2023a)
- data filtering (using a Hampel filter, Hampel, 1974) and transformation (to the psychoacoustic Bark scale, Traunmüller, 1990)
- further analysis and plotting in R (see Albrecht, 2023b, 2023c)

statistical hypotheis testing

- assumptions of statistical tests (e.g. ANOVA) have to be considered and tested
  - normality using the Shapiro-Wilk test (Shapiro & Wilk, 1965)
  - homoscedasticity using Levene's test (Levene, 1960)
- Kruskal-Wallis test as a non-parametric alternative to ANOVA (Kruskal & Wallis, 1952)
- reporting effect sizes to make tests meaningful, e.g. Kruskal-Wallis test effect size η^2 via rstatix package (Kassambara, 2023)

Features of Chinese English

Definition

Chinese English is a developing variety of English, which is subject to ongoing codification and normalization processes. It is based largely on the two major varieties of English, namely British and American English. It is characterized by the transfer of Chinese linguistic and cultural norms in discourse, syntax, pragmatics, lexis, and phonology. (Albrecht, 2021)



Figure 1: 'China linguistic map' U.S. Central Intelligence Agency, marked as public domain, Wikimedia

Features of Chinese English

- discourse: e.g. “ancestral hometown discourse” or discourse of “face” (Xu, 2020)
- syntax: three types of variation (Xu, 2008)
  - preference: use of imperatives
  - innovation: unmarked usage of nominalization
  - transference: null-subject/object utterances
- pragmatics: general-particular pattern (*theme* and *rheme*, see Halliday & Matthiessen, 2013) vs. problem-solution pattern (M. Wang & Li, 1993)
- lexis
  - Chinese loanwords in English: *dimsum, guanxi* (Xu, 2020)
  - words nativized into Chinese English: *propaganda, cadre, comrade, individualism* (Chang, 2001; Xu, 2020)
  - common English words used differently: *all-around* (Liang & Li, 2017)
- phonology: vowels and consonants + suprasegmentals (“rhythm”)

The Chinese English Vowel Space Across Three Provinces

RQ1 What are the vowel space features of Chinese English from Guangdong, Guangxi, and Yunnan?

RQ2 What regional or gender differences are there in the Chinese English vowel spaces under investigation in this study?

RQ3 Does the method of data elicitation (reading scripts vs. interviews) impact the vowel spaces?

Vowel Spaces: Results

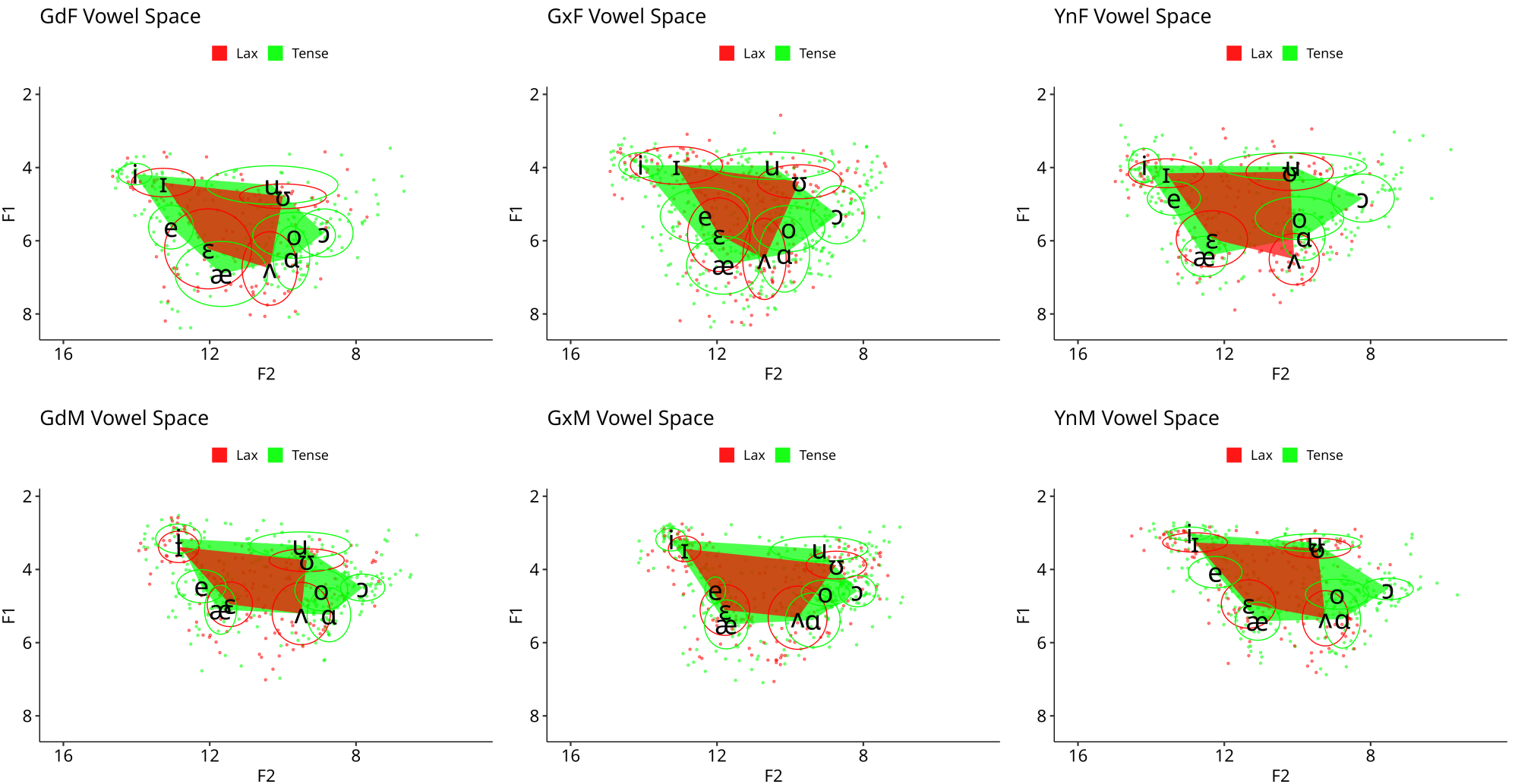


Figure 2: Vowel spaces of Chinese English subnational varieties

Merger Analysis

- vowel area overlap (based on Albrecht et al., 2022)
  - calculate vowel space area (i.e. convex polygon hull), calculate overlap percentage of areas, set threshold for potential mergers, e.g. 50% overlap
  - GOOSE-FOOT merges for almost all groups, FLEECE-KIT for some groups and DRESS-TRAP especially in Guangdong data
- support vector machine (SVM) binary classification task
  - follows Chan and Fuchs (fc.), split data in training set and validation set, run 50 times, average κ values denoting model fit and accuracy
  - FLEECE-KIT and GOOSE-FOOT merge for almost all groups
- statistical testing of variables
  - Kruskal-Wallis test (Kruskal & Wallis, 1952) as a non-parametric alternative to ANOVA
  - province only partially influences mergers
  - gender influences mergers both in the **combined data set** and in individual province data sets
  - data elicitation method partially significant in the combined data set, mostly not significant in other data sets

Vowel Spaces: Conclusion

- Chinese English vowel space has a consistent, unique, non-native shape and the tense-lax distinction is not very pronounced, with likely mergers of FLEECE-KIT and GOOSE-FOOT (RQ1)
- small differences in the vowel spaces by province and larger differences in terms of gender (RQ2) but size and shape very similar
- most likely data elicitation method does not impact the Chinese English vowel space (RQ3)

Durational Variation across inner/outer/expanding Circle Varieties of English

RQ1 How is the language background of English speakers from three variety circles reflected in their production of rhythm as expressed through durational variability?

RQ2 How does regional durational variation within a non-native variety compare to duration variation across countries?

RQ3 How does the data elicitation method (reading vs. interview) influence durational variability?

Durational Variation: Data and Metrics

- data of English varieties from all three circles (cf. Kachru, 1985)
  - inner circle: American, (Northern) British, Scottish, Irish English
  - outer circle: Nigerian English (Yoruba and Igbo)
  - expanding circle: Czech, Italian, Chinese (Guangdong, Guangxi, Yunnan) English
- two types of data:
  - reading data: linguistic reading passages (see Deterding, 2006, for a discussion of different reading passages) and TTS training data script (Taubert et al., 2022)
  - interview data: questions based on the list provided by Tagliamonte (2006)
- interval measures
  - ΔV, ΔC, %V (Ramus et al., 1999)
  - VarcoV, VarcoC (Dellwo, 2006)
- pairwise variability indices
  - nPVI-V (Low et al., 2000)
    $nPVI = 100 \times \left(\sum_{k=1}^{m-1} \left|(d_k - d_{k+1}) \,/\, ((d_k + d_{k+1}) \,/2)\right|\right) / (m-1)$
  - rPVI-C (Grabe & Low, 2002) $rPVI = \left(\sum_{k=1}^{m-1} \left|d_k - d_{k+1}\right|\right) / (m-1)$

Durational Variation: Results

Variety	nPVI-V	rPVI-C	VarcoV	VarcoC	ΔC	ΔV	%V
BrE	59.77	66.52	53.38	64.52	59.28	46.31	38.31
ScE	56.88	68.56	52.61	59.11	60.82	42.53	37.03
IrE	57.29	62.25	52.14	60.80	55.88	42.82	39.26
AmE	56.16	107.91	75.72	57.74	115.91	54.18	37.46
NiE-Y	59.04	90.69	60.54	86.98	77.22	53.34	38.56
NiE-I	58.80	84.83	59.21	66.34	72.28	52.01	40.41
CzE	63.66	112.46	71.56	76.88	93.59	75.81	41.65
ItE	60.04	128.84	75.20	76.42	113.90	64.61	36.82
ChE-Guangdong	76.25	92.45	91.34	127.05	81.05	100.14	41.49
ChE-Guangxi	80.72	89.29	100.89	133.74	79.03	117.32	48.62
ChE-Yunnan	84.26	99.46	108.42	140.52	87.61	131.44	51.64

Table 1: Results of all speech rhythm metrics for the reading data

Variety	nPVI-V	rPVI-C	VarcoV	VarcoC	ΔC	ΔV	%V
NiE-Y	64.4	83.19	77.94	107.12	69.17	55.85	24.96
CzE	68.65	104.95	91.77	109.04	90.52	75.27	25.18
ItE	73.11	96.36	82.78	109.75	82.2	87.33	29.16
ChE-Guangdong	69.76	114.29	121.84	94.14	94.66	87.09	30.07
ChE-Guangxi	68.55	136.28	151.90	98.78	113.92	87.51	24.86
ChE-Yunnan	71.59	149.54	255.96	101.29	129.24	105.93	28.36

Table 2: Results of all speech rhythm metrics for the interview data

Durational Variation: Conclusion

- increased language experience, operationalized by varieties belonging to different circles, leads to durational variability more closely approximating the variability of inner circle varieties (RQ1)
- while the variation within the Chinese varieties is relatively large, the variation across countries seems to be larger (RQ2)
- plots and tables suggest the influence of data elicitation methods, not confirmed by statistical testing (RQ3)
- no correlation between metric values and articulation rates

Variety	syllables/second	phones/second
AmE	4.28 (0.51)	10.90 (1.19)
BrE	4.38 (0.64)	8.89 (2.01)
ScE	4.45 (0.60)	9.07 (1.92)
IrE	4.64 (0.62)	9.53 (1.99)
NigE-Y	4.10 (0.28)	9.39 (1.96)
NigE-I	4.19 (0.39)	9.02 (1.36)
CzE	3.85 (0.46)	7.37 (1.26)
ItE	4.36 (0.35)	8.10 (1.03)
ChE-GD	3.75 (0.30)	5.11 (1.01)
ChE-GX	3.42 (0.38)	4.62 (1.00)
ChE-YN	3.44 (0.37)	4.47 (0.98)

Table 3: Articulation rate and phonation rate of read data

Variety	syllables/second	phones/second
NigE-Y	3.97 (0.15)	4.10 (2.11)
CzE	3.75 (0.43)	5.76 (0.77)
ItE	3.85 (0.57)	5.83 (0.83)
ChE-GD	3.32 (0.60)	4.94 (0.98)
ChE-GX	3.09 (0.39)	4.04 (0.59)
ChE-YN	3.26 (0.52)	4.11 (0.91)

Table 4: Articulation rate and phonation rate of interview data

Learning with a non-native Agent

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 (cf. Rey & Steib, 2013)

studies measure:

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

Note

The stimuli are based on factual knowledge about a subject, not language learning.

Measuring Linguistic Credibility: Methods

- participants: 74 Italians and 15 Czechs
- survey administered online via LimeSurvey
  - stimulus based on linguistic paper “Lemons and Watermelons: Visual Advertising and Conceptual Blending” by Delibegović Džanić and Berberović (2021)
  - participants were randomly assigned local variety or American English control condition
  - participants rated voice (credibility, warmth, competence, ease of understanding, likableness)
  - test component
    - general vocabulary (multiple choice)
    - academic vocabulary (multiple choice)
    - transfer learning questions (one short, one long open question, one multiple-choice question)
  - demographic questions
  - questions on media use and language exposure
- analysis in R, using chi-square tests (multiple choice) and non-parametric Wilcoxon signed-rank test (open question grades)

Measuring Linguistic Credibility: Conclusion

- no significant improvement in learning outcomes due to an agent speaking a non-native variety
- the Italian participants performed better than the Czech in the multiple-choice questions, but the groups scored similarly in the open-choice questions
- there might be a hidden English level difference despite similar demographics
- no influence of virtual assistant usage
- the COMPETENCE characteristics (*credibility*, *competence*, *ease of understanding*) were rated higher than the WARMTH characteristics (*likableness*, *warmth*)

Summary of Results

- FLEECE-KIT and GOOSE-FOOT mergers
- tense and lax vowel areas rather similar in size
- CLOTH vowel relatively far back (as in Q. Wang, 2023)
- ChE exhibits different durational variation than native varieties
- CzE and ItE differ in durational variation, suggesting L1 influence
- no correlation between DVMs and speech rate found
- automatic vowel formant analysis is a valid and reliable method
- non-native agent rated as warm, but did not improve learning

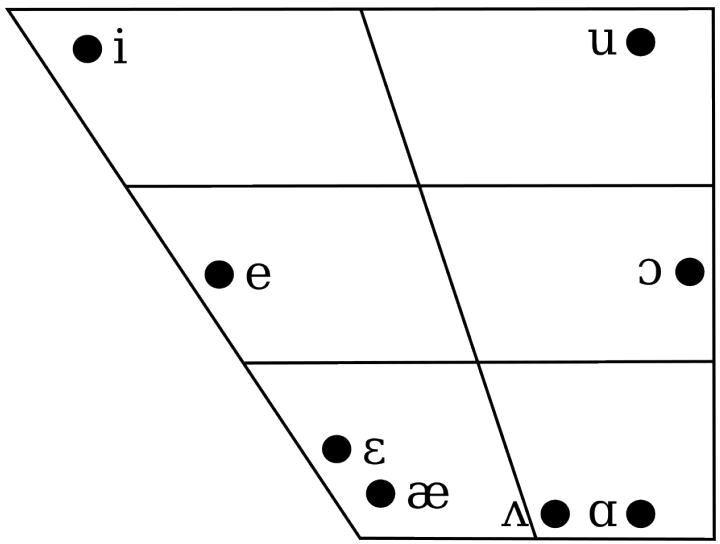


Figure 3: Proposed, tentative Chinese English vowel inventory

Limitations

- data-related limitations
  - Covid-19 impact on data collection
  - participants mainly university students
- theoretical and methodological limitations
  - investigating L1 influence of L2 speech requires a SLA theory that permits that
  - measuring a surface phenomenon allows conclusions of underlying language features and cross-linguistic influence
  - isolating a variable and its effect in an experiment is challenging

Summary

Limitations imposed by external factors, such as a global pandemic, have been mitigated successfully. The methodological limitations have to be kept in mind when interpreting results.

Outlook

- further linguistic research
  - add data from other parts of China, esp. from Northern and Western China
  - investigate the North-South divide in Chinese varieties in Chinese English
  - geotagged data for language maps (cf. Siebenhaar & Hahn, 2019)
  - include written data (from social media) to investigate lexis and grammar
  - grapheme-to-phoneme mapping
- further research on credibility and learning
  - using a validated scale to improve the comparability of results (currently developed in D03, Tamboli, in prep.)
  - different format of stimuli (e.g. podcast)
  - multimodal stimuli
  - giving the agent an appearance (cf. anthropomorphism)
  - acceptance of synthesized non-native English outside educational setting

References I

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh University Press.
Albrecht, S. (2021).Current research on the linguistic features of Chinese English. *World Englishes*, weng.12572. https://doi.org/10.1111/weng.12572
Albrecht, S. (2023a). Praat scripts for automated vowel formant measurements. https://doi.org/10.5281/ZENODO.8384068
Albrecht, S. (2023b). Analysis Scripts for "Durational Variation across inner/outer/expanding Circle Varieties of English". https://doi.org/10.5281/ZENODO.10041445
Albrecht, S. (2023c). Analysis Scripts for "The Chinese English Vowel Space Across Three Provinces". https://doi.org/10.5281/ZENODO.8409882
Albrecht, S., Tamboli, R., Taubert, S., Eibl, M., Rey, G. D., & Schmied, J. (2022).Towards a Vowel Formant Based Quality Metric for Text-to-Speech Systems: Measuring Monophthong Naturalness. *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 1–6. https://doi.org/10.1109/CIVEMSA53371.2022.9853712
Arvaniti, A. (2009).Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*(1-2), 46–63. https://doi.org/10.1159/000208930
Bertinetto, P. M. (1989).Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue De Phonétique Appliquée*, *91*(93), 99–130.
Bozkurt, B., Ozturk, O., & Dutoit, T. (2003).Text design for TTS speech corpus building using a modified greedy selection. *Eighth European Conference on Speech Communication and Technology*.
Chan, H. L., & Fuchs, R. (fc.).Revisiting the Vowels of Hong Kong English - The post-handover Generation. *ICPhS*.
Chang, J. (2001). Chinese Speakers. In M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 310–324). Cambridge University Press.
Cummins, F. (2002).Speech rhythm and rhythmic taxonomy. *Proceedings of Speech Prosody 2002*, 121–126.
Dauer, R. M. (1983).Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4
Delibegović Džanić, N., & Berberović, S. (2021). Lemons and Watermelons: Visual Advertising and Conceptual Blending. In L. D'Angelo, A. Mauranen, & S. Maci (Eds.), *Metadiscourse in Digital Communication* (pp. 115–132). Springer International Publishing. https://doi.org/10.1007/978-3-030-85814-8_6
Dellwo, V. (2006).Rhythm and Speech Rate: A Variation Coefficient for deltaC. https://doi.org/10.5167/UZH-111789
Deterding, D. (2006).The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, *36*(2), 187–196.
Fuchs, R. (2023). Analysing the speech rhythm of New Englishes: A guide to researchers and a case study in Pakistani, Philippine, Nigerian and British English. In G. Wilson & M. Westphal (Eds.), *New Englishes, New Methods*. Benjamins.
Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). De Gruyter Mouton. https://doi.org/10.1515/9783110197105.2.515
Halliday, M. A. K., & Matthiessen, C. M. (2013). *Halliday's introduction to functional grammar*. Routledge.
Hampel, F. R. (1974).The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*(346), 383–393. https://doi.org/10.1080/01621459.1974.10482962

References II

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: English language in the outer circle. In R. Quirk Randolph & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–36). Cambridge University Press.

Kassambara, A. (2023). *Rstatix: Pipe-friendly framework for basic statistical tests*. Manual.

Kruskal, W. H., & Wallis, W. A. (1952).Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics* (7. ed). Cengage Learning.

Lehiste, I. (1977).Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–264.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin & H. Hotelling (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Liang, J., & Li, D. C. (2017). Researching Collocational Features: Towards China English as a Distinctive New Variety. In Z. Xu, D. He, & D. Deterding (Eds.), *Researching Chinese English: The State of the Art* (pp. 61–75). Springer.

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.

Low, E. L., Grabe, E., & Nolan, F. (2000).Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and Speech*, *43*(4), 377–401. https://doi.org/10.1177/00238309000430040301

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

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116).

O'Dell, M. L., & Nieminen, T. (1999).Coupled oscillator model of speech rhythm. *Proceedings of the XIVth International Congress of Phonetic Sciences*, *2*, 1075–1078.

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press.

Ramus, F., Nespor, M., & Mehler, J. (1999).Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

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

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). Edward Arnold.

Shapiro, S. S., & Wilk, M. B. (1965).An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Siebenhaar, B., & Hahn, M. (2019).Vowel space, speech rate and language space. *Proceedings of the 19th International Congress of Phonetic Sciences. Melbourne, Australia*.

References III

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.
Tamboli, R. (in prep.).Establishing Conversational Pedagogical Agents as Credible Knowledge Providers: The Case of Synthsized Indian English.
Taubert, S., Sternkopf, J., Kahl, S., & Eibl, M. (2022).A Comparison of Text Selection Algorithms for Sequence-to-Sequence Neural TTS. *2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 1–6. https://doi.org/10.1109/ICSPCC55723.2022.9984283
Traunmüller, H. (1990).Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, *88*(1), 97–100. https://doi.org/10.1121/1.399849
Wang, M., & Li, J. (1993).An investigation on English discourse patterns used by Chinese students. *Waiyu Jiaoxue Yu Yanjiu [Foreign Language and Teaching Research]*, 59–64.
Wang, Q. (2023).An instrumental investigation of the vowel inventory of China English. *Asian Englishes*, *25*(1), 111–132. https://doi.org/10.1080/13488678.2022.2158265
White, L., & Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven & A. Chen (Eds.), *Oxford handbook of language prosody* (pp. 167–182). Oxford University Press.
Xu, Z. (2008).Analysis of Syntactic Features of Chinese English. *Asian Englishes*, *11*(2), 4–31.
Xu, Z. (2020). Chinese English. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes* (pp. 265–297). Routledge.

Production and Perception of non-native English in China

Focus on Sociophonetic Variation by Humans and Artificial Agents

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

PhD Colloquium

12.12.2023

HYBRID SOCIETIES

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Data Collection and Covid-19

Challenge

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

Alternative forms of data collection, broadening our approach:

- added Nigerian English, leveraging social network of visiting scholars from Nigeria (in-group interviews)
- added Italian English, leveraging personal connections
- added Czech English, via Marina's project
- recorded Chinese students in Chemnitz
- Chinese colleagues provided all 30 planned recordings

Merger Analysis

Group	Gender	TRAP-DRESS	FLEECE-KIT	GOOSE-FOOT
ChE combined		54.65%	78.28%	87.26%
ChE	female	50.39%	27.24%	79.01%
	male	52.93%	78.69%	71.18%
Guangdong	female	27.61%	32.85%	95.21%
	male	40.54%	80.15%	62.83%
Guangxi	female	30.34%	27.68%	66.14%
	male	85.04%	64.31%	45.29%
Yunnan	female	12.27%	34.79%	68.23%
	male	11.34%	39.58%	86.57%

Table 5: Results of the vowel area overlap analysis

Merger Analysis

Group	Gender	TRAP-DRESS (κ / accuracy)	FLEECE-KIT (κ / accuracy)	GOOSE-FOOT (κ / accuracy)
ChE combined		0.67 / 0.84	0.26 / 0.63	0.37 / 0.67
ChE	female	0.65 / 0.83	0.35 / 0.67	0.36 / 0.68
	male	0.70 / 0.85	0.28 / 0.64	0.43 / 0.71
Guangdong	female	0.82 / 0.92	0.35 / 0.67	0.29 / 0.64
	male	0.50 / 0.75	0.16 / 0.57	0.54 / 0.77
Guangxi	female	0.57 / 0.79	0.42 / 0.71	0.49 / 0.75
	male	0.63 / 0.82	0.31 / 0.65	0.49 / 0.75
Yunnan	female	0.89 / 0.94	0.17 / 0.55	0.25 / 0.63
	male	0.84 / 0.93	0.24 / 0.61	0.14 / 0.54

Table 6: Results of the SVM classification analysis. Small kappa values indicate poor separability.

Merger Analysis

	TRAP-DRESS		FLEECE-KIT		GOOSE-FOOT	
	Female	Male	Female	Male	Female	Male
ChE combined	Y/N		Y/Y		Y/Y	
ChE	Y/N	Y/N	N/Y	Y/Y	Y/Y	Y/N
Guangdong	N/N	N/N	N/Y	Y/Y	Y/Y	Y/N
Guangxi	N/N	Y/N	N/N	Y/Y	Y/N	Y/N
Yunnan	N/N	N/N	N/Y	N/Y	Y/Y	Y/Y

Table 7: Combined indicators of mergers (overlap analysis / SVM analysis)

Statistical Testing of Variables

- province only partially influences mergers
  - GOOSE-FOOT F1 highly significant
  - FLEECE-KIT F2 highly significant
  - effect size small in both cases
- gender influences mergers both in the combined data set and in individual province data sets
- data elicitation method partially significant in the combined data set, mostly not significant in other data sets

Data Set	Variable	Formant	Merger	χ^2	p	Significance	η^2
ChE	elicitation	F1	TRAP-DRESS	13.54	<.001	***	0.01
			FLEECE-KIT	41.33	<.001	***	0.04
			GOOSE-FOOT	5.55	<.001	*	0.01
		F2	TRAP-DRESS	0.91	.4		0
			FLEECE-KIT	0.31	.64		0
			GOOSE-FOOT	46.99	<.001	***	0.06
	gender	F1	TRAP-DRESS	361.44	<.001	***	0.37
			FLEECE-KIT	395.77	<.001	***	0.43
			GOOSE-FOOT	296.47	<.001	***	0.4
		F2	TRAP-DRESS	63.02	<.001	***	0.06
			FLEECE-KIT	192.53	<.001	***	0.21
			GOOSE-FOOT	87.87	<.001	***	0.12
	province	F1	TRAP-DRESS	0.98	.65		0
			FLEECE-KIT	2.19	.4		0
			GOOSE-FOOT	16.64	<.001	***	0.02
		F2	TRAP-DRESS	0.04	.98		0
			FLEECE-KIT	24.41	<.001	***	0.02
			GOOSE-FOOT	5.53	.09		0.01
Guangdong	elicitation	F1	TRAP-DRESS	3.4	.09		0.01
			FLEECE-KIT	4.77	.04	*	0.01
			GOOSE-FOOT	0.12	.74		0
		F2	TRAP-DRESS	0.2	.68		0
			FLEECE-KIT	2.57	.14		0.01
			GOOSE-FOOT	24.55	<.001	***	0.11
	gender	F1	TRAP-DRESS	150.62	<.001	***	0.51
			FLEECE-KIT	121.1	<.001	***	0.46
			GOOSE-FOOT	126.84	<.001	***	0.6
		F2	TRAP-DRESS	25.13	<.001	***	0.08
			FLEECE-KIT	45.14	<.001	***	0.17
			GOOSE-FOOT	15.77	<.001	***	0.07
Guangxi	elicitation	F1	TRAP-DRESS	0.69	.47		0
			FLEECE-KIT	14.14	<.001	***	0.03
			GOOSE-FOOT	2.5	.14		0.01
		F2	TRAP-DRESS	2.62	.14		0
			FLEECE-KIT	0.28	.64		0
			GOOSE-FOOT	22.15	<.001	***	0.07
	gender	F1	TRAP-DRESS	100.18	<.001	***	0.27
			FLEECE-KIT	168.9	<.001	***	0.44
			GOOSE-FOOT	74.55	<.001	***	0.25
		F2	TRAP-DRESS	1.88	.21		0
			FLEECE-KIT	50.07	<.001	***	0.13
			GOOSE-FOOT	60.68	<.001	***	0.2
Yunnan	elicitation	F1	TRAP-DRESS	17.38	<.001	***	0.06
			FLEECE-KIT	26.11	<.001	***	0.09
			GOOSE-FOOT	7.64	.01	**	0.03
		F2	TRAP-DRESS	0.3	.64		0
			FLEECE-KIT	2.7	.14		0.01
			GOOSE-FOOT	4.92	.04	*	0.02
	gender	F1	TRAP-DRESS	112.7	<.001	***	0.38
			FLEECE-KIT	115.72	<.001	***	0.42
			GOOSE-FOOT	112.23	<.001	***	0.49
		F2	TRAP-DRESS	50.66	<.001	***	0.17
			FLEECE-KIT	92.15	<.001	***	0.34
			GOOSE-FOOT	20.17	<.001	***	0.34

Figure 4: Results for vowel formants from Kruskal-Wallis tests. (Note: elicitation and gender $df = 1$, province $df = 2$. Notation: * significant ($p < .05$), ** very significant ($p < .01$), *** highly significant ($p < .001$). Effect sizes: $0.01 \leq \eta^2 < 0.06$ (small effect), $0.06 \leq \eta^2 < 0.14$ (moderate effect) and $\eta^2 \geq 0.14$ (large effect))

Speech Rhythm

- conventionally, languages have been categorized as syllable or stress-timed (Abercrombie, 1967; Pike, 1945)
- the third category of mora timed languages (like Japanese, Ladefoged & Johnson, 2014) has been abandoned (Fuchs, 2023)
- speech rhythm could be
  - a surface phenomenon realized by some linguistic units or processes (Abercrombie, 1967)
  - a motor control phenomenon realized by coupled oscillators (O'Dell & Nieminen, 1999)
  - a perceptual phenomenon where an aperiodic signal is perceived as periodic (Lehiste, 1977)
  - a result of abstract structuring of points in time, manifested by sufficiently salient linguistic elements as grouping structure (Arvaniti, 2009)
  - a result of a systematic relationship between segmental identity and segmental and prosodic context, and surface duration patterns (Dauer, 1983; Low et al., 2000; Ramus et al., 1999)

Durational Variation

- categorization based on equal length of units (isochrony) could not be supported empirically (Bertinetto, 1989; Dauer, 1983; Roach, 1982)
- an investigation of surface timing features does not constitute an investigation of rhythm (Arvaniti, 2009; Cummins, 2002)
- evidence of cross-linguistic variation in the “durational marking of stress contrast” (White & Malisz, 2020, p. 171)

Durational Variation

Investigating durational variation can show cross-linguistic variation without imposing a definition of rhythm or a typology of languages.

Durational Variation Metrics

- languages differ strongly in the amount of vowel reduction and the complexity of consonant clusters (Dauer, 1983)
- vowel duration and intensity might be used by infants to infer the rhythm class of their native language (Mehler et al., 1996)
- syllabification rules differ by language and are contested even within languages (Low et al., 2000)
- interval measures
  - ΔV, ΔC, %V (Ramus et al., 1999)
  - VarcoV, VarcoC (Dellwo, 2006)
- pairwise variability indices
  - nPVI-V (Low et al., 2000)
    $nPVI = 100 \times \left(\sum_{k=1}^{m-1} \left|\left(d_k - d_{k+1}\right) / \left(\left(d_k + d_{k+1}\right)/2\right)\right|\right) / \left(m-1\right)$
  - rPVI-C (Grabe & Low, 2002) $rPVI = \left(\sum_{k=1}^{m-1} \left|d_k - d_{k+1}\right|\right) / \left(m-1\right)$

Durational Variation: Results

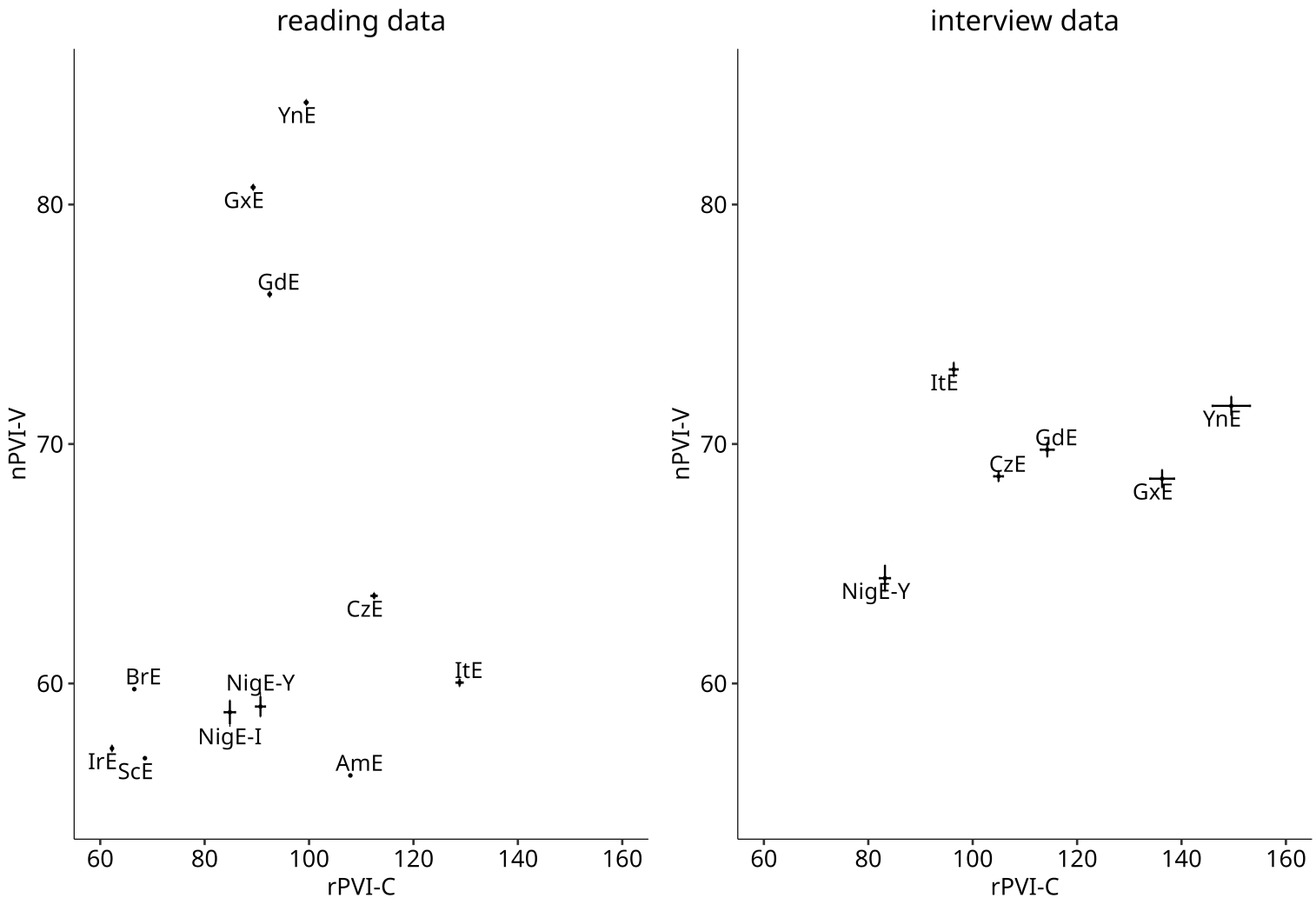


Figure 5: nPVI-V - rPVI-C Chart for reading and interview data

Durational Variation: Results

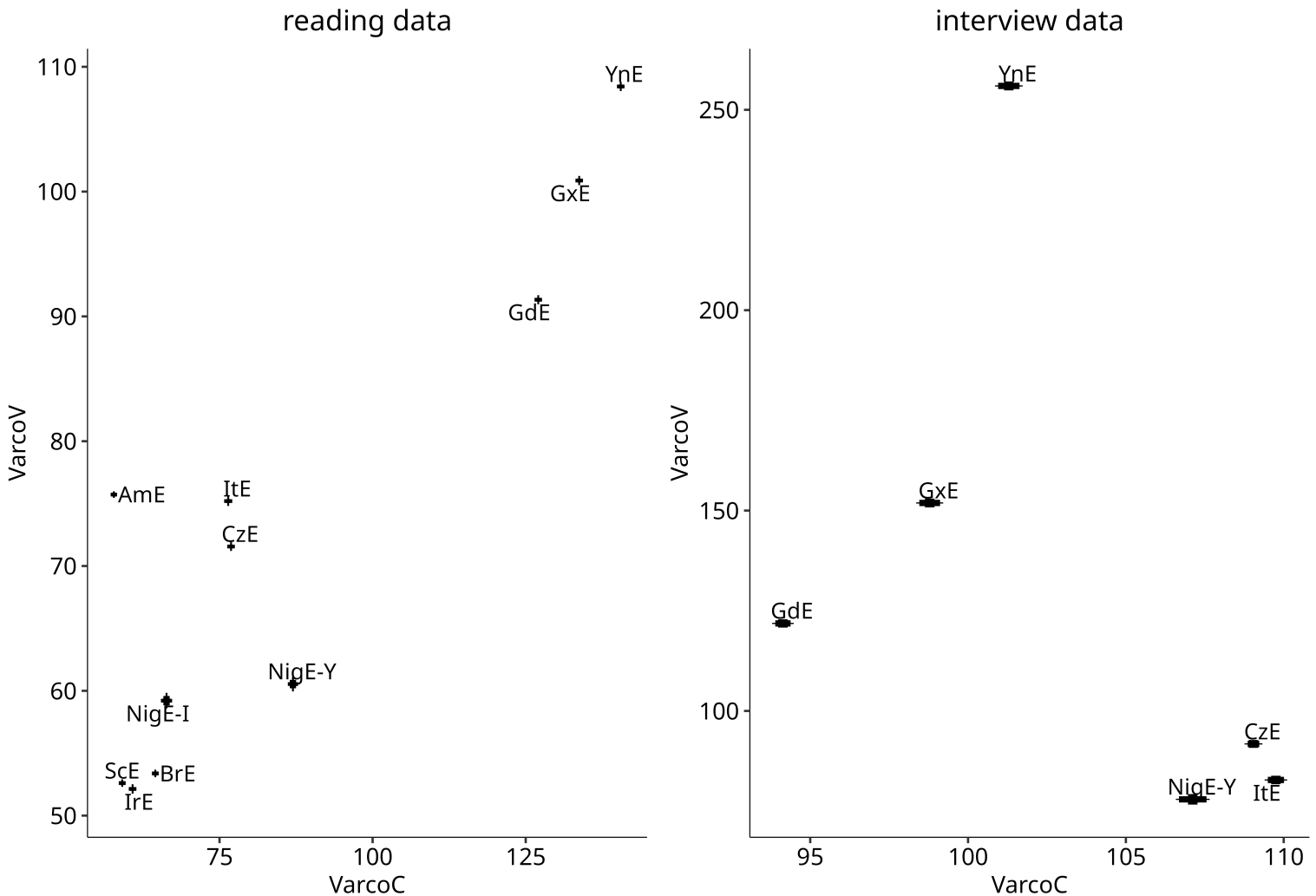


Figure 6: VarcoV - %V Chart for reading and interview data

Durational Variation: Results Data Elicitation Method

- test for normality with the Shapiro-Wilk test (Shapiro & Wilk, 1965), from the R rstatix package (Kassambara, 2023)
- test for homogeneity of variances, using Levene's test (Levene, 1960)
- based on the results of these tests, applied the Kruskal-Wallis test (Kruskal & Wallis, 1952) as a non-parametric alternative to ANOVA

DVM	x^2	p value	Significance	η^2
nPVI-V	2.273	0.132		0.085
rPVI-C	2.919	0.088		0.128
VarcoV	6.313	0.012	*	0.354
VarcoC	2.273	0.132		0.085
ΔV	1.707	0.191		0.047
ΔC	2.273	0.132		0.085
%V	11.000	<0.001	***	0.667

Table 8: Results for the influence of data elicitation method of DVMs from Kruskal-Wallis tests. (Note: elicitation method df = 1. Notation: * significant (p <.05), ** very significant (p <.01), *** highly significant (p <.001). Effect sizes: 0.01 ≤ η^2 <0.06 (small effect), 0.06 ≤ η^2 <0.14 (moderate effect) and η^2 ≥ 0.14 (large effect))

Measuring Linguistic Credibility: Results

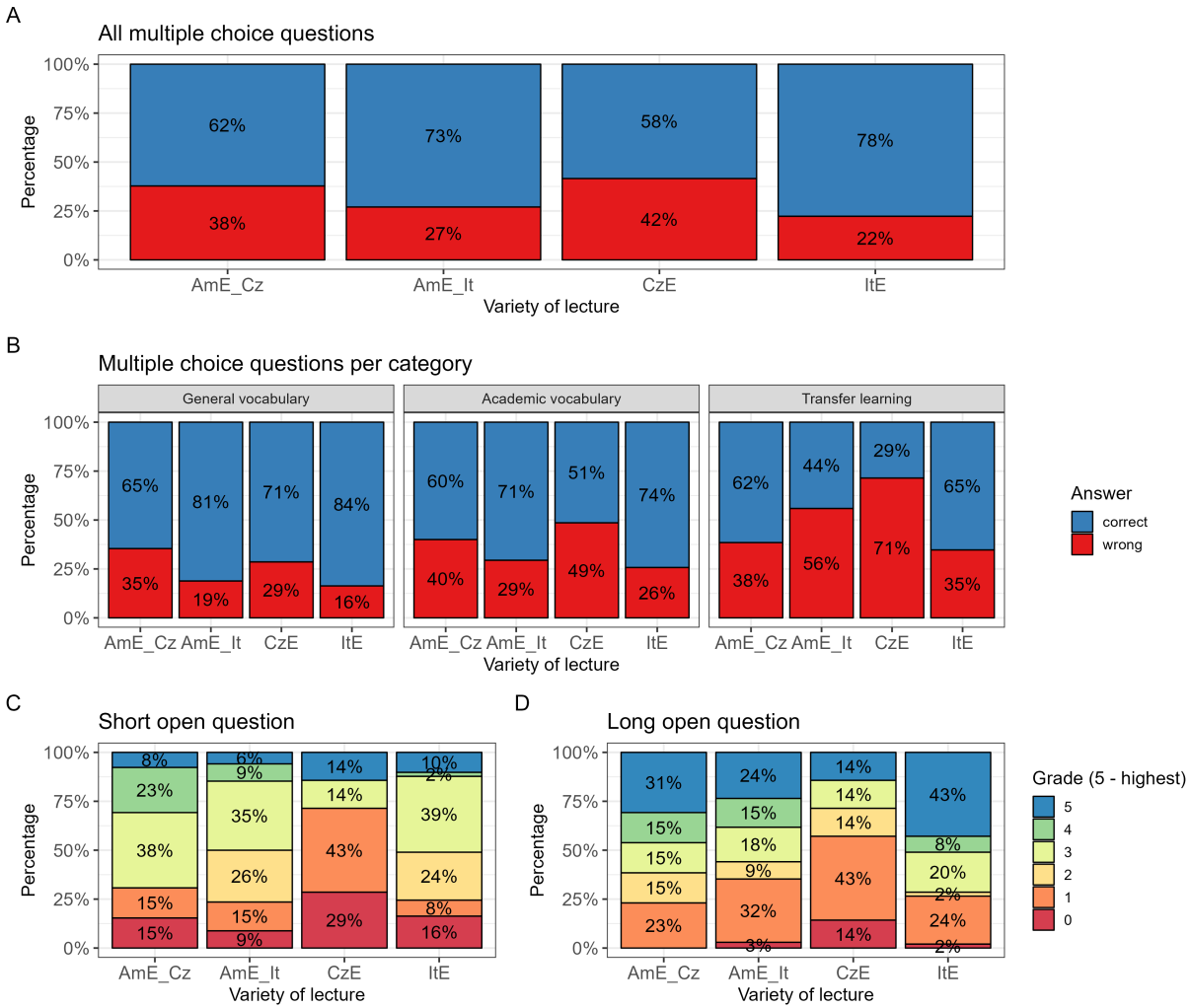


Figure 7: Overview of the responses to the multiple choice and open questions per lecture variety.

Measuring Linguistic Credibility: Results

- language groups differ in terms of correctness [$\chi^2(1, N = 979) = 18.8, p < .001$], Italians perform better
- significant difference between the impact of the different lecture varieties on the correct and wrong responses [$\chi^2(3, N = 979) = 20.8, p < .001$]
- when only the Italian data is tested, the number of correct and wrong responses does not differ based on the variety of the lecture [$\chi^2(1, N = 814) = 0.646, p = .422$]
- the groups differ in their answer correctness of the general vocabulary questions [$\chi^2(1, N = 445) = 10.7, p = .001$]
- no difference in open questions [$W = 473, p = .368$]
- no influence of virtual assistant use [$\chi^2(1, N = 979) = 0.165, p = .685$]

Measuring Linguistic Credibility: Results

Group 1	Group 2	N1	N2	W	p	p adj.	Sign.
competence	credibility	89	89	3894.5	0.843	1	ns
competence	ease of understanding	89	89	3657.5	0.364	1	ns
competence	likableness	89	89	5411	1.34×10^{-5}	8.04×10^{-5}	***
competence	warmth	89	89	5760.5	5.64×10^{-8}	5.08×10^{-7}	***
credibility	ease of understanding	89	89	3691.5	0.42	1	ns
credibility	likableness	89	89	5435.5	1×10^{-5}	7×10^{-5}	***
credibility	warmth	89	89	5780.5	4.21×10^{-8}	4.21×10^{-7}	***
ease of understanding	likableness	89	89	5422.5	1.37×10^{-5}	8.04×10^{-5}	***
ease of understanding	warmth	89	89	5709.5	1.77×10^{-7}	1.42×10^{-6}	***
likableness	warmth	89	89	4159.5	0.548	1	ns

Table 9: Results of Wilcoxon signed-rank tests comparing characteristic ratings.

Vowel Formant TTS Quality Metric

- most TTS systems have some form of quality metric built-in
- tacotron2: visual comparison of mel spectrograms

Idea

Can we use our vowel space measurement pipeline to establish a linguistically motivated TTS quality metric?

- two data sets:
  - original: recordings used to train TTS system
  - inferred: speech generated by the TTS system
- plot vowel space for both
- draw a convex polygon around means, central 75%, and all data points
- calculate the overlap of the polygons

Published at CIVEMSA 2022 (Albrecht et al., 2022) and received Best Paper Award 2nd Place

Vowel Formant TTS Quality Metric

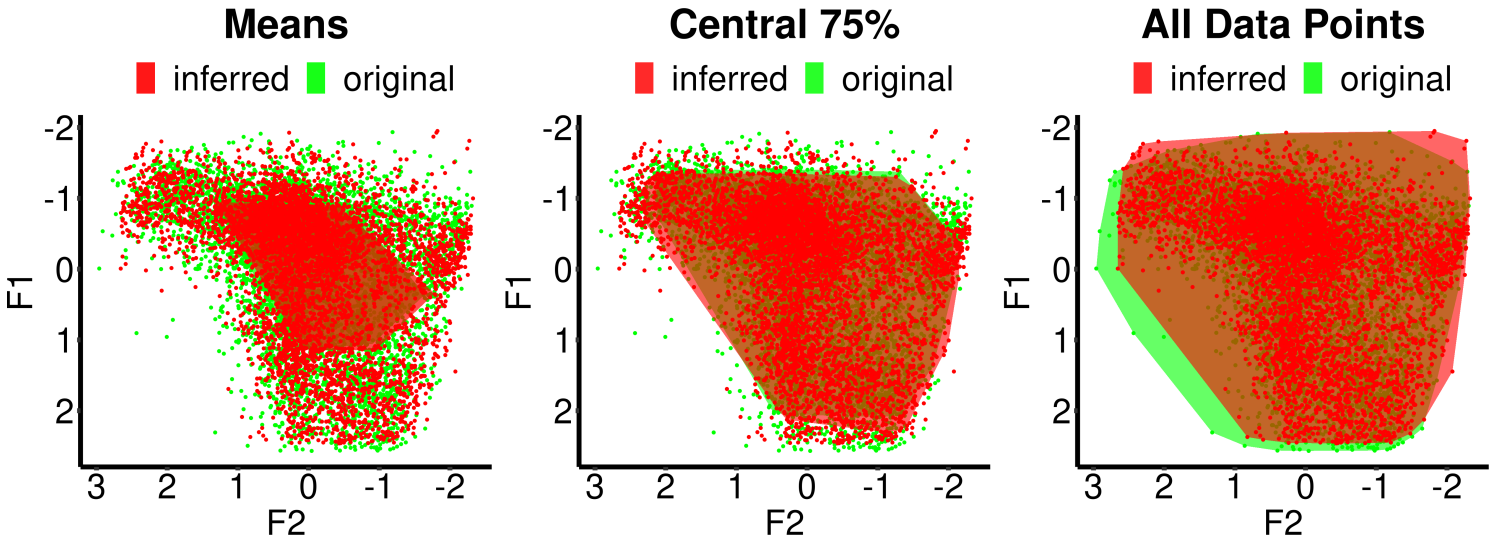


Figure 8: LJ Speech Vowel Space, plotted as F1 - F2 space (Lobanov Normalized, Hampel Filtered)

Dataset	Overlap
Phoneme Averages	93.20%
Central 75%	97.70%
All Data Points	91.50%

Table 10: Vowel Space Overlap

Vowel Formant TTS Quality Metric

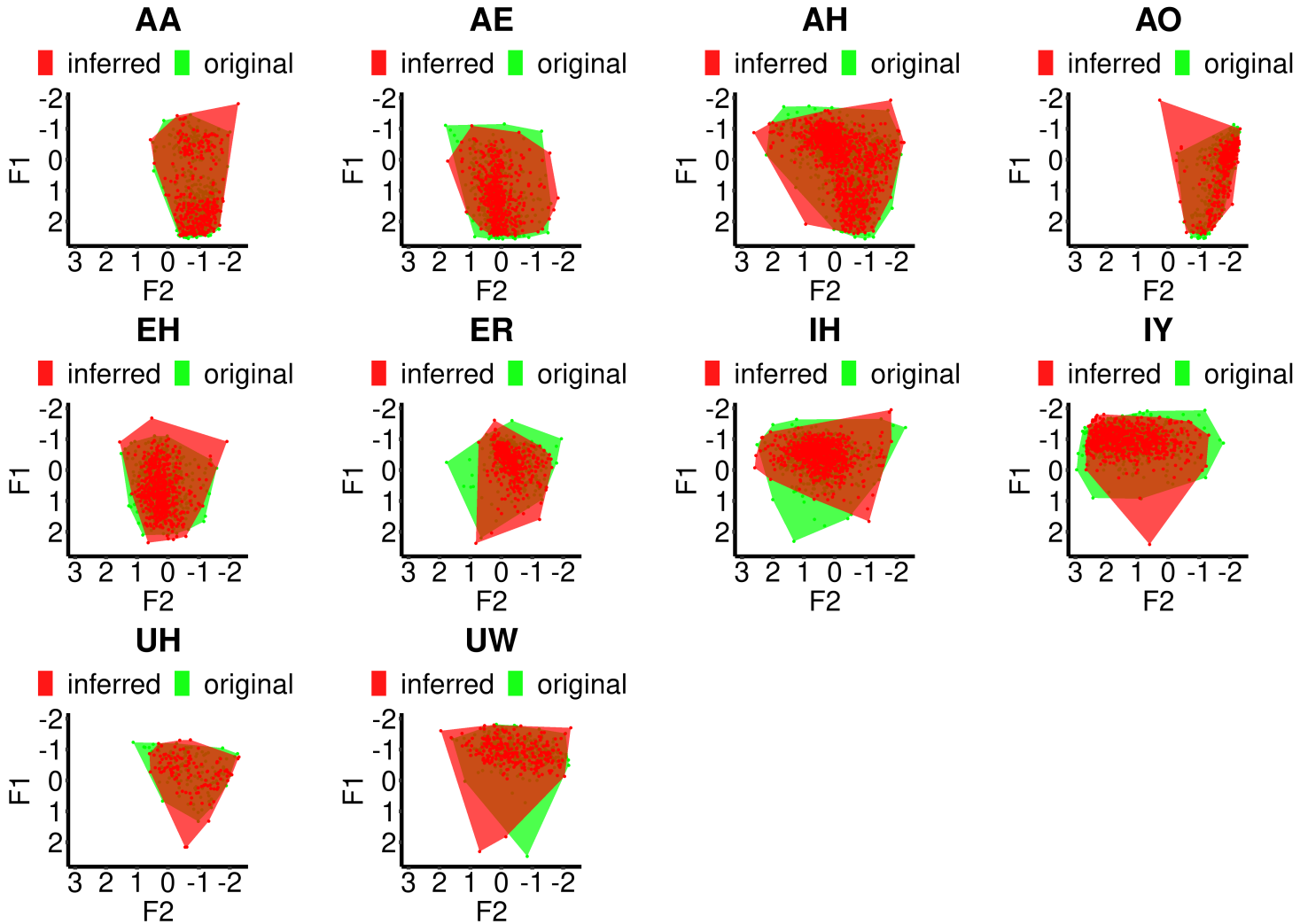


Figure 9: LJ Speech Phoneme Space All Data Points (Lobanov Normalized, Hampel Filtered)

Vowel Formant TTS Quality Metric

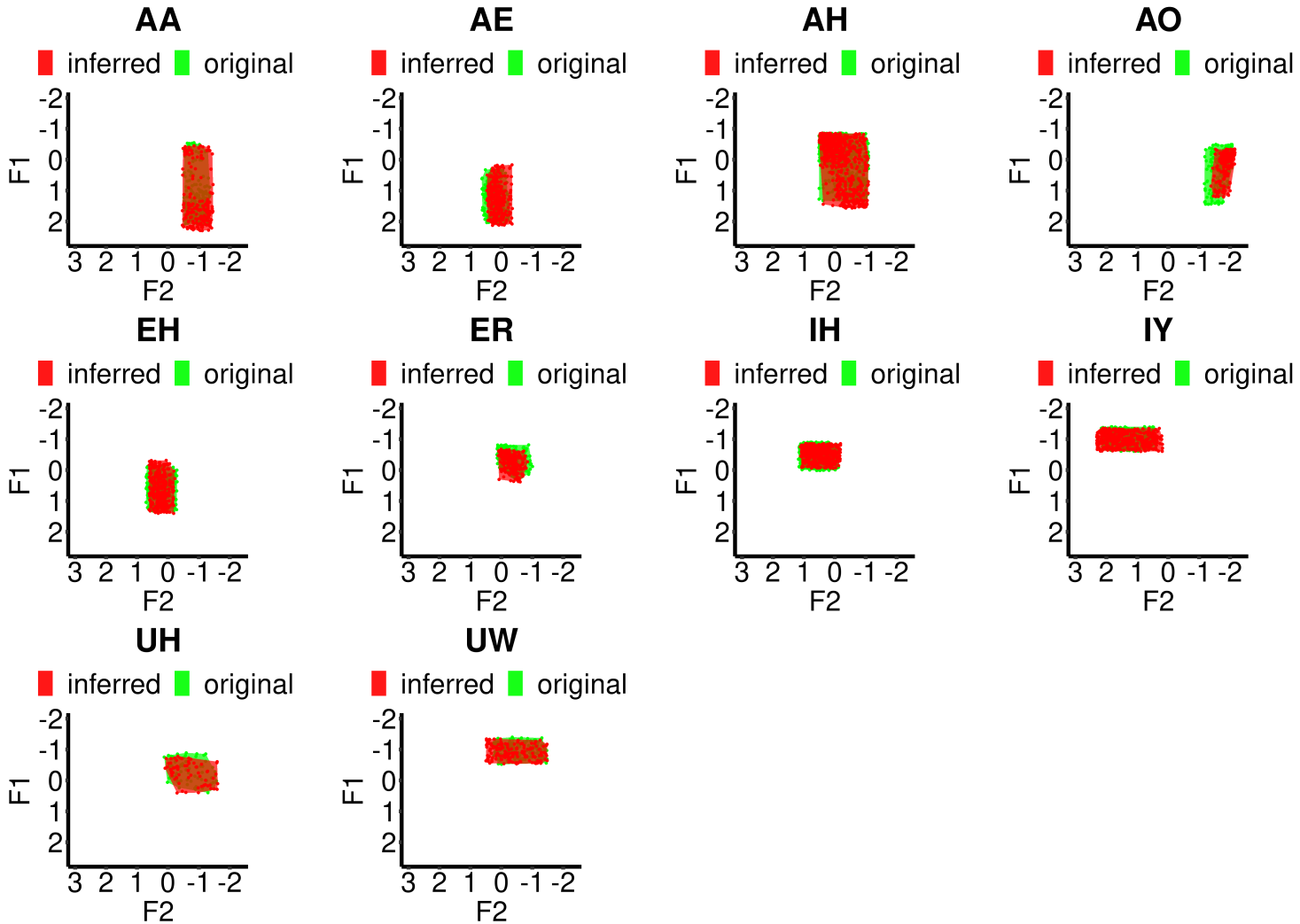


Figure 10: LJ Speech Phoneme Space Central 75% of All Data Points (Lobanov Normalized, Hampel Filtered)

Vowel Formant TTS Quality Metric

Phoneme		Central 75%	All Data Points
AA	(ɑ)	98.22%	95.84%
AE	(æ)	81.07%	86.69%
AH	(ʌ)	95.22%	93.24%
AO	(ɔ)	56.54%	98.03%
EH	(ɛ)	83.24%	92.04%
ER	(ɝ)	67.60%	72.20%
IH	(ɪ)	84.93%	74.91%
IY	(i)	93.61%	81.24%
UH	(ʊ)	86.52%	91.41%
UW	(u)	92.19%	87.44%

Table 11: Phoneme Space Overlap

Production and Perception of non-native English in China

Focus on Sociophonetic Variation by Humans and Artificial Agents

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

PhD Colloquium

12.12.2023

HYBRID SOCIETIES

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation