AUGMENTED REALITY IN LISTENING AND PRONUNCIATION TRAINING

Tormod Aagaard and Mikhail Fominykh

Norwegian University of Science and Technology

tormod.aagaard@ntnu.no, mikhail.fominykh@ntnu.no

Language teachers have always harnessed the powers of situated, multimodal communication, for example, by exploiting the iconic and indexical potential in gestures. The effectiveness of some of these pedagogical practices for improving pronunciation skills have also been corroborated in recent studies (Baills, F., Alazard-Guiu, C., & Prieto, P. 2022). Situated and multimodal interaction and communication are well supported by Augmented Reality (AR), an emerging technology that seamlessly adds digital content to the physical world (Williams&Ortega 2020). AR can support digital objects and characters programmed to be aware of the learner and the surrounding physical environment. AR thus can be used as a delivery platform for embodied conversational agents (ECA), which has maintained an interest among researchers and practitioners in computer assisted language learning for a long time (Wik and Hjalmarsson 2009). In the Augmented Reality Instructional Design for Language Learning project (ARIDLL https://aridll.eu), we explore the possibilities of using AR in language learning and experiment with them in practice.

The acoustic signal of spoken language is in most situations part of a total communicative package involving visual information from the articulators, gestures, body pose and gaze of the speaker. The McGurk effect (McGurk&MacDonald 1976) is a well-known example of how visual and acoustic signals are processed as a whole, but we are also now seeing increasing evidence for a tight link between gestures and speech, to the extent that the two modalities can be seen as parts of a unitary system (Biau, E., Morís Fernández, L., Holle, H., Avila, C., & Soto-Faraco, S. 2016, Kelly, S. D., Özyürek, A., & Maris, E. 2010, Marstaller & Burianova 2015). Moreover, the speech act is often supported by the affordances of a physical context. We propose an experimental AR-supported learning experience design that includes multimodal three-dimensional contextualized listening and pronunciation tasks.

References:

Baills, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy. Applied Linguistics, 43(4), 776-804. https://doi.org/10.1093/applin/amac010

Biau, E., Morís Fernández, L., Holle, H., Avila, C., & Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio–visual alignment are modulated by the communicative nature of the stimuli. NeuroImage, 132, 129-137. https://doi.org/https://doi.org/10.1016/j.neuroimage.2016.02.018

Chan, Marsha J. Embodied Pronunciation Learning: Research and Practice CATESOL Journal, v30 n1 p47-68 2018

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. Psychological Science, 21(2), 260–267. https://doi.org/10.1177/0956797609357327

Marstaller, L., & Burianová, H. (2015). A common functional neural network for overt production of speech and gesture. Neuroscience, 284, 29–41. https://doi.org/10.1016/j.neuroscience.2014.09.067

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. Nature, 264(5588), 746-748. https://doi.org/10.1038/264746a0

Parmaxi, A. and A. A. Demetriou (2020). "Augmented reality in language learning: A state-of-the-art review of 2014–2019." Journal of Computer Assisted Learning 36(6): 861-875.

Wik, Preben, and Anna Hjalmarsson (2009). 'Embodied conversational agents in computer assisted language learning', Speech Communication, 51: 1024-37.

Williams, A. S., & Ortega, F. R. (2020). Understanding Gesture and Speech Multimodal Interactions for Manipulation Tasks in Augmented Reality Using Unconstrained Elicitation. Proc. ACM Hum.-Comput. Interact., 4(ISS), Article 202.

PHONETIC VARIATION IN THE REALIZATION OF PHONOLOGICAL TONAL ACCENTS IN PROFICIENT L2 SPEAKERS OF NORWEGIAN

Teodor Ekblad Aagaard and Aïda Leistad Thomassen

Department of Linguistics and Scandinavian Studies, University of Oslo

t.e.aagaard@iln.uio.no, a.l.thomassen@iln.uio.no

Previous studies of Norwegian L2 intonation have treated realizations of tonal accents as categorical variables, i.e. either "correct" or "incorrect" realizations of accent 1 and accent 2, based either on L1 listener judgements (e.g. Hognestad, 2017) or on specific acoustic criteria (e.g. Steien & Van Dommelen, 2018). A problem with the former approach is that an L2 speaker might systematically distinguish accent 1 and accent 2 words intonationally, without L1 listeners recognizing the phonological contrast. A problem with the latter approach is that it is not clear exactly which acoustic variables should be measured.

We present an ongoing study of L2 intonation in which we combine L1 listener judgements with measuring a range of actual phonetic properties of accent phrases (APs), such as f0 range, alignment of f0 minima, etc. We aim to answer the following questions:

1. In APs judged correct with respect to tonal accent, what are the phonetic differences between accent 1 and 2?
2. How do APs judged incorrect differ phonetically from those judged correct?
3. In APs judged incorrect, are there still phonetic differences between accent 1 and 2?

We present preliminary data from a pilot study that suggest that L2 speakers who produce APs judged correct with respect to tonal accent, do not always employ the same phonetic strategies as L1 speakers. In addition, some L2 speakers

systematically distinguish between the accents by means of tone in APs that L1 listeners judge incorrect.

We hope that this study can inspire further investigation into the topic of tonal accents in an L2 perspective, and thus give us more knowledge about what phonetic factors are the most important when producing tonal accents. This might be valuable for educational settings of pronunciation instruction.

References:

Hognestad, J.K. (2017). Prosodiske strategier i norsk som andrespråk. *NOA – Norsk som andrespråk, 33* (1), 63–84.

Steien, G. B. & Van Dommelen, W. A. (2018). The production of Norwegian tones by multilingual non-native speakers. *International Journal of Bilingualism, 22* (3), 316–329.

GAMES COULD TEACH CHILDREN TO SPEAK THE NORDIC LANGUAGES

Yaroslav Getman[1], Nhan Phan[1], Ragheb Al-Ghezi[1], Ekaterina Voskoboinik[1], Mittul Singh[1], Tamás Grósz[1], Mikko Kurimo[1], Giampiero Salvi[2,5], Torbjørn Svendsen[2], Sofia Strömbergsson[3], Anna Smolander[4], and Sari Ylinen[4]

[1]Department of Information and Communications Engineering, Aalto University, Finland
[2]Signal processing, Norwegian University of Science and Technology, Norway
[3]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Sweden
[4]Logopedics, Welfare Sciences, Faculty of Social Sciences, Tampere University, Finland
[5]KTH Royal Institute of Technology, EECS, Stockholm, Sweden

Corresponding author: yaroslav.getman@aalto.fi

The advances in AI research have propelled the application of speech technology and gamification to the area of Computer-Assisted Language Learning (CALL). This offers students the means to refine their language skills, notably pronunciation, beyond the confines of a traditional classroom setting. In this talk we will report on recent results [1] obtained within the Teflon project [2]. We adapted state-ot-the-art automatic speech recognition technology to the tasks of online speech recognition and pronunciation assessment in a gamified mobile application intended for children. We tested the system on two categories of young speakers. The first group consists of Swedish children participating in speech therapy because they are diagnosed with Speech Sound Disorder. The second group consists of L2 children learning Swedish and Finnish words.

Our results show that the recently proposed end-to-end speech recognition models may be used to provide feedback to the students, and they are particularly convenient because they can perform speech recognition and some form of pronunciation assessment simultaneously. Because of the black-box nature of such models, however, the pronunciation assessment is limited to a global score. In order to provide an explanation of the model inference, we investigated the use of input

attribution algorithms which estimate the influence of each speech segment to the final decision. These insights allow the development of more accurate and trustworthy methods and may potentially be used to provide more detailed feedback to the students.

References:

[1] Getman, Y. et al. (2023). Developing an AI-Assisted Low-Resource Spoken Language Learning App for Children. *IEEE Access*, vol. 11, pp. 86025-86037. doi: 10.1109/ACCESS.2023.3304274.

[2] https://teflon.aalto.fi/

FLUENCY AND PRONUNCIATION IN LANGUAGE ASSESSMENT: HOW DO LANGUAGE TESTERS DEFINE THEM AND WHAT DO RATERS PAY ATTENTION TO?

Ari Huhta, invited speaker

University of Jyväskylä

ari.huhta@jyu.fi

The presentation discusses how pronunciation and fluency have been defined in the assessment of foreign or second language (L2). Since assessment can have a considerable washback on learning and teaching, it useful to look at how major language tests and examinations view these two key aspects of speaking and is known about how raters assess them. The talk is based on two kinds of analyses. The first involves a systematic analysis of practical language assessment instruments such as rating scales used by human raters; these provide us with the operational definitions of fluency and pronunciation that raters should pay attention to in their assessments. The analysis produces an overview of which features of pronunciation and fluency are included in the rating scales often vs rarely. The second type of analysis focuses on what research has found raters to actually focus on when they rate L2 learners' fluency and pronunciation. A summary is also given of the features that automated speaking assessment systems consider when evaluating L2 speech and how those compare with the features assessed by human raters. Furthermore, the presentation reports on the findings of two recent Finnish studies that investigated fluency and pronunciation. One of them investigated, e.g., the recognition of the examinees' first language by the raters and its effects on their ratings, while the other developed automated speech recognition and evaluation system for L2 Finnish and L2 Swedish.

The analysis of fluency is based (with some updates) on the chapter Fluency in Language Assessment by Huhta, Kallio, Ohranen and Ullakonoja in the book Fluency in L2 Learning and Use by Lintunen, Mutta and Peltonen (Multilingua Matters, 2019). A similar analysis of pronunciation in L2 speaking assessment was conducted for the purpose of this talk.

CAN DANISH SENIORS LEARN TO IDENTIFY NONNATIVE TONES?

Sidsel Holm Rasmussen and Ocke-Schwen Bohn

Department of English, Aarhus University

sira@cc.au.dk, ocke.bohn@cc.au.dk

Previous studies have reported succesful training of naïve listeners on nonnative tonal contrasts (e.g., Wang and Kuhl 2003; Sadakata and McQueen 2014; Godfroid, Lin, and Ryu 2017), but little is known about L1 Danish listeners' ability to acquire L2 tones. Additionally, current models of nonnative speech perception hypothesize that novel speech contrasts can be learned throughout an individual's lifetime (e.g., SLM-r by Flege and Bohn 2021), but very few studies have tested this prediction. Addressing these two points, the current study examines the perceptual acquisition of Mandarin Chinese tones by older adults (age 60+) in an internet-based auditory training experiment. A group of younger participants (age 20-40) will also be recruited for training, just as we will recruit two age matched control groups who will undergo no training.

Training groups will be asked to complete 10 session with 120 trials of High Variability Phonetic Training with corrective feedback in their own homes over the course of ca. 3 weeks. The perceptual training sessions are administered through the web-based program PERCY (Draxler 2014), through which we can monitor the progress of each trainee. Post-tests will establish potential benefits of training and will further probe any generalizability of training to untrained syllables and speakers.

This presentation discusses methodological considerations and introduces in more detail the experimental design of the current study. While data collection will most likely not be completed before the end of October 2023, preliminary data from the study will be presented. This work is part of a larger ongoing project "Perceptual Flexibility in Old Age" in which we investigate the phonological learning abilities of seniors to fill a gap in the field of L2 acquisition research.

References:

Draxler, Christoph. 2014. "Online Experiments with the Percy Software Framework - Experiences and Some Early Results." In Proceedings of the Ninth International Conference on Language Resources and Evaluation, 235–340. Reykjavik Iceland.

Flege, James Emil, and Ocke-Schwen Bohn. 2021. "The Revised Speech Learning Model (SLM-r)." In Second Language Speech Learning, edited by Ratree Wayland, 1st ed., 3–83. Cambridge University Press. https://doi.org/10.1017/9781108886901.002.

Godfroid, Aline, Chin-Hsi Lin, and Catherine Ryu. 2017. "Hearing and Seeing Tone Through Color: An Efficacy Study of Web-Based, Multimodal Chinese Tone Perception Training: Multimodal Tone Training." Language Learning 67 (4): 819–57. https://doi.org/10.1111/lang.12246.

Sadakata, Makiko, and James M. McQueen. 2014. "Individual Aptitude in Mandarin Lexical Tone Perception Predicts Effectiveness of High-Variability Training." Frontiers in Psychology 5 (November). https://doi.org/10.3389/fpsyg.2014.01318.

Wang, Yue, and Patricia K. Kuhl. 2003. "Evaluating the 'Critical Period' Hypothesis: Perceptual Learning of Mandarin Tones in American Adults and American Children at 6, 10 and 14 Years of Age." In Proceedings of the 15th International Congress of Phonetic Sciences, edited by Maria-Josep Solé, Daniel Recasens, and Joaquin Romero, 1537–40. Barcelona: Causal Productions.

L2 ENGLISH COMPREHENSIBILITY AND ACCENTEDNESS: WHAT DISTINGUISHES L1 FINLAND-SWEDISH AND L1 FINNISH LEARNERS?

Elina Tergujeff

University of Jyväskylä

elina.tergujeff@jyu.fi

After Munro and Derwing (1995) discovered that second language (L2) comprehensibility (i.e., perceived ease of understanding) and accentedness (perceived strength of foreign accent) are partly independent constructs, applied linguists have tried to gain more in-depth knowledge concerning speech features that contribute to comprehensibility in particular. So far, studies have associated fluency features and speech prosody to comprehensibility, whereas segmental accuracy has been found more strongly linked with accentedness (e.g., Saito et al., 2016). Little attention has been paid to comparing L2 comprehensibility or accentedness between different speaker groups. As a notable exception, Tergujeff (2021) discovered that English-speaking listeners found L1 Finland-Swedish learners' English easier to understand and less accented than L1 Finnish learners' English, despite the speakers' equal oral language proficiency assessment (B1-B2 on the CEFR scale). This finding gave inspiration to the present study, which explores if possible differences in fluency and/or segmental accuracy might explain the differences in comprehensibility and accentedness between L1 Finland-Swedish and L1 Finnish learners of English.

In this study, B1 and B2-level speech samples from Tergujeff (2021) were used as material. These were approximately 20-second samples of semi-spontaneous L2 English speech elicited from L1 Finland-Swedish (*n*=20) and L1 Finnish (*n*=20) teenagers. The speech samples were measured for 12 speech features including token frequency, speed of delivery, pausing, repairs and segmental accuracy. The measurements were contrasted between the two speaker groups by means of descriptive statistics and Mann-Whitney U. The results reveal statistically significant differences in token frequency, speech and articulation rate, and mean length of run: L1 Finland-Swedish speakers on average produced more words in total, spoke faster and in longer stretches between pauses. Consequently, the findings suggest

that fluency (rather than segmental accuracy) may partly explain why listeners find Finland-Swedes' English more comprehensible and less accented than Finnish speakers' English.

References:

Munro, M. J., Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning, 45(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x.

Saito, K., Trofimovich, P., Isaacs, T., (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. Applied Psycholiguistics, 37(2), 217–240. https://doi.org/10.1017/S0142716414000502.

Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s. System, 100, 1–11. https://doi.org/10.1016/j.system.2021.102567

FROM A MULTILINGUAL PERSPECTIVE: CROSS-LINGUISTIC INFLUENCE IN THE ACQUISITION OF L3 PHONETICS AND PHONOLOGY

Magdalena Wrembel (invited speaker)

Adam Mickiewicz University

magdala@amu.edu.pl

A complex linguistic landscape in the modern world has led to the development of a wider perspective in language acquisition research, going beyond the second language. A growing body of studies into the acquisition of third language phonetics/phonology demonstrates an inherent complexity of the field reflected, among others, in multidirectional dynamic cross-linguistic influence. This talk aims to provide a state-of-the-art overview of related studies as well as some theoretical and methodological considerations in research on L3 phonological acquisition.

The contribution will present new insights into the field that stem from an ongoing Polish-Norwegian project investigating patterns of speech of L1 Polish/L2 English/L3 Norwegian speakers (*Across-domain investigations in multilingualism: Modelling L3 acquisition in diverse settings - ADIM*). The overview will cover a range of studies including foreign accentedness ratings, perception (vowel assimilation paradigms and EEG evidence) as well as production studies investigating spectral overlap and separation in the three vocalic systems. The discussion of selected results will focus on developmental trajectories of L2 and L3 phonologies; complex cross-linguistic interactions

over time and the production and perception interface. Finally, potential implication for L2/L3 pronunciation pedagogy will be drawn.

LEARNERS' ATTITUDES TOWARDS NORWEGIAN ACCENTS

Magdalena Wrembel & Kamil Malarski

Adam Mickiewicz University in Poznań

magdala@amu.edu.pl, kamil.malarski@amu.edu.pl

Notwithstanding some popular opinions, accents in Norway are stratified socially (Johnsen 2015), e.g., the Western Oslo accents are perceived as more prestigious than the working-class or multicultural Oslo accents (Aasheim 1995, Johnsen 2015, Svendsen and Røyneland 2008), or accents in Northern Norway (Sollid 2014). Despite previous research on language attitudes to Norwegian accents (Gulbrandsen 1977, Lund 2006), we lack a more up-to-date perspective accounting for the context of multicultural and multilingual speakers or learners of Norwegian.

In this study, therefore, we aim at bridging this gap, comparing the perceptions towards Norwegian accents by three groups of respondents, including Polish learners of Norwegian living in Poland, Polish learners of Norwegian residing in Norway, as well as Norwegian native speakers as controls. We are interested in across-group differences to address the question to what extent the changing perceptions of these accents are dependent on social attributions of the speakers representing the accents. Through an online survey in Qualtrics, the respondents listened to ten samples of read speech (*Nordavinden og sola*) coming from five regions in Norway, i.e. the Tromsø area, Trondheim, Stavanger, Kristiansand, and Oslo, as well as four non-native accents of Norwegian of different strength. Each region was represented with two speech samples of middle-aged native speakers of Norwegian. The respondents were asked to identify the region of origin and to evaluate the samples according to perceived level of education, intelligence, nativeness and other criteria. Based on the preliminary results, we investigate the following: 1) Do learners of Norwegian attribute similar aesthetic judgments to Norwegian speech as Norwegian speakers do? 2) Are some accents of Norwegian perceived differently than others? 3) Are there any acoustic correlates of these judgments like high-pitched voice, female/male voice, the presence of uvular /r/ phonemes?

References:

Aasheim, S. 1995. "Kebab-norsk": Framandspråkleg påverknad på ungdomsspråket i Oslo. Master thesis, Oslo: University of Oslo.

Gulbrandsen, P. P. 1977. "Hvor går de øst-norske bymåla?" [Where are the urban dialects in Eastern Norway going?]. Språklig Samling 18: 14–17.

Johnsen, S. S. 2015. "Dialect change in South-East Norway and the role of attitude in diffusion", Journal of Sociolinguistics 19/5, 2015: 612–642

Lund, C.. 2006. 'Vi sier ikke Freksta, vi sier Freiksta': En variasjons- og endringsstudie av bymålet i Fredrikstad [A study of variation and change in the city dialect of Fredrikstad]. Unpublished MA thesis. Oslo, Norway: University of Oslo.

Sollid, H. 2014. "Hierarchical Dialect Encounters in Norway", Acta Borealia 31 (2): 111–130

Svendsen, B. A., & Røyneland, U. 2008. "Multiethnolectal facts and functions in Oslo, Norway", International Journal of Bilingualism, 12(1–2), 63–83.

FOREIGN ACCENTED SPEECH REFLECTED IN L2 WRITTEN PRODUCTION

Elisabeth Zetterholm

Linköping University

elisabeth.zetterholm@liu.se

A foreign accent is usually inevitable among adult second language learners (Moyer, 2013). The speaker's first language can be identified in accented speech based on transfer of linguistic features, especially in the early stages of second language acquisition (Jarvis & Pavlenko, 2010). The foreign accent might also affect written production in L2 in such a way that spelling errors can be related to the speaker's pronunciation in L2. The aim of this study was to investigate in what way segmental mispronunciation in foreign accented speech can be reflected in written production (misspellings).

Recordings of adult second language learners of Swedish were used for acoustic and auditory analyses. A dictation practice was conducted for the analysis of spelling errors in L2 written production. Spelling errors were categorized based on known pronunciation errors related to phonological transfer (e.g. Bannert, 1990; Zetterholm, 2022; in print). Results are of interest for didactic implementation in the teaching of Swedish as a second language.

The analyses indicate that there is a connection between accented speech and written production, e.g. exchange of the vowels y/i (*lyssnar* (listening) is pronounced and spelled *lisnar) and the consonants p/b (*packar* (packing) is pronounced and spelled *bakar). This is especially found among speakers with Arabic or Somali as their L1. Swedish fronted rounded vowels such as /y/ is often an articulation problem for L2 learners, regardless of L1, and found in both oral and written productions in this study. The exchange of p/b might depend on phonological differences between languages. Misspellings can also be compared between young first language learners and second language learners in the early stages of L2 acquisition (Manjón-Cabeza Cruz & Sosiński, 2021; Zetterholm, forthc.).

References:

Bannert, R. (1990). *På väg mot svenskt uttal*. Lund: Studentlitteratur.

Jarvis, S. & Pavlenko, A. (2010). *Crosslinguistic influence in language and cognition*. New York/London: Routledge.

Manjón-Cabeza Cruz, A. & Sosiński, M. (2021). Moroccan immigrants learning Spanish writing (compared with L1 children). *Journal of Second Language Writing* 51.

Moyer, A. (2013). *Foreign accent. The phenomenon of non-native speech*. Cambridge: Cambridge University Press.

Zetterholm, E. (2022). Uttal, fonologi och skrivande hos yngre flerspråkiga elever. In D. von Börtzell-Szuch & S. Vuorenpää (Eds.). *Läs- och skrivundervisning utifrån elevers varierade behov*, p. 53–79. Stockholm: Gleerups.

Zetterholm, E. (in print). Phonological transfer in oral and written production among adult L2 learners of Swedish.

Zetterholm, E. (forthc.). Spelling in the early stages of writing: a comparative study between adult L2 immigrants and L1 and L2 young children.