Fonetiikan päivät, Joensuu, 25.–26.08.2022.

Puhetta kaikille

Abstraktit

Table of Contents

Plenaari: Research and Technology Transfer in German Universities of Applied Sciences – Experiences in the Area of AI and Speech Technologies

Mathias Walther, Technical University of Applied Sciences (TH) Wildau

Germany is a modern knowledge society whose prosperity is largely based on the innovative strength of its economy. In order to successfully keep up the pace in international competition, technical, economic and social innovations must be realised. Universities play a central role in this process. In the context of the necessary modernisation of the European higher education system, the European Commission also sees universities as being centrally located in the knowledge triangle that spans the areas of teaching, research and innovation. On the one hand, the aim is to take on the technology needs of companies, especially small and medium-sized enterprises (SMEs), and to satisfy them through research and development (R&D) solutions from the university. On the other hand, R&D results from the university can be transferred to industry via the transfer service and developed into new products there. AI and Speech Technologies is an emerging field from different perspectives. The science behind is demanding and practical use is still in its infancy. Hence this is an interesting area for technology and knowledge transfer as well as government funding.

The talk will illustrate knowledge and technology transfer projects from both sides with examples: past experience as a researcher in a company and current experience as a professor at TH Wildau.

Transfer is understood as the mutual interactions between the university and their environment. This understanding of transfer is consistent with the third mission, which the universities pursue in addition to research and teaching. That means transfer is more than pure technology transfer, and not only companies, but also institutions, e. g. associations, public institutions and non-profit organizations can be partners in technology transfer. The transfer of knowledge and technology can have several forms, according to the project's needs. Cooperation can be initiated as research and/or development projects. These activities can also incorporate student's work e. g. in a thesis or a semester project. In the same vein a beneficial form of transfer for companies as well as students are dual studies, which have been experiencing a real boom for years. It does not confront young people with the decision: either study or in-company training, but enables a combination of theoretical university training and practical training in the company. The demand for modern AI and speech technology is high in SMEs, though their resources are limited. At this point sharing information and presentations are especially important. The TH Wildau is interested in various exchanges between people from science and companies or institutions at conferences, in networks, etc. This also can be done via academic continuing education which means cooperation in the training, further education and training of employees from the SMEs. Another way is consultancy and R&D services of TH Wildau scientists including use of infrastructures. In this case, TH Wildau can be a testbed, which is possibly supported by the Brandenburg innovation voucher. Usually after a longer process of R&D, licensing, exploitation of patents can be form of cooperation. The patent service offers professors, employees and students advice on all intellectual property issues with a focus on inventions and patents. Finally, the transfer of knowledge and technology can lead to spin-off-companies. Transfer activities of the TH Wildau are regional, national and international. The university is committed to its role in the development of the region. The university is also embedded in the capital region, which includes the federal states of Berlin and Brandenburg. Zooming into the region shows that the TH Wildau diverse and close cooperation relationships with partners from business, science and company in the southern part of Brandenburg and the airport region plays an important role. TH Wildau supports the Brandenburg region in boosting its competitiveness in the digital era, by acting as a contact point for two European Initiatives, the I4MS and the AI4EU. The I4MS is the initiative promoted by the European Commission to foster the digital innovation of manufacturing SMEs and midcaps in Europe in order to boost their competitiveness in the digital era. The AI4EU is the initiative promoted by the European Commission to create a supportive platform on AI for European organisations for sharing a repository of AI knowledge and align experts in AI research to support innovation and technology transfer to SMESs and low-tech SMEs.

Työpaja: Introduction to PsychoPy workshop

Eugenia Rykova, UEF / TH Wildau

1. What is PsychoPy? Installation. Builder vs Script.
2. Builder. Experiment structure and components. Experiment properties.
3. Building a routine with an audio-stimulus and response buttons.
4. Building a routine with an image-stimulus and recording an audio as a response.
5. Adding script elements. Editing the script.
6. Running experiment on the web.

An Automatic Speaking Assessment System for Spontaneous L2 Speech in Under-Resourced Languages (talk)

Ragheb Al-Ghezi [1], Katja Voskoboinik [1], Yaroslav Getman [1], Clara Akiki [1], Anna von Zansen [2], Raili Hildén [2], Ari Huhta [3], Heini Kallio [3], Mikko Kuronen [3], and Mikko Kurimo [1]

1 Aalto University, 2 University of Helsinki, 3 University of Jyväskylä

Developing automatic systems for assessing spontaneous spoken utterances is important for second language learning, because it promotes and democratizes self-regulated learning and can serve as an auxiliary tool in language proficiency assessment and teacher training. While such systems are typically developed for languages with a large number of learners such as English, the languages with fewer learners such as Finnish and Swedish remain at a disadvantage due to the lack of training data. Nevertheless, due to recent advancements in self-supervised machine learning methods (Devlin, J et al, 2018; Conneau, A. et al,. 2020; Baevski, A. et al., 2020; Al-Ghezi et al., 2021) it is now possible to develop automatic speech recognition systems without a large amount of annotated training data. This means that it could also be now feasible to develop automatic speaking assessment systems also for under-resourced languages.

In the DigiTala project, we have designed and implemented an automatic speaking assessment system for spontaneous L2 speech in Finnish and Finland Swedish. Furthermore, we have developed tools to be able to provide accurate personalized feedback to the language learners. In this presentation, we briefly describe and evaluate the main components of the system.

The automatic assessment system comprises several machine learning models: (1) an Automatic Speech Recogniser that converts the spoken utterances into written transcripts; (2) Lexico-grammatical Accuracy and Range Evaluators to the transcribed responses by leveraging textual features; (3) Pronunciation and Fluency Evaluators to the transcribed responses by leveraging acoustic and prosodic features; (4) a Task Accomplishment Evaluator to evaluate test-taker adherence to the task assignment; and (5) a separate Evaluator for the final CEFR-like score representing the holistic overall speaking proficiency.

In this work the performance of the automatic system is evaluated in terms of system-human agreement. Each of the comprising machine learning models and the overall score for each response in the test set is compared to the corresponding results provided by human raters. The performance of the automatic speech-to-text conversion is measured using word and character error rate compared to a manual transcript. Furthermore, the most important input features for the machine learning models are determined in order to be able to return more accurate feedback for the students and teachers.

References:

1. Al-Ghezi, R., Getman, Y., Rouhe, A., Hildén, R., & Kurimo, M. (2021). Self-Supervised End-to-End ASR for Low Resource L2 Swedish. *Proc. Interspeech 2021*, 1429-1433.
2. Al-Ghezi, R., Getman, Y., Singh, M., & Kurimo, M. (2022). Automatic Rating of Spontaneous Speech for Low-Resource Languages. (in review)
3. Al-Ghezi, R., Vosboinik, K., Getman, Y., von Zansen, A., Kallio, H., Clara, A., Kuronen, M., Huhta, A., & Hilden, R. (2022). Automatic speaking assessment of Spontaneus L2 Finnish and Swedish. Manuscript submitted for publication.(in review)
4. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *arXiv preprint arXiv:2006.11477*.
5. Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.
6. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

FinSyn: A large scale corpus of Finnish speech material for text-to-speech synthesis

Atte Asikainen, Tuukka Törö, Juraj Šimko, Martti Vainio, Krister Lindén, Antti Suni
University of Helsinki

In order to generate high quality speech output, modern end-to-end deep-learning-based text-to-speech synthesis systems require a relatively high amount of good quality speech data. The Language Bank of Finland (Kielipankki) and the Phonetics and Speech Synthesis Group of the University of Helsinki recently collected a large corpus of Finnish spoken data, designed to satisfy the requirements for building state-of-the-art speech synthesis systems in Finnish language.

The corpus consists of recordings of two professional female speakers of Finnish, approximately 25 hours for each. To facilitate synthesis of prosodically rich output, the corpus contains prosodically varied material, covering a wide range of speaking styles from formal reading of Wikipedia text to informal discussions between the voice talents. Most of the material consists of longer stretches of communicative narratives rather than more traditional phonetically balanced isolated sentences.

We will present the ideas behind and the details of the corpus design as well as some relevant aspects of the recording setup, and post-processing of the recorded material. In addition, we will demonstrate several examples of speech synthesized using the corpus, primarily focusing on control of prosodic and stylistic variation of the generated speech.

Developing an Automatic Speaking Assessment System for L2 Speech (poster with demo)

Yaroslav Getman [1], Ragheb Al-Ghezi [1], Katja Voskoboinik [1], Clara Akiki [1], Anna von Zansen [2], Raili Hildén [2], Ari Huhta [3], Heini Kallio [3], Mikko Kuronen [3], and Mikko Kurimo [1]

1 Aalto University, 2 University of Helsinki, 3 University of Jyväskylä

This presentation describes the main steps we took to implement an automatic speaking assessment system for second language learners in the DigiTala project. So far, we have developed systems for two languages, Finnish and Finland Swedish. In addition to the necessary tools for offline assessments we have also implemented an online system using a Moodle plugin that will be demonstrated in the conference. Our poster describes the steps taken to develop the system:

1. **Designing suitable speaking tasks** for data collection and speaking assessment for the target groups. In total, the Swedish and the Finnish tests included 22 and 43 unique tasks, respectively, to match various skill levels of the students.
2. **Collecting training data for Automatic speech recognition (ASR) and Evaluators for the analytic evaluation dimensions.** For collecting speech data, a specially tailored students' Moodle system was developed and utilized in the collection of the new Finnish data in 207 high school and 116 Aalto University students The total amount of speech data for Finnish includes 4849 samples and 19.5 hours
3. **Designing the transcription guidelines and transcribing the speech data.** The transcribers were also instructed about special tags and symbols for almost correct words, words from foreign languages, as well as non-speech events and various disfluencies.
4. **Designing the rating scales and manually rating the student responses** with the teachers' Moodle system that was developed for the project. The recordings were distributed among the raters with a partial overlap in order to be able to monitor the inter-rater reliability, as well as compute fair average scores using Many-facet Rasch measurement (MFRM).
5. **Training and testing the ASR and Evaluators.** The ASR system was implemented using monolingual and multilingual pre-trained self-supervised models and PyTorch implementation of end-to-end CTC-based ASR system (Al-Ghezi et al 2021). Each analytic criterion involved a separate machine learning model trained with the fair average scores provided by the raters. In addition, a separate system was trained to predict the holistic score (Al-Ghezi et al 2022).
6. **Implementing an ASR-based automatic speaking assessment server back-end** that is able to return the ASR transcript, evaluation scores for the analytic dimensions and holistic overall score for the test-takers in real-time. It can also compute more accurate feedback based on the measured feature valuess.
7. **Implementing a demonstration system front-end with a Moodle-plugin** interface (von Zansen et al 2022) for ASR and Evaluator API. This serves as the front-end for the test-takers to use the back-end speaking assessment server.
8. **Collecting feedback from the test-takers and teachers.** To collect the feedback we use online questionnaires and face-to-face interviews. Initial analysis of the feedback will be available for the presentation.

References:

1. Al-Ghezi, R., Getman, Y., Rouhe, A., Hildén, R., & Kurimo, M. (2021). Self-Supervised End-to-End ASR for Low Resource L2 Swedish. *Proc. Interspeech 2021*, 1429-1433.
2. Al-Ghezi, R., Voskoboinik, K., Getman, Y., von Zansen, A., Kallio, H., Clara, A., Kuronen, M., Huhta, A., & Hilden, R. (2022). Automatic speaking assessment of Spontaneus L2 Finnish and Swedish. Manuscript submitted for publication. [in review]
3. von Zansen, A., Alanen, T., Al-Ghezi, R., Erkkilä, J., Harjunpää, T., Heijala, M., Kallio, H. (2022). DigiTala Moodle plugin. https://github.com/aalto-speech/moodle-mod_digitala

Namibialaisten kielten tallentaminen ja kielten arvostuksen lisääminen haastattelujen ja tanssin keinoin

Katja Haapanen, Antti Saloranta, Kimmo U. Peltola, Henna Tamminen ja Maija S. Peltola

Fonetiikka ja Learning, Age & Bilingualism -laboratorio, Turun yliopisto

Namibiassa puhutaan noin kolmeakymmentä eri kieltä, joista osa on alkuperäisiä bantu- ja khoisankieliä, ja osa on siirtomaa-aikaisia germaanisia kieliä. Maan virallinen kieli on englanti, jonka lisäksi kymmenen yleisintä paikalliskieltä ovat virallisen aseman saaneita koulukieliä, joita käytetään opetuksessa koulun ensimmäisillä luokilla. Ylemmillä luokilla opetuskieleksi vaihtuu englanti (Norro, 2021, 2022). Monilla alkuperäiskielillä on vain vähän puhujia ja ne ovat vaarassa kuihtua. Lisäksi monien paikalliskielten arvostus englannin virallisen valtaaseman rinnalla on alhaista.

Tanssi uhanalaisten kielten ja foneettisen maailman tulkkina (T&T&F) on Turun yliopiston fonetiikan Learning, Age & Bilingualism -laboratorion (LAB-lab) kolmevuotinen hanke, jota rahoittaa Koneen säätiö. Hankkeen yhtenä päätavoitteena on tallentaa Namibian uhanalaisimpia kieliä sekä lisätä paikalliskielten arvostusta ja yleistä kielitietoisuutta. Tämän tavoitteen saavuttamiseksi tallennamme suullisia kertomuksia haastattelujen muodossa historiallisista tapahtumista eri paikalliskielten puhujilta. Haastattelut käsittelevät kulttuurihistoriallisesti ja folkloristisesti kiinnostavia aiheita liittyen Suomen ja suomalaisten läsnäoloon Namibiassa. Haastatteluilla on sekä kielitieteellistä että kulttuurihistoriallista arvoa ja ne arkistoidaan sekä suomalaisten että namibialaisten tutkijoiden käyttöön. Lisäksi haastattelumateriaalia käytetään kielitietoisuutta ja kielten arvostusta korostavassa tanssiteatteriesityksessä, jonka toteuttaa Tanssiteatteri ERI.

Haastattelut aloitettiin Namibiassa keväällä 2022. Ensimmäisellä aineistonkeruumatkalla tutkimukseen osallistui kuusi vapaaehtoista haastateltavaa. Haastatteluissa vapaaehtoisia pyydettiin kertomaan omia muistojaan ja kokemuksiaan tutkimuksen teemoihin liittyen. Tutkimme kielten foneettisia piirteitä analysoimalla ääninäytteitä akustisesti ymmärtääksemme eri kielitaustojen luomia lähtökohtia englanninkieliseen opetukseen namibialaisissa kouluissa. Aineistonkeruuta jatketaan koko hankkeen ajan ja tavoitteena on tallentaa mahdollisimman monipuolisesti eri paikalliskieliä foneettista analyysia, arkistointia ja kielitietoisuuta korostavaa tanssiesitystä varten.

Posteriesityksessämme esittelemme alustavaa haastatteluaineistoa, ääninäytteitä sekä tutkimuksen seuraavia vaiheita tallenteiden foneettiseen analyysiin liittyen.

Lähteet

Norro, S. (2022). Factors affecting language policy choices in the multilingual context of Namibia: English as the official language and medium of instruction. *Apples-Journal of Applied Language Studies*.

Norro, S. (2021). Kielten kirjo haasteena Namibian kouluissa. *Kieli, koulutus ja yhteiskunta*, *12*(4).

Building open-source speech technology for low-resource minority languages with Sámi as an example – tools, methods and experiments

Katri Hiovain–Asikainen, Universitet i Tromsø/University of Helsinki

The current paper will describe ongoing work for developing open-source speech technology applications for two Sámi languages, Lule and North Sámi. Lule and North Sámi are neighboring languages, spoken in the northernmost parts of Scandinavia. While Lule Sámi is spoken in Norway and Sweden, North Sámi is spoken in three countries: Norway, Sweden and Finland. For both languages, generally all speakers are bilingual in Sámi and at least one of the majority languages: Norwegian, Swedish or Finnish. The two languages are structurally similar, and after some training, they are mutually intelligible to some extent. However, as a part of language revitalization and preservation as well as accelerating digitalization, separate languages need separate language and speech technology tools to meet the needs of modern language users.

Lule and North Sámi differ remarkably in terms of the amount of speakers or language users. According to Ethnologue (Lewis, 2009), North Sámi has by far the largest number of language users: 25 000 in all three countries. Lule Sámi, on the other hand, has considerably fewer speakers: total of 2000 in both countries it is spoken in. All Sámi languages are classified as endangered by UNESCO (Moseley, 2010) and Lule Sámi as severely endangered. Perhaps consequently, as North Sámi has most language users among the Sámi languages, it has also most language resources and tools available. An infrastructure of dictionaries, morphological analyzers, spell checkers and language learning tools etc. have been maintained and developed since 2001 by the Divvun and Giellatekno groups[1].

A Text-to-speech tool is made to be able to synthesize intelligible speech output from any unseen text input in a particular language. A key objective for developing speech technology tools for indigenous languages generally is to meet the needs of modern language users in all language communities equally. For the Sámi languages, this would mean equal possibilities to use Sámi in the same digital contexts as the majority languages are being used. In this way, developing speech and language technology tools for the Sámi languages also contribute to the revitalisation of these languages. Additionally, speech technology tools are important for many language users, also those with special needs. These include language learners, people with dyslexia, vision impaired individuals, (native) users of the language that are not used to read Sámi etc. Additionally, speech technology is bringing more accessibility to many kinds of contents and utilities: a user can for example choose to listen to the news instead of reading the text, or a speech synthesis tool could be integrated into an online dictionary to allow listening to the correct pronunciations of the words.

The first Text-to-speech (TTS) tool for the Sámi languages was developed in 2015 for North Sámi by Divvun and Acapela. This tool is produced as closed-source and thus neither the framework used to develop the tool nor the speech corpus used for it are publicly available. Also, the company has ended support for certain operating systems. For this reason, we are now working on a modern, open-source TTS system that could be openly available for anyone who wants to develop speech technology for minority languages. At this stage, we have designed and collected a text corpus specifically for developing speech technology applications, namely Text-to-speech (TTS) and Automatic speech recognition (ASR) for the Lule and North Sámi languages. A new Lule Sámi speech corpus is going to be built from scratch during the year 2022. We have also piloted and experimented with different speech synthesis technologies using a miniature speech corpus as well as developed tools for effective processing of large spoken corpora. Additionally, we discuss effective and mindful use of the speech corpus and also possibilities to use found/archive materials for training an ASR model for these languages.

References:

1) https://github.com/giellalt, https://divvun.no/fi/, https://giellatekno.uit.no/

Lewis, M. P. (2009). *Ethnologue: Languages of the world*. SIL international.

Moseley, C. (Ed.). (2010). *Atlas of the World's Languages in Danger*. Unesco.

Suomen vokaalien artikulografinen pilottitutkimus: Carstensin AG501:n käytöstä kvantaaliteorian tutkimuksessa

Satu Hopponen, Alexandre Nikolaev ja Marianne Hyppönen, UEF

Itä-Suomen yliopisto hankki äskettäin Puheen ja kielentutkimuksen laboratorioon uuden artikulografin. Se on Carstens Medizinelektronikin mallia AG501 (2014), joka on tarkempi kuin saman valmistajan vanhempi AG500-laite (Sigona et al. 2018). Aloimme heti suunnitella kuinka käyttäisimme artikulografia kielen ja huulien liikkeiden kartoittamiseen suomen vokaalien ääntämyksessä. Koska lähtökohtanamme on Stevensin kvantaaliteoria (mm. Stevens 1972; Stevens & Keyser 2010), meitä kiinnostavat erityisesti ääntöväylän liikkeiden ja vokaalien kolmen ensimmäisen formantin väliset suhteet. Kyseisen teorian empiirinen tutkimus on keskittynyt englantiin, joten sen oletusten tutkiminen vokaalijärjestelmältään erilaisen kielen kontekstissa on tarpeen. Artikulografi on oivallinen väline tällaiseen tutkimukseen, sillä se ei aiheuttane suurta muutosta koehenkilön ääntämykseen ja ääni voidaan samalla tallentaa hyvälaatuisena.

Noudatimme Wangin ym. julkaisussa (Wang et al. 2016) kuvailtuja sensorien paikkoja. Ainoan muutoksen, jonka jouduimme laitemallista johtuen tekemään, oli referenssisensorien paikka (he käyttivät laseja, me korvien takana olevia os mastoideuksia ja nenänvartta). Käytimme yhteensä neljäätoista sensoria. Takimmaisen kieleen liimattavan sensorin paikka osoittautui ongelmaksi. Aikuisen osallistujan suu ei välttämättä aukene riittävästi ja/tai hän ei kykene työntämään kieltään ulos niin pitkälle, että liimaaminen onnistuu ilman kohtuutonta epämukavuutta tai oksennusrefleksin aktivaatiota. Niinpä takimmainen sensori sijoitettiin hieman etisempään paikkaan. Sensorit pysyivät hyvin paikallaan ja kerätty data oli laadultaan hyvää.

Tässä kokeilussa oli vain yksi osallistuja; varsinaisessa tutkimuksessa heitä tulee olemaan enemmän. Tulevaisuudessa aiomme lisätä Wangin ym. kuvailemaan sensoriasetelmaan yhden lisää, jotta voimme kartoittaa myös alaleuan liikkeet. Tämä on mahdollista, sillä käytössämme olevista sensoripalikoista jäi kaksi sensoria vapaaksi. Hypoteesimme on, että näemme suomessa samoja tai samankaltaisia ilmiöitä, joita kvantaaliteoria kuvailee, mutta oletamme, että pyöreiden etuvokaalien [y ø] osalta asiat saattavat olla toisin.

Viitteet:

Sigona, Francesco, Massimo Stella, Antonio Stella, Paolo Bernardini, Barbara Gili Fivela & Mirko Grimaldi. 2018. Assessing the position tracking reliability of Carstens' AG500 and AG501 electromagnetic articulographs during constrained movements and speech tasks. *Speech Communication* 104. 73–88. https://doi.org/10.1016/j.specom.2018.10.001.

Stevens, Kenneth N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Edward David & Peter B. Denes (eds.), *Human communication: A unified view* (InterUniversity Electronics Series Vol. 15), 51–66. New York: McGraw-Hill.

Stevens, Kenneth N. & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19. http://dx.doi.org/10.1016/j.wocn.2008.10.004.

Wang, Jun, Ashok Samal, Panying Rong & Jordan R. Green. 2016. An Optimal Set of Flesh Points on Tongue and Lips for Speech-Movement Classification. *Journal of Speech, Language, and Hearing Research* 59. 15–26.

2014. *AG501 Manual (PD-01.03)*. Carstens Medizinelektronik. http://articulograph.de.

Kognitiivisen kuormituksen vaikutukset helikopterilentäjän puheen prosodiapiirteisiin

Marianne Hyppönen, UEF

Lentomiehistön kommunikaatio puheen muodossa on keskeinen osa ilmailua ja sen selkeys ja ymmärrettävyys on yhteydessä lentoturvallisuuteen. Lentotehtäviin liittyy usein kognitiivista kuormitusta, jonka on havaittu vaikuttavan ihmisen tiedonkäsittelyyn havainnoinnin, loogisen päättelyn, muistin ja oppimisen alueilla sekä puheen prosodiapiirteisiin. Koska puheen ymmärrettävyys liittyy lentoturvallisuuteen, on tärkeätä selvittää kognitiivisen kuormituksen vaikutukset puheen prosodiapiirteisiin.

Prosodisella lähenemisellä (engl. *prosodic entrainment*) tarkoitetaan puheen prosodiapiirteiden samankaltaisuuden lisääntymistä tai vähenemistä puhekumppanien välillä keskustelun aikana. Koska kommunikaatio puheen muodossa liittyy jokaiseen lentotehtävään, on todennäköistä, että niiden aikana esiintyy prosodista lähenemistä. Kognitiivisen kuormituksen vaikutuksia prosodiseen lähenemiseen ei ole aiemmin tutkittu.

Koneoppimista käytetään luomaan malleja, jotka tuottavat ennusteita uuden datan käyttäytymisestä. Tutkimuksessani on tarkoitus luokitella kerätyt puhenäytteet sen perusteella, esiintyykö niissä riskejä lentoturvallisuudelle käyttäen logistista regressiota. Samalla tarkastellaan, millaiset puheen prosodiset piirteet tai niiden yhdistelmät esiintyvät eri tilanteissa. Saatua tietoa on mahdollista hyödyntää ilmailussa riskitilanteiden tunnistamisessa.

Tutkimuskysymykset ovat seuraavat:

1. Millaisia muutoksia lentomiehistön puheen prosodiapiirteissä esiintyy kognitiivisen kuormituksen kasvaessa?
2. Miten kognitiivisen kuormituksen muutokset vaikuttavat prosodiseen lähenemiseen?
3. Millaiset prosodiapiirteet ja niiden yhdistelmät ovat tyypillisiä kognitiivisesti kuormittaville tilanteille?

Tutkimusaineisto tätä tutkimusta varten kerätään tallentamalla helikopterilentäjien puhetta lentosimulaattorissa suoritettavien lentotehtävien aikana. Aineiston keruu on suunniteltu tapahtuvan yrityksessä nimeltä Coptersafety Oy, joka tarjoaa simulaattorilentokoulutusta helikopterilentäjille.

Tallennetut puhenäytteet annotoidaan, jotta akustisten mittausten suorittaminen niille mahdollistuisi. Tämän jälkeen näytteistä mitataan puheen perustaajuus (puheen perustaajuuden keskiarvo, keskihajonta, vaihteluväli, HNR, jitter, shimmer), puheen intensiteetti (puheen intensiteetin keskiarvo, keskihajonta, vaihteluväli, energia <1kHz, painopiste) sekä puhenopeus (puhenopeus, artikulaationopeus, taukojen määrä ja kesto). Kyseiset muuttujat on valittu, koska aiemmissa tutkimuksissa on havaittu yhteyksiä kognitiivisen kuormituksen ja näiden puheen prosodiapiirteiden välillä.

Prosodisen lähenemisen astetta tarkastellaan edellä esiteltyjen muuttujien suhteen puheenvuorojen alussa ja lopussa keskustelun edetessä, jotta saataisiin selville, lisääntyykö vai väheneekö samankaltaisuus niissä keskustelukumppanien välillä. Kognitiivisen kuormituksen astetta lentotehtävien aikana seurataan mittaamalla lentäjien sykettä. Taustamuuttujina tarkastellaan lentäjien sukupuolta, ikää, lentotuntien määrää, yhdessä lennetyt lentotuntien määrää sekä lentäjän äidinkieltä.

Les voyelles nasales du français produites par les étudiants finnophones – Ranskan nasaalivokaalit suomenkielisten opiskelijoiden tuottamina

Akseli Häärä & Michael O'Dell
Tampereen yliopisto & Helsingin yliopisto

Tässä tutkielmassa tutkitaan sitä, kuinka suomea äidinkielenään puhuvat ranskan kielen yliopistoopiskelijat tuottavat ranskan kielen nasaalivokaaleita. Nasaalivokaalit ovat merkittävä ominaispiirre ranskan fonologiassa, kun taas suomen kielessä niitä ei esiinny lainkaan. Tämän vuoksi olisi tärkeää selvittää, esiintyykö nasaalivokaaleiden tuottamisessa ongelmia, kun suomea äidinkielenään puhuvat ranskan L2-opiskelijat tuottavat niitä. Tutkimuksen tarkoituksena olisi selvittää, millaisia nämä mahdolliset ääntämisongelmat ovat. Näiden ongelmien tunnistaminen voisi jatkossa auttaa nasaalivokaaleiden opetuksessa suomea äidinkielenään puhuville L2-ranskan opiskelijoille, sillä opettaja voisi varautua näihin ongelmiin jo ennalta.

Luonteeltaan tutkimus on kontrastiivisen fonetiikan tutkimus, jossa natiivipuhujien tuottamia äänteitä verrataan tutkittavan ryhmän, eli suomalaisten ranskan kielen yliopisto-opiskelijoiden tuottamiin äänteisiin. Kontrastiivisen analyysin menetelmän mukaisesti on tutkimuksessa kontrastiivisen analyysin hypoteesi. Tutkimuksessa oletetaan, että ääntämisessä ilmenee joitakin ongelmia, sillä suomen kielessä ei esiinny nasaalivokaaleita merkityksiä erottavana tekijänä. Hypoteesi on, että suomen kielen vastaavat oraaliset vokaalit aiheuttavat mahdollisesti interferenssiä, jonka seurauksena koehenkilöiden tuottamiin nasaalivokaaleihin tulee piirteitä suomen oraalivokaaleista.

Tutkimuksen aineisto koostuu kahden ranskan kielen natiivin ja 11 suomea äidinkielenään puhuvan ranskan kielen opiskelijan tuottamista yksitavuisista sanoista, jotka sisältävät ranskan kielen nasaalivokaaleita /ɑ̃/, /ɔ̃/ ja /ɛ̃/ tai niiden oraalisia variantteja. Äänitteitä tarkastellaan Praatohjelman avulla, ja tarkastelussa kiinnitetään huomiota vokaalien akustisiin piirteisiin, joita mitataan Stylerin kehittämällä skriptillä (ks. Styler 2017). Näin pyritään löytämään mahdollisia eroavaisuuksia natiivien ja L2-opiskelijoiden tuotosten välillä. Päivillä kerrotaan, minkälaisia eroja löytyy ja mitä niistä voi päätellä.

VIITTEET

Styler, Will (2017) ”On the Acoustical Features of Vowel Nasality in English and French”. *Journal of the Acoustical Society of America*. 142(4):2469-2482.

DigiTala – suullisen kielitaidon automaattinen arviointi kotimaisissa kielikonteksteissa

Heini Kallio[1], Anna von Zansen[2], Ragheb Al-Ghezi[3], Ekaterina Voskoboinik[3], Yaroslav Getman[3], Ari Huhta[1], Mikko Kuronen[1], Mikko Kurimo[3], Raili Hildén[2]

Jyväskylän yliopisto[1], Helsingin yliopisto [2], Aalto-yliopisto [3]

DigiTala-hanke on Helsingin yliopiston, Aalto-yliopiston ja Jyväskylän yliopiston yhteinen hanke, jossa kehitetään automaattiseen puheentunnistukseen perustuvaa digitaalista apuvälinettä suomen ja ruotsin suullisen kielitaidon arviointiin [1]. Hanke yhdistää pedagogiikan, puheteknologian ja fonetiikan asiantuntijoita. Esitelmässämme kerromme projektin etenemisestä sekä foneettisten ja kielipedagogisten tutkimusten päätuloksista [2,3,4,5]. Esittelemme myös automaattisen arviointityökalun prototyypin [6].

Hanke alkoi vuonna 2015, ja Suomen Akatemia rahoittaa tutkimushanketta 2019–2023. Projektin aikana on kehitetty useita digitaalisia puhumisen kokeita sekä arviointikriteereitä suullisen kielitaidon ulottuvuuksille sekä kerätty suuri puhe- ja arviointiaineisto. Digitaalisiin puhekokeisiin on osallistunut suomea ja ruotsia toisena kielenä opiskelevia lukiolaisia sekä akateemisia S2alkeisoppijoita (N = 710). Lisäksi aineistoa on saatu Yleisten Kielitutkintojen puhekokeista.

Puhe- ja arviointiaineistojen avulla on tutkittu ihmisarvioihin vaikuttavia puheen piirteitä, kuten temporaalista sujuvuutta, rytmiparametreja ja f0:n muutoksia [2,3]. Painopiste on spontaanissa suomen- ja ruotsinkielisessä puheessa, joissa akustisten piirteiden ja havaitun sujuvuuden tai taitotason välisiä yhteyksiä on tutkittu vasta vähän.

Sekä kielenoppijoilta että asiantuntija-arvioijilta on kerätty kokemuksia digitaalisesta puhumisen kokeesta ja käsityksiä automaattisesta arvioinnista. Tutkimuksista selviää, miten oppijat ja arvioijat suhtautuvat automaattiseen suullisen kielitaidon arviointiin [4,5]. Lisäksi tutkimusten avulla kehitetään digitaalista suullisen kielitaidon arviointiprosessia käyttäjäystävällisemmäksi.

[1] Kautonen, M., & von Zansen, A. (2020). *DigiTala research project: automatic speech recognition in assessing L2 speaking.* Kieli, koulutus ja yhteiskunta, 11 (4). https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitalaresearch-project-automatic-speech-recognition-in-assessing-l2-speaking

[2] Kallio, H., Suviranta, R., von Zansen, A. & Kuronen, M. (2022). *Creaky voice and fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish.* Proceedings of Speech Prosody 2022.

[3] Kallio, H., Kautonen, M., & Kuronen, M. (arvioitavana). *Prosody and fluency of Finland Swedish as a second language: investigating global parameters for automated speaking assessment.*

[4] von Zansen, A., Hildén, R., & Sneck, M. (tulossa). *Lukiolaisten käsitykset ja heidän antamansa palaute suullisen kielitaidon automaattisesta arvioinnista.*

[5] von Zansen, A., Kallio, H., Sneck, M., Kuronen, M., Huhta, A. & Hildén, R. (arvioitavana). *Ihmisarvioijien näkemyksiä suullisen kielitaidon automaattisesta arvioinnista, digitaalisesta arviointiprosessista sekä puhesuorituksista arvioitavista ulottuvuuksista.*

[6] von Zansen, A., Alanen, T., Al-Ghezi, R., Erkkilä, J., Harjunpää, T., Heijala, M., & Kallio, H. (2022). *DigiTala Moodle plugin*. https://github.com/aalto-speech/moodle-mod_digitala

Investigating language-specific coarticulatory patterns: A comparative ultrasound study of Finnish and English

Sonja Dahlgren[1], Pertti Palo[2], Minnaleena Toivola[3]

Free University of Bozen-Bolzano[1], Indiana University Bloomington[2], University of Helsinki[3]

We investigate a hypothesis for languages to be categorised based on their coarticulation directions i.e. whether coarticulation mainly occurs from consonant to vowel or from vowel to consonant, and whether this is connected to specific vowel reduction and stress types. In the current study we report our preliminary findings of tongue ultrasound studies on Finnish and (Standard American) English coarticulatory patterns, and compare the results to a previous study of Greek coarticulation (Scobbie and Sfakianaki 2013). Previous speech studies offer evidence of language-specific preferences for the direction of coarticulation (Öhman 1966, Manuel 1999, Traunmüller's 1999, Hardcastle and Hewlett 1999). Our hypothesis is that coarticulatory preferences mainly result from the ratio between vowels and consonants in the phoneme inventory (cf. Maddieson 2013), creating differing needs of phonological contrast and leading to language-specific word formational and inflectional patterns. For example, Finnish with its eight vowels and 11-12 (dialect dependent) native consonants (Suomi et al. 2008: 25) is *vocalic* and Arabic with its three to five vowels but up to 28 consonants (e.g. Ryding 2005: 25-26) is *consonantal*.

The current study adds to the evidence of language-specific coarticulatory patterns. We compare two further *vocalic* languages to the *vocalic* Modern Greek using closely corresponding utterances: L1 Finnish and L1 (Standard American) English. The chosen languages span the continuum of stress- vs. syllable-timing and differ in vowel reduction. Finnish has traditionally been considered syllable-timed but also shows signs of stress-timing (O'Dell and Nieminen 1998) and (Modern) Greek lies in between syllable- and stress-timed stress types (Grabe and Low 2002: 515-546); neither language has phonemic vowel quality reduction. In contrast, English is stress-timed with phonemic vowel quality reduction (e.g. Auer 2001: 1391-1393). According to our preliminary findings, English has similar vowel-to-consonant coarticulation to Finnish and Greek, which means that neither the stress system of a language, nor its tendency for vowel reduction, can necessarily predict its coarticulatory bias.

References

Auer, Peter. 2001. Silben- und akzentzählende Sprachen. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible (eds.), *Language typology and language universals. An international handbook*, 1391-1399. Berlin: De Gruyter.

Grabe, Esther & Ee Ling Low. Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology* 7. 515-546.

Hardcastle, William J. & Nigel Hewlett (eds.). 1999. *Coarticulation. Theory, data and techniques*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2013. Consonant-Vowel ratio. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Manuel, Shannon. 1999. Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In Hardcastle, William J. & Nigel Hewlett (eds.), *Coarticulation. Theory, data and techniques*, 179-198. Cambridge: Cambridge University Press.

O'Dell, Michael L. & Tommi Nieminen. 1998. Reasons for an underlying unity in rhythm dichotomy. In Proceedings of the Finnic Phonetics Symposium, August 1 1-14, 1998, Parnu, Estonia, Linguistica Uralica 34, 178-185.

Öhman, Sven EG. 1966. Coarticulation in VCV utterances: Spectrographic measurements.

The Journal of the Acoustical Society of America 39 (1). 151-168.

Ryding, Karin C. 2005. A *reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Scobbie, James & Anna Sfakianaki. 2013. A single speaker study of Modern Greek: The effect of stress, syllable position and consonantal context on vowels, and of vowel context on intervocalic consonants. Ultrafest VI.

Suomi, Kari, Juhani Toivanen & Riikka Ylitalo. 2008. *Finnish Sound Structure – Phonetics, phonology, phonotactics and prosody*. STUDIA HUMANIORA OULUENSIA. University of Oulu.

Traunmüller, Hartmut. 1999. Coarticulatory effects of consonants on vowels and their reflection in perception. In *Proceedings from the XIIth Swedish Phonetics Conference,* 141-144.

Studying vocal intensity information using machine learning – Database and preliminary experiments

Manila Kodali[1], Sudarsana Reddy Kadiri[1], Laura Laaksonen[2], and Paavo Alku[1]

1: Department of Signal Processing and Acoustics, Aalto University, Finland.

2: Tampere Wireless Headset Audio Lab, Finland Research Center, Huawei Technologies Oy (Finland) Co., Ltd., Tampere, Finland. paavo.alku@aalto.fi

In everyday speech communication, speakers regulate vocal intensity on many occasions, for example, to emphasise something or to be heard over a long distance or to signal vocal emotions such as anger or sadness. Vocal intensity is typically quantified using Sound Pressure Level (SPL), which can be measured easily by recording a standard calibration signal with speech and by comparing the RMS (root mean square) of the recorded speech signal with that of the calibration tone. Unfortunately, speech recordings are today mostly conducted without using a constant mouth-to-microphone distance and without recording the SPL calibration signal. Therefore, the original vocal intensity information (e.g. SPL) of speech is not available in most current speech databases and speech signals are typically treated using arbitrary amplitude scales (e.g. by scaling the maximum amplitude value of the sound waveform to be 1.0). However, despite being presented on arbitrary amplitude scales, the speech waveform contains acoustic cues of vocal intensity categories used by the speaker. In the current study, we study Machine Learning (ML) -based methods to automatically classify vocal intensity category when speech is expressed using an arbitrary amplitude scale. Based on a new speech database, four vocal intensity categories (soft, normal, loud, and very loud) are studied. Support Vector Machine (SVM) and Convolutional Neural Network (CNN) were used to develop different automatic ML-based classification systems using Mel-Frequency Cepstral Coefficients (MFCCs) and Melspectra as features. The SVM classifier with MFCCs showed the best classification accuracy of about 62%.

Puheen sujuvuus suomessa toisena kielenä

Liisa Koivusalo[1], Heini Kallio[2], Minnaleena Toivola[1]

[1]Helsingin yliopisto

[2]Jyväskylän yliopisto

Toisen kielen (L2) puheen tutkimuksissa sujuvuutta tutkitaan usein puheesta mitattavilla temporaalisilla piirteillä, joita ovat esimerkiksi puheen nopeus, tauot, korjaukset ja toistot [1]. Nopea, vähän epäröintiä ja keskeytyksiä sisältävä puhe mielletään usein sujuvaksi, ja toisen kielen oppimisen alkuvaiheessa puhe on epäsujuvampaa [2]. Koivusalon maisterintutkielmassa tutkittiin lukiolaisten L2-suomen foneettista sujuvuutta puheesta mitattavien foneettisten sujuvuuspiirteiden sekä sujuvuus- ja taitotasoarvioiden avulla [3]. Tutkimuksessa on hyödynnetty puhe- ja arviointiaineistoa, joka on kerätty DigiTala-tutkimusprojektissa [4]. Analysoitu aineisto sisälsi 53 spontaania puhenäytettä lukiolaisilta, jotka puhuvat suomea toisena kielenä. Jokaisen puhenäytteen sujuvuus ja yleinen taitotaso oli arvioitu. Puhenäytteisiin annotoitiin hiljaiset ja täytetyt tauot, korjaukset ja toistot sekä yksittäiset sanat. Annotoitujen intervallien kestoista laskettiin useita foneettisia sujuvuuspiirteitä jokaiselle puhenäytteelle.

Foneettisten sujuvuuspiirteiden vaikutusta ihmisarvioihin tutkittiin lineaaristen regressiomallien avulla. Puhenopeus, artikulaationopeus, pitkät ja lyhyet hiljaiset tauot ja yhtenäisten puhejaksojen väliset keskeytykset ennustivat sujuvuus- ja taitotasoarvioita parhaiten. Puhenopeus oli paras yksittäinen sujuvuusarvioiden ennustaja, mutta sujuvuus- ja taitotasoarvioiden vaihtelua selitti myös erilaisten sujuvuuspiirteiden yhdistelmät. Tulokset ovat samankaltaisia kuin aiemmissa sujuvuustutkimuksissa [2,5,6,7]. Tutkielman tulokset lisäävät tietoa L2-suomen foneettisesta sujuvuudesta, mitä on tutkittu aiemmin suhteellisen vähän. Tuloksia voidaan hyödyntää esimerkiksi L2-suomen opetuksessa tai puheen automaattisten arviointityökalujen kehittämisessä.

[1] Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

[2] Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164.

[3] Koivusalo, L. (2022). Phonetic fluency in Finnish as a second language: Acoustic analysis of high school students' spontaneous speech. Helsingin yliopisto. Maisterintutkielma. http://urn.fi/URN:NBN:fi:hulib202206152522

[4] Kautonen, M. & von Zansen, A. (2020). DigiTala research project: Automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta, 11*(4).

[5] Kallio, H., Suviranta, R., Kuronen, M. & von Zansen, A. (2022). Creaky voice and utterance fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish. *Proceedings of Speech Prosody 2022*.

[6] Kallio, H., Šimko, J., Huhta, A., Karhila, R., Vainio, M., Lindroos, E., Hildén, R. & Kurimo, M. (2017). Towards the phonetic basis of spoken second language assessment: temporal features as indicators of perceived proficiency level. *AFinLA-e: Soveltavan kielitieteen tutkimuksia,* (10), 193–213. https://doi.org/10.30660/afinla.73137

[7] Préfontaine, Y., Kormos, J. & Johnson, D. E. (2016). How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing, 33*(1), 53–73. https://doi.org/10.1177/0265532215579530

Puheen prosodisten piirteiden muutokset varhaisessa Alzheimerin taudissa

Marianne Kosin, Sonja Hakkarainen, Jessica Lamminsivu, Tiina Ihalainen, Nelly Penttilä
Logopedia, yhteiskuntatieteiden tiedekunta, Tampereen yliopisto

Johdanto: Alzheimerin tautiin liittyvät kognitiiviset ja kielelliset oireet voivat heijastua puheen prosodisiin piirteisiin: intonaatioon, rytmikkyyteen ja painotuksiin [1]. Akustinen puhesignaalin automaattinen analysointi, muun muassa edellä mainittujen prosodisten piirteiden osalta, onkin osoittautunut sensitiiviseksi menetelmäksi Alzheimerin taudin varhaisessa tunnistuksessa [1, 2]. Tutkimuksissa taudille ominaiseksi tunnistettuja dysprosodisia piirteitä ovat monotoninen [1] ja epärytmikäs puhe [3]. Akustisesti tarkasteltuna monotoninen puhe ilmenee perustaajuuden (F_0) madaltumisena ja vähäisempänä vaihteluna (F_0 *SD*) [1]. Epärytmikkään puheen taustalla on puolestaan hidastunut artikulaationopeus [1], tavujen keston pidentyminen [3] ja vierekkäisten tavujen keston tavallista suurempi vaihtelu (nPVI, *normalized pairwise variability index*). SPI (*novel syllabic prosody index*) on puheen painotuksen mittari [4]. SPI:n suurempi arvo viittaa painokkaasti tuotettuun tavuun, jolloin perustaajuus ja tavun kesto kasvavat sekä energia jakaantuu laajemmin, painottuen korkeammille taajuuksille (epb1kHz, *energy proportion below 1 kHz*). Siten Alzheimerin taudissa madaltunut perustaajuus [1] ja pidentynyt tavujen kesto [3] voivat aiheuttaa muutoksia myös puheen painotuksiin.

Menetelmät: Tutkimuksessa tarkastellaan puheen prosodisia muutoksia varhaista Alzheimerin tautia sairastavilla henkilöillä (AT-ryhmä; $n = 3$). AT-ryhmän tuloksia verrataan neurologisesti terveisiin puhujiin (NT-ryhmä; $n = 3$, ikä > 65 v). Aineisto muodostuu puhetallenteista ($N = 6$), joissa tutkittavat suorittavat sarjakuvakerrontatehtävän. Puhetallenteet litteroidaan Praatohjelmistolla ja analysoidaan prosodiaskriptillä [4]. Tutkimuksessa tarkasteltavia muuttujia ovat perustaajuus F_0 (*ka*, *vv, SD*), artikulaationopeus (tavua/s), vierekkäisten tavujen keston vaihtelu (nPVI) ja puheen painotukset (SPI).

Tulokset: Tutkimuksen tulokset julkaistaan Fonetiikan päivillä 2022.

[1] Martinez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., & Meilán, J. J. G. (2021). Ten years ofresearch on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: a systematic review article. *Frontiers in Psychology, 12*, 620251. doi: https://doi.org/10.3389/fpsyg.2021.620251

[2] Pulido, M. L. B., Hernández, J. B. A., Ballester, M. A. F., González, C. M. T., Mekyska, J., & Smékal, Z. (2020). Alzheimer's disease and automatic speech analysis: a review. *Expert Systems With Applications, 150*, 113213. doi: https://doi.org/10.1016/j.eswa.2020.113213

[3] Martínez-Sánchez, F., Meilán, J. J. G., Vera-Ferrandiz, J. A., Carro, J., Pujante-Valverde, I. M., Ivanova, O., & Carcavilla, N. (2017). Speech rhythm alterations in Spanish-speaking individuals with Alzheimer's disease. *Aging, Neuropsychology, and Cognition, 24*(4), 418–434. doi: https://doi.org/10.1080/13825585.2016.1220487

[4] Tavi, L. & Werner, S. (2020). A phonetic case study on prosodic variability in suicidal emergency calls. *The International Journal of Speech, Language and the Law, 27*(1), 59–74. doi: https://doi.org/10.1558/ijsll.39667

Parkinson-potilaiden poikkeavat puheen piirteet SMR-tehtävässä

Marianne Kosin, Leena Rantala, Nelly Penttilä
Logopedia, yhteiskuntatieteiden tiedekunta, Tampereen yliopisto

Johdanto: Oraalisen diadokokinesian SMR-tehtävä (*sequential motion rate*) on puheterapeuttien käyttämä maksimaalista puhemotorista suorituskykyä mittaava testi, jolla arvioidaan muun muassa Parkinsonin tautiin liittyvää hypokineettista dysartriaa [1]. Aiemmat tutkimukset ovat osoittaneet Parkinson-potilaiden SMR-tehtävässä mitatun maksimaalisen tavutoistonopeuden sekä hidastuneen [2] että nopeutuneen neurologisesti terveisiin puhujiin verrattuna [3]. Nämä erot mitatuissa maksimaalisissa tavutoistonopeuksissa ovat mahdollisesti yhteydessä kehon motoristen oireiden vaikeusasteeseen [2]. SMR-tutkimuksissa on raportoitu myös artikulaation epätarkkuutta Parkinson-potilailla.

Menetelmät: Tutkimuksessa tarkasteltiin idiopaattista Parkinsonin tautia (PT) sairastavien henkilöiden (*n* = 19, iän *ka* = 68,5 v) kykyä suoriutua oraalisen diadokokinesian SMRtehtävästä. Lisäksi tutkittiin, eroaako motorisilta oireilta lievää (PTa-ryhmä, *n* = 11) ja keskivaikeaa Parkinsonin tautia sairastavien (PTb-ryhmä, *n* = 8) tutkittavien suoriutumiskyky toisistaan. PT-, PTa- ja PTb-ryhmiltä saatuja tuloksia verrattiin neurologisesti terveisiin puhujiin (NT-ryhmä; *n* = 19, iän *ka* = 62,3 v). Tutkittavilta kerätyistä puhetallenteista (*N* = 38) mitattuja muuttujia olivat maksimaalinen tavutoistonopeus, sujumattomuustyypit ja niiden määrä. Tutkimusryhmien välisiä eroja analysoitiin tilastollisin menetelmin.

Tulokset: PT-ryhmän maksimaalinen tavutoistonopeus (*ka* = 7,0 tavua/s) ei eronnut NTryhmästä (*ka* = 7,1 tavua/s). Puolestaan PTb-ryhmän maksimaalinen tavutoistonopeus (*ka* = 6,4 tavua/s) oli merkitsevästi hitaampi kuin PTa-ryhmän (*ka* = 7,5 tavua/s, $p < ,05$). PTryhmässä ilmeni runsaasti virheellistä artikulaatiota (sijalukujen *ka* = 21,97) ja toistoa (sijalukujen *ka* = 21,5), jotka erottivat ryhmän NT-ryhmästä (sijalukujen *ka* = 17,03 / 17,5, $p < ,05$). PTa-ryhmän tuottamien sujumattomuuksien määrä (*md* = 2) oli merkitsevästi suurempi NT-ryhmään verrattuna (*md* = 0, $p < ,05$).

Johtopäätökset: Tutkimustulokset osoittivat, että SMR-tehtävässä Parkinsonin taudille tunnuksenomaisia puheen muutoksia ovat artikulaatiovirheet ja toistot, joita ilmenee runsaasti erityisesti motorisilta oireilta lievässä taudissa. Lisäksi tulokset antavat viitteitä siitä, että lievässä Parkinsonin taudissa maksimaalinen tavutoistonopeus on suurempi kuin keskivaikeassa taudissa. Näin ollen voidaan päätellä Parkinsonin tautiin liittyvien, SMRtehtävässä mitattujen, puheen poikkeavien piirteiden muuttavan muotoaan, kun kehon motoriset oireet etenevät lieväasteisista keskivaikeisiin; maksimaalinen tavutoistonopeus hidastuu ja sujumattomuuksien määrä vähenee.

[1] Karlsson, F., Schalling, E., Laakso, K., Johansson, K., & Hartelius, L. (2020). Assessment of speechimpairment in patients with Parkinson's disease from acoustic quantifications of oral diadochokinetic sequences. *The Journal of the Acoustical Society of America, 147*(2), 839–851. doi: https://doi.org/10.1121/10.0000581

[2] Novotný, M., Rusz, J., Čmejla, R., & Růžička, E. (2014). Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(9), 1366–1378. doi: https://doi.org/10.1109/TASLP.2014.2329734

[3] Skodda, S. (2015). Steadiness of syllable repetition in early motor stages of Parkinson's disease. *Biomedical Signal Processing and Control, 17*, 55–59. doi: https://doi.org/10.1016/j.bspc.2014.04.009

Palataaliapproksimantin koartikulaatio suomenkielisessä spontaanissa puheessa

Oskari Laakkonen & Michael O'Dell

oskari.laakkonen@tuni.fi michael.odell@helsinki.fi
Tampereen yliopisto ja Helsingin yliopisto

Tutkimus käsittelee soinnillisen palataalisen approksimantin koartikulaatiota. Tutkimuksessa käsitellyt rakenteet on rajattu V1-j-V2 sekä C-j-V -tyyppisiin jaksoihin. Tutkimus koostuu kahdesta osiosta: kvantitatiivisesta osiosta, jossa mitataan käsitellyt foneettiset sekvenssit, ja havaintokokeesta, jossa koehenkilöt arvioivat kuulemaansa. Aineistona on käytetty spontaania suomenkielistä puhetta ja aineiston käsittelyyn on käytetty Praat-ohjelmaa.

Luonnollisessa puheessa äänteet ovat alati vuorovaikutussuhteessa toisiinsa ja väistämättä vaikuttavat toisiinsa enemmän tai vähemmän. Tästä johtuen puheessa esiintyy toistuvasti koartikulaatiota eli vierekkäisten äänteiden mukautumista toisiinsa tai osittain päällekkäistä artikulaatiota.

Soinnillinen palataalinen approksimantti on suomen kielen foneemeista ainoa palataalinen konsonantti sekä soinnillisen labiodentaalisen approksimantin [ʋ]:n lisäksi ainoa puolivokaali. Approksimantit muistuttavat samalla artikulaatiopaikalla esiintyviä frikatiiveja, mutta niiden lausumisessa on vapaampi väylä ilmavirralle ja muodostuu huomattavasti vähemmän frikatiivihälyä. Approksimantit myös muistuttavat vokaaleja vastaavanlaisella artikulaatio paikalla ja huulien pyöreysasteella. Soinnillista palataalista approksimanttia on luonnehdittu suppean lavean etuvokaalin [i] ei-syllabiseksi variantiksi.

Tutkimuksen tavoitteena on perehtyä tarkemmin j-foneemin variaatioon ja arvioida sen vaikutusta ympäristössä esiintyvään vaihteluun.

Avainsanat: koartikulaatio, palataaliapproksimantti

Viitteet:

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. Language and speech, 36(2-3), 171-195.

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. Coarticulation: Theory, data and techniques, 730.

Maddieson, I. & Emmorey, K. (1985). Relationship between Semivowels and Vowels: Cross-Linguistic Investigations of Acoustic Difference and Coarticulation. Phonetica 42(4): 163-174.

SayEst - Android mobile app for practicing Estonian pronunciation

Pärtel Lippus, Katrin Leppik, Anton Malmi
University of Tartu

SayEst is an Android app for training the pronunciation and perception of Estonian vowels and consonants. Most currently available materials for learning Estonian as a second language concentrate on grammar and vocabulary, while very little attention is paid to pronunciation training. SayEst is a development of the prototype described in [1] and aims to fill this gap by offering a tool to work on the intricate details of the pronunciation of Estonian. The app can be used as supplementary material in a classroom setting by teachers, or learners can practice on their own at their own pace outside the classroom. SayEst is available with an English and Russian interface in the Play Store and free of charge.

The app is divided into two units for training vowels and consonants. There are blocks of lessons for training individual sounds or minimally contrastive sound pairs within each training unit. Each lesson includes a theoretical video and games based on discriminating the sounds in minimal pairs (e.g., ***u**ks*-***ü**ks*, *t**oo***-*t**öö***). There are four types of games:

1. Exposure – the task is to listen and repeat words and compare their pronunciation with a recording of native speakers' pronunciation.
2. Discrimination – the task is to listen to a word and choose which of the two written word forms is correct.
3. Pronunciation – the task is to read and pronounce two words. The pronunciation is evaluated by an automatic speech recognition tool [2].
4. Mixed Mode – this game includes the tasks from all previous games in random order.

In addition to the speech training, the app is used for crowdsourcing an Estonian L2 learner's dataset. Upon users' consent, their productions, progress, and choices in the app will be saved and used to analyze the acquisition of Estonian by non-native speakers. On the one hand, this will give us information on whether the app helps the learners improve their pronunciation. On the other hand, we will be able to collect a large dataset of Estonian learners' production and perception of Estonian vowels and consonants, which has been difficult to obtain in lab conditions [3, 4].

References

[1] K. Leppik and C. Tejedor-García, "Estoñol, a computer-assisted pronunciation training tool for Spanish L1 speakers to improve the pronunciation and perception of Estonian vowels," *Eesti ja soome-ugri keeleteaduse ajakiri. J. Est. Finno-Ugric Linguist.*, vol. 10, no. 1, pp. 89–104, Dec. 2019, http://dx.doi.org/10.12697/jeful.2019.10.1.05.

[2] T. Alumäe, O. Tilk, and Asadullah, "Advanced rich transcription system for Estonian speech," in *Human Language Technologies - the Baltic Perspective : Proceedings of the Eighth International Conference*, 2018, pp. 1–8, http://dx.doi.org/10.3233/978-1-61499-912-6-1.

[3] A. Malmi and P. Lippus, "Russian L1 speakers' palatalization in Estonian and the effect of phonetic speech training," *Est. Pap. Appl. Linguist.*, vol. 17, pp. 211–230, 2021, http://dx.doi.org/10.5128/ERYa17.12.

[4] K. Leppik, P. Lippus, and E. L. Asu, "The production of Estonian vowels in three quantity degrees by Spanish L1 speakers," in Proc. of the 19th ICPhS, Melbourne, Australia 2019, pp. 1154–1158. Available: https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2019/papers/ICPhS_1203.pdf

Lasten fonologisen prosessoinnin taitojen yhteys vieraan kielen äänteiden oppimiseen

[1]Lähteinen, Sonja, [2]Peltola, Kimmo U., [3]Alku, Paavo & [2]Peltola, Maija S.
[1]Logopedia, Psykologian ja logopedian laitos, Turun yliopisto
[2]Fonetiikka ja Learning, Age & Bilingualism –laboratorio (LAB-lab), Tietotekniikan laitos, Turun yliopisto
[3]Signaalin käsittelyn ja akustiikan laitos, Aalto yliopisto

Fonologisen prosessoinnin taidot huomioidaan usein esimerkiksi lukemisen vaikeuksia arvioidessa. On kuitenkin perusteltua epäillä, että vaikeudet äidinkielen kielellisissä perustaidoissa vaikuttaisivat myös vieraan kielen tuoton oppimiseen. Vieraan kielen äänteiden oppimista tarkastelevien teorioiden, mallien ja tutkimustulosten perusteella on selvää, että äidinkielen äännejärjestelmä vaikuttaa oleellisesti vieraan kielen äänteiden oppimiseen ja että äänteiden havaitseminen ja tuottaminen linkittyvät toisiinsa (esim. Flege et al. 1987, Best ja Strange 1992). Tässä tutkimuksessa etsittiin mahdollisia yhteyksiä äidinkielen fonologisen prosessoinnin vaikeuksien ja vieraan kielen äänteiden oppimisen välille. Päämääränä oli selvittää, peilautuvatko fonologisen prosessoinnin haasteet uusien äänteiden tuoton oppimiseen.

Tutkimukseen osallistui 17 koehenkilöä. Tutkittavat olivat iältään keskimäärin 8;3-vuoden ikäisiä suomenkielisiä lapsia, jotka osallistuivat tutkimukseen kahtena peräkkäisenä päivänä. Harjoittelu- ja testausjakso noudatti Peltola et al (2017) tutkimusta varten kehitettyä kuuntele ja toista -protokollaa, jossa koehenkilöt kuuntelevat ja tuottavat mallin mukaisesti semisynteettisiä (Alku et al. 1999) epäsanoja. Kohdesanassa esiintyy suomenkielisille vaikeasti opittava suppea pyöreä keskivokaali /ʉ/ epäsanassa /tʉ:ti/. Kontrollisanana toimii /ty:ti/, jossa esiintyy suomen kielen suppea pyöreä etuvokaali /y/. Mittauskertoja oli neljä ja niissä koehenkilöt äänsivät kuulemiaan sanoja kymmenen kertaa kumpaakin. Harjoitteissa sanat toistettiin 30 kertaa. Tuotoksista analysoitiin akustisesti (Praat) ensimmäisen tavun vokaalin F1, F2 ja F3 arvot ja mittaustulokset analysoitiin toistettujen mittausten ANOVAlla. Toisen tutkimuspäivän lopuksi koehenkilöiden fonologisen prosessoinnin taitoja tarkasteltiin Äänteiden prosessointi -osatestillä, joka sisältyy NEPSY II – lasten neuropsykologiseen tutkimukseen (Korkman ym., 2008). Tämän perusteella lapset jaettiin kahteen ryhmään (Ryhmä 1 = ei prosessoinnin vaikeuksia, Ryhmä 2 = prosessoinnissa vaikeuksia). Lopulta molemmista testeistä saadut tulokset yhdistettiin.

Tilastolliset analyysit paljastivat sanan päävaikutuksen ($F(1, 15) = 8.107$, $p = .012$) sekä sessio x formantti interaktion ($F(6, 10) = 7.804$, $p = .003$). Tärkeimpänä löydöksenä oli sessio x formantti x ryhmä interaktio sanalle /tʉ:ti/ ($F(6, 10) = 3.252$, $p = .049$), mikä osoitti ryhmien kehittyvän eri tavoin. Kun ryhmät analysoitiin erikseen Ryhmällä 1 ei paljastunut tilastollisia muutoksia harjoittelun suhteen, mutta sanat tuotettiin systemaattisesti erillään ($F(1, 9) = 7.791$, $p = .021$). Ryhmällä 2 havaittiin session päävaikutus ($F(3, 4) = 9.982$, $p = .025$) ja lisäksi kontrollisanan /ty:ti/ analysoinnissa paljastui formantin ja session interaktio ($F(6, 1) = 449.58$, $p = .036$). Yllättäen analyysi paljasti Ryhmällä 2 myös kontrollisanan session päävaikutuksen ($F(3, 4) = 57.902$, $p = .001$). Tämä ryhmä muovaa näin ollen oletusten vastaisesti kontrolliäännettä kohdeäänteen sijaan. Tutkimuksen päätulos onkin, että heikot äidinkielen fonologisen prosessoinnin taidot johtavat erilaiseen suoriutumisprofiiliin vieraan kielen äänteen oppimista mittaavassa tehtävässä. Lapset, jotka suoriutuivat keskimääräistä heikommin Äänteiden prosessointi -testistä, muokkasivat kuuntele ja toista -tehtävässä poikkeuksellisesti kontrolliäännettä /y/ kohdeäänteen /ʉ/ sijaan. Tutkimuksen kliininen merkittävyys kohdistuu lasten äidinkielen äänteellisten taitojen huomiointiin ja tukemiseen sekä kielenopetuksessa että puheterapiatyössä uusia äänteitä harjoitellessa.

LÄHTEET

Alku, P., Tiitinen, H. & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology, 110*, 1329–1333.

Best, C. T. & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305–330.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* 47–65.

Korkman, M., Kirk, U. & Kemp, S. L. (2008). *NEPSY-II – Lasten neuropsykologinen tutkimus.* Marit Korkman ja Psykologien Kustannus Oy.

Peltola, K. U., Alku, P. & Peltola, M. S. (2017). Non-native speech sound production changes even with passive listening training. *Linguistica Lettica, 25,* 158–172.

Estonian Elderly Speech: corpus collection and some prosodic characteristics

Einar Meister and Lya Meister
Tallinn University of Technology

Aging involves several degenerative changes in the human body including the parts of the speech production mechanism – the respiratory system, larynx, and the oral cavity (Linville, 2001). As a result, differences in several acoustic parameters between elderly and middle-aged voices have been reported for different languages, including changes in formants, fundamental frequencies (raising in males and lowering in females), voice quality changes (increased breathiness, jitter, and shimmer), and slower speaking rate (e.g., Albuquerque et al., 2019, Bóna, 2014, Eichhorn et al., 2017, Torre and Barlow, 2009).

These acoustic differences have been reported to affect the performance of the ASR systems resulting in higher WER in recognizing the speech of older (60–80 years old) adults compared to WER of younger (20–60 years old) adults (Werner et al. 2019). As speakers aged over 60 years are rarely present in the training corpora, the acoustic models trained with speech samples from middle-aged speakers are not suitable to recognize speech produced by elderly voices (Hämäläinen et al., 2014).

The Estonian elderly speech corpus targets to extend the existing speech corpora with speech samples from speakers aged over 60 years. The corpus will be used for training speech recognition systems and socio-phonetic studies addressing the changes in voice and speech characteristics of this age group. The number of speakers in the corpus is 200 balanced by gender and age groups and it contains spontaneous speech samples elicited during the interviews/conversations guided by an interviewer.

The talk will introduce the design and collection of the Estonian elderly speech corpus and reports the preliminary results of acoustic analysis of some prosodic characteristics – fundamental frequency (F0) and speech tempo – depending on age and gender.

References

Albuquerque, L., Oliveira, C., Teixeira, A., Sa-Couto, P., Figueiredo, D. (2019). Age-related changes in European Portuguese vowel acoustics. INTERSPEECH 2019, Graz, Austria, pp. 3965–3969.

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style, *J. Acoust. Soc. Am.* August 2014, 136(2), EL116-21. DOI: 10.1121/1.4885482.

Eichhorn, J. T., Kent, R. D., Austin, D., Vorperian H. K. (2017). Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women, *Journal of Voice*, 32(5): 644.e1–644.e9. DOI: 10.1016/j.jvoice.2017.08.003.

Hämäläinen, A., Avelar, J., Rodrigues, S., Dias, M. S., Kolesiński, A., Fegyó, T., Németh, G., Csobánka, P., Ting, K. L. H., Hewson, D. (2014). The EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14), May 26-31, Reykjavik, Iceland.

Linville, S. E. (2001). *Vocal aging*, Singular Thomson Learning.

Torre P, Barlow J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders,* 42, 324–333.

Werner, L., Huang, G., Pitts, B. J. (2019). Automated Speech Recognition Systems and Older Adults: A Literature Review and Synthesis. *Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting*, DOI: 10.1177/1071181319631121.

Tavujen ryhmittymiseen vaikuttavistä tekijöistä

Michael O'Dell, Tommi Nieminen, Joonas Vakkilainen
Helsingin Yliopisto ja Tampereen Yliopisto

Ihminen on taipuvainen tulkitsemaan toistuvia tapahtumia siten, että tapahtumat ryhmittyvät rytmisiksi yksiköiksi. Näin käy yleisesti havaitsemisessa ja tietenkin myös puhetta kuunnellessa. Vaikka jotkin akustiset piirteet saattavat universaalisti edistää tällaista ryhmittymistä, puheen tapauksessa myös kokemus omasta äidinkielestä vaikuttaa siihen, mitkä piirteet vaikuttavat eniten ja millä tavalla ne vaikuttavat (ks. Crowhurst 2018, ja siinä olevat lähteet). Crowhurst (2018) osoitti että toistuvien tavujen akustiset piirteet (mm. kesto, *F*0, narina) vaikuttavat englanninkielisten ja espanjankielisten kuulijoiden ryhmittämiseen osittain samalla tavalla, mutta osittain eri tavalla silloin, kun leksikaalisesta informaatiosta ei ole apua.

Suoritamme suomenkielisillä kuuluijoilla samankaltaisia havaitsemiskokeita, joita Crowhurst (2018) käytti tutkimuksissaan. Oletamme, että suomenkielisten rytminen ryhmittyminen osoittautuu myös osittain samankaltaiseksi, mutta myös osittain erilaiseksi verrattuna muiden kielten vastaaviin tuloksiin. Päivillä kerromme tarkemmin tuloksista ja pohdimme erojen mahdollisia syitä.

Viitteet

Crowhurst, Megan J. (2018). "The Influence of Varying Vowel Phonation and Duration on Rhythmic Grouping Biases among Spanish and English Speakers". *Journal of Phonetics* 66, s. 82–99.

Timing the starting and stopping of speech – An ultrasound study

Pertti Palo[1], *Steven M. Lulich*[1], *and Giulia Orlando*[2]

[1]Indiana University, [2]University of Helsinki

Background: The *inverse* correlation between acoustic reaction time and acoustic duration of the onset consonant in speech reaction time experiments has been known for some time (Fowler, 1979: Experiment 4). Replicating an acoustic experiment by Rastle et al. (2005) with articulatory methodology Palo (2019) was able to show that the inverse correlation of acoustic reaction time and the acoustic duration of the onset consonant (Onset Duration = OD), arises from the Articulatory onset to Acoustic onset Interval (AAI). Palo was further able to show that the AAI is also affected by articulation rate. These results are illustrated by the schematic in Fig. 1. In summary, articulatory reaction time appears to be a constant delay which is unaffected by phonetic content of the utterance while Acoustic Reaction time and the AAI – *are* affected by the phonetic parameters.

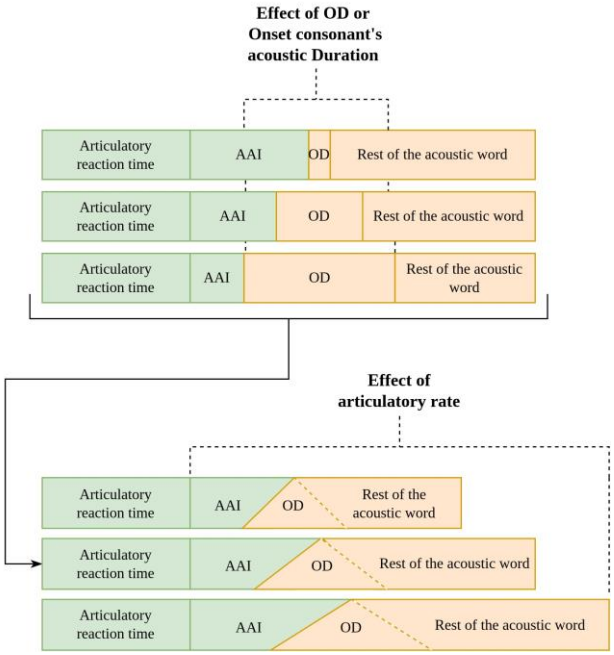

Figure 1: Timing model of speech initiation in delayed naming experiments based on results of Palo (2019).

This study: We report preliminary results from a new study which expands the number of participants (Palo (2019) only had 3 speakers) and also examines utterance end timing. The purpose of this study is to replicate and verify the earlier onset timing results and explore the timing of the final gesture of the utterance in relation to acoustic segment timing. The latter goal is motivated by the observation that speech gestures are

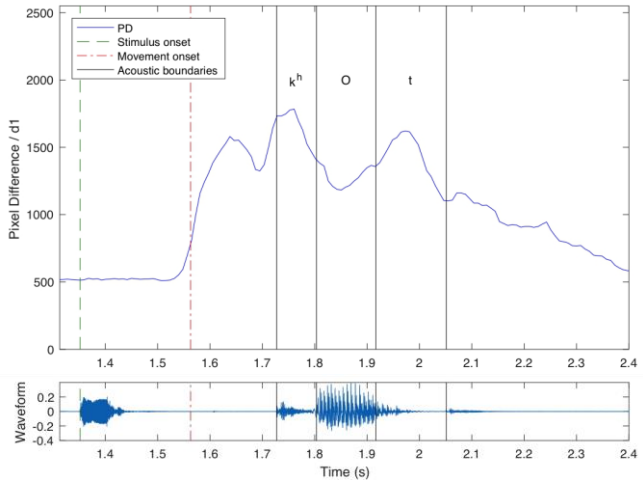

Figure 2: A pixel difference contour with the acoustic waveform: Participant says 'caught'.

evident in most of the earlier data when analysed with Pixel Difference (Figure 2). This offers a window to explore the potential systematicity in the timing of the final speech gesture.

Materials: The participants in this study are 18-30-year-old native speakers of American English. We record the speakers in a delayed naming experiment which combines simultaneous acquisitions of audio and UTI and is controlled with AAA.

Each trial begins with the target word displayed on a computer screen. The participant is instructed read the word internally while remaining at rest until they hear the go-signal (50 ms long 1 kHz beep), which is played after a random delay of 1.2-1.8 s from the beginning of the UTI recording. They are instructed to produce the target word as soon as possible after the beep.

Discussion: We expect to confirm the earlier findings on speech initiation and expect to have interesting data to discuss regarding the end of the utterance.

References

Fowler, C. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, 25(5):375 – 388.

Palo, P. (2019). *Measuring Pre-Speech Articulation*. PhD thesis, Queen Margaret University, Edinburgh.

Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083 – 1095.

Tanssi uhanalaisten kielten ja foneettisen maailman tulkkina (T&T&F) – Namibian kielten tallentaminen, elvyttäminen ja englannin opetuksen kehittäminen fonetiikan näkökulmasta

Maija S. Peltola

Fonetiikka ja Learning, Age & Bilingualism –laboratorio (LAB-lab), Tietotekniikan laitos, Turun yliopisto

Tässä esitelmässä päämääränä on kertoa kolmivuotisen Koneen Säätiön rahoittaman hankkeen tavoitteista ja toiminnoista.

Namibiassa puhutaan kymmeniä eri kieliä, joista osa on alkuperäisiä bantu- ja khoisankieliä, ja osa siirtomaavallan aikaisia germaanisia kieliä. Englanti on Namibian ainoa virallinen kieli ja sen lisäksi osa yleisimmistä kielistä on niin sanottuja koulukieliä, joita käytetään kouluopetuksessa ennen täysin englanninkieliseen opetukseen siirtymistä. T&T&F-hankkeen tavoitteissa yhdistyvät kieliperinnön tallentaminen, uhanalaisten kielten elvyttäminen ja pakollisen kielen oppimisen helpottaminen.

Kieliperinnön säilyttämiseksi ja namibialaisen foneettisen maailman taltioimiseksi nauhoitamme ja videoimme uhanalaisimpia kieliä ja arkistoimme ne siten, että ne ovat tulevaisuudessakin tutkimuskäytössä. Nauhoituksissa on osioita, joissa pyydämme informantteja kertomaan arkipäiväisistä asioista sekä suomalaisten vaikutuksista namibialaiseen kulttuuriin. Näiden tarinoiden sisällöistä saamme lyhyet englannin kieliset käännökset, joita kulttuurihistorioitsijat ja folkloristit käyttävät tutkimuksissaan. Samojen nauhoitusten avulla saamme lisätietoa kielten äännejärjestelmistä, mikä toimii pohjana englannin kielen ääänneharjoitusten laatimiselle. Näiden nauhoitusten pohjalta esiin nousseita Namibialaisten kielten erityisominaisuuksia nostetaan esiin myös nauhoittamalla loruja ja riimittelyjä, jotka ovat mukana rakentamassa Tanssiteatteri ERIn hankkeeseen sisältyvää produktiota. Produktiossa nostetaan foneettinen maailma esiin liikkeeseen, teoreettisesti ajatellen tavallaan palataan puheen havaitsemisen motorisen teorian ja viitotun kielen kautta kommunikaation juuriin. Tämä kielitietoisuutta tukeva tanssiproduktio esitetään myös streamina Namibian yliopistossa Windhoekissa.

Namibian kielten ja foneettisen maailman tutkimuksen ohella hankkeessa kerätään myös LAB-labin kansainvälisten tutkimusyhteistyölaboratorioiden mukaista ääntämisen oppimisen akustista materiaalia. Eri kielitaustaiset ja eri-ikäiset koehenkilöt osallistuvat testauksiin, joissa opetellaan kuuntele ja toista –menetelmällä ääntämään sekä/y/-/ɔ/ vokaalikontrastia että kesto-oppositioita epäsanakonteksteissa. Päämääränä on tutkia, miten erilaiset äidinkielten äännejärjestelmät vaikuttavat uusien foneettisten piirteiden omaksumiseen ja Namibian kielet tarjoavat hankkeelle laajan äännejärjestelmien kirjon.

Vaikuttaako vauvan neurokehityksellinen tila äidin puhetyyliin? – Hoivapuheen analyysi italiankielisellä vuorovaikutusaineistolla

Okko Räsänen[1], Manu Airaksinen[2], Fabrizia Festante[3], Viviana Marchi[3], Olena Chorna[3] & Andrea Guzzetta[3]

[1]Tampereen yliopisto, [2]BABA center, HUS, Helsinki, [3]IRCCS Fondazione Stella Maris, Pisa, Italia

Puhujat mukauttavat puhetyyliään riippuen puheen vastaanottajasta. Yksi tähän liittyvistä ilmiöistä on vanhemman vauvalle suuntaama puhe, ns. hoivapuhe, joka poikkeaa tyyliltään selvästi normaalista keskustelusta. Hoivapuheen aiempi tutkimus on pääasiassa keskittynyt hoivapuheen ominaisuuksiin eri kielissä ja eri ikäisillä lapsilla. Jonkun verran on tutkittu myös lapsen yksilökohtaisten tekijöiden, kuten kielenkehityksen tason (D'Odorico et al., 1999), heikkokuuloisuuden (Lovcevic et al., 2020) tai neurologisten kehityshäiriöiden (lähinnä autismikirjon häiriöt; Woolard et al., 2022) vaikutuksia vanhemman tuottamaan hoivapuheeseen. Ymmärryksemme hoivapuheen mukautumisesta lapsen yksilölliseen kehitystasoon ja tilannekohtaisiin tarpeisiin on kuitenkin edelleen vajavaista.

Tässä tutkimuksessa selvitettiin lapsen neurokehityksellisen tilan, eli ns. aivoterveyden, vaikutusta äitien tuottamaan hoivapuheeseen. Aineistona käytettiin italiankielisiä ääninauhoituksia, joissa äitejä pyydettiin vuorovaikuttamaan 4,5-kuukautisten vauvojensa kanssa. Nauhoitukset tehtiin joko laboratoriossa tai kotona käyttäen samaa nauhoitusasetelmaa. Aineisto koostuu yhteensä 28 vauva-äiti parista, joista 14 oli nauhoitusten aikaan tutkittavana lapsen epäillyn neurologisen ongelman takia (ns. riskiryhmä) ja 14 paria toimi terveenä verrokkiryhmänä (ns. kontrollit). Lisäksi riskiryhmä voitiin jakaa myöhemmin toteutuneen kliinisen arvion perusteella kahteen aliryhmään: 1) lapsiin, jotka osoittautuivat terveiksi tai joiden neurologiset ongelmat olivat lieviä (N = 7; esim. lievä CP-vamma ilman kognitiivisia ongelmia), sekä 2) lapsiin joilla esiintyi vakavia neurologisia ongelmia (N = 7; esim. vakava CP- ja kehitysvamma, vaikeat aistitoimintojen häiriöt, autismi).

Tutkimuksen ensimmäisessä vaiheessa äitien hoivapuhetta analysoitiin usealla akustisella mittarilla (intonaatio, lausekestot, puhenopeus yms.) ja mittauksia verrattiin kontrolli- ja riskiryhmän välillä. Ainoa merkittävä ero löytyi soinnillisen puheen määrästä puheen kokonaiskestoon suhteutettuna: soinnin määrä väheni systemaattisesti kontrolleista riskiryhmäläisiin, ja riskiryhmän sisällä lievistä vakaviin tapauksiin. Koska soinnillisuus on yhteydessä fonaatiotapaan, ja koska hoivapuhe on aiemmin yhdistetty lisääntyneeseen kuiskaamiseen (esim. Fernald & Simon, 1984) sekä vuotoisaan (ns. pehmeään) fonaatioon (Miyazawa et al., 2017), tutkimuksen toisessa vaiheessa keskityttiin fonaatiotavan tarkempaan analysiin. Kuiskauksen, vuotoisan ja modaalisen fonaation määrää verrattiin ryhmien välillä kahden annotaattorin arvioihin perustuen. Tuloksena havaittiin vuotoisan fonaation olevan merkittävästi yleisempää neurologisesti vakavasti sairaille lapsille suunnatussa puheessa kuin muissa vertailuryhmissä. Lisäksi kuiskaaminen oli yleisempää vakavassa neurologisessa ryhmässä kontrolleihin verrattuna.

Tulokset osoittavat, että lapsen neurokehitykselliseen tilaan liittyvät tekijät vaikuttavat äidin suulliseen vuorovaikutustapaan jo esikielellisillä lapsilla. On syytä olettaa, että neurologiset häiriöt vaikuttavat vauvojen omaan käyttäytymiseen vuorovaikutustilanteessa, ja siten myös heijastuvat äitien puhetapaan. Lisäksi vakavista neurologisista häiriöistä kärsivät lapset voivat olla herkkiä voimakkaille aistimuksille, joka pehmentää äitien käyttämää puhetapaa. Selittäviä tekijöitä ei kuitenkaan pystytty tässä tutkimuksessa analysoimaan tarkemmin.

Viitteet

D'Odorico, L., Salerni, N., Cassiba, R., & Jacob, V. (1999). Stability and change of maternal speech to Italian infants from 7 to 21 months of age: a longitudinal study of its influence on early stages of language acquisition. *First Language*, 19, 313–346.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104–113.

Lovcevic, I., Kalashnikova, M., & Burnham, D. (2020). Acoustic features of infant-directed speech to infants with hearing loss. *Journal of the Acoustical Society of America*, 148, 3399–3416.

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, 166, 84–93.

Woolard, A., Lane, A., Campbell, L., Whalen, O., Swaab, L., Karayanidis, F., Barker, D., Murphy, V., & Benders, T. (2022). Infant and child-directed speech used with infants and children at risk or diagnosed with autism spectrum disorder: a scoping review. *Review Journal of Autism and Developmental Disorders*, 9, 290–306.

aphaDIGITAL – Avatar-based digital speech therapy solution for aphasia patients: evaluation phase

Eugenia Rykova (UEF/TH Wildau), Mathias Walther (TH Wildau), Elisabeth Zeuner (MLU)

Aphasia, literally translated from Ancient Greek as „speechlessness", is an acquired language disorder due to focal brain injury. The most common reason of aphasia is a stroke, which affects about 270,000 people in Germany [1] and 25,000 people in Finland [2] every year. 30% of stroke patients suffer from aphasia. Recently, aphasia awareness probably grew because of the actor Bruce Willis' diagnosis [3].

Aphasia affects some or all language modalities, which makes communication difficult and decreases the quality of life. Speech and language therapy (SLT) improves functional communication. High intensity and duration of SLT bring certain benefits [4]. However, not all the patients with aphasia have access to sufficient SLT (e.g., due to lack of specialists or geographical remoteness). Research shows the efficiency of supplementing in-person therapy with independent usage of digital therapy solutions [5].

AphaDigital project [6] focuses on developing an SLT application for German-speaking patients with aphasia. In distinction from the existing German apps [7], in aphaDigital app an avatar-based SLT helper provides detailed feedback based on speech, text, and image processing mechanisms. Thus, speech and lip movements are analysed to provide feedback on pronunciation, and the helper gives not only an auditory but also an articulatory exemplary model to the user. Speech recognition with further text processing are used for higher-level feedback: semantic and grammatic. At the current stage of the project, the main goal is assessment of existing speech recognition solutions and semantic networks. Furthermore, different avatarhelpers are evaluated in a pilot study.

Several existing open-source speech recognition solutions (i.a., IMS-Speech [9]) are evaluated in two directions. First, we are looking for a solution that provides a phone/phoneme-level granularity, in other words, is to certain extent independent from existing vocabulary of the language. This is necessary to track phonetic and phonological mistakes of the speaker. On the other hand, language-dependent (sort of forcedaligned) speech recognition is useful for the analysis of semantic errors. The latter is carried out with the help of a semantic network: in a naming task, a word pronounced by the speaker is recognised and compared to the target word. In other words, their semantic relations and distance are analysed. This analysis allows a differentiated feedback upon an error. For example, when a patient with aphasia names apple a fruit, which is a hypernym of the word apple, she gets a prompt to be more specific. German wordnets (i.a., GermaNET [10]) are evaluated for semantic error analysis. The experiments are carried out with both artificially constructed examples and speech samples from AphasiaBank [11].

Four avatars are created for further evaluation: two women and two men of different ages. For each of the avatars, a set of phrases was recorded with a corresponding voice. Avatars are incorporated into a short (five trials) food-themed picture-naming exercise (choose the answer from four written options), designed in PsychoPy [8]. After assessing the experiment with 11 age-matched non-brain-damaged individuals, four patients with mild aphasia (57-62 y.o.) complete four exercises, with a different avatar-based helper each, and rate the helper according to 10 Likert-scales: general, naturalness, likeability, appearance, comprehensibility, voice, responsiveness, eye contact, support, and motivation.

The experiments described above are to be carried out in June-July 2022. The results will be reported in form of a poster.

[1] Deutscher Bundesverband der akademischen Sprachtherapeuten. (2016). Aphasie Informationen für Betroffene und Angehörige [Information on aphasia for affected individuals and relatives] [Brochure]. German Federal Association of Academic Speech Therapists.

[2] Reinikka-Uitto, P. (2019). Afaattisten puhujien kertova kieli. Kolmen puhujaryhmän sarjakuvakertomukset [The narrative language of aphasic speakers. Comics-based stories by three groups of speakers]. Master's Thesis. University of Tampere.

[3] Franklin, J. (2022). Understanding aphasia, the condition impacting Bruce Willis' acting career. Retrieved from https://www.npr.org/2022/03/31/1089806228/what-is-aphasia-explained?t=1650963692597

[4] Brady, M.C., Kelly, H., Godwin, J., and Enderby, P. (2016). Speech and language therapy for aphasia following stroke. The Cochrane Database of Systematic Reviews 2016, 6: 4–7.

[5] Braley, M., Pierce, J.S., Saxena, S., De Oliveira, E., Taraboanta, L. Anantha, V., Lakhan, S.E., and Kiran, S. (2021). A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia. Frontiers in Neurology, 12. [6] APHADIGITAL: Entwicklung einer digitalen, dezentralen sprachtherapeutischen Versorgung [Development of digital, decentralized speech therapy solutions]. Retrieved from https://inno-tdg.de/projekte/aphadigital/

[7] Griffel, J., Leinweber, J., Spelter, B., and Roddam, H. (2019) Patient-centred design of aphasia therapy apps: a scoping review. Aphasie und verwandte Gebiete | Aphasie et domaines associés, 2: 6-21.

[8] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods.

[9] Denisov, P., & Vu, N.T. (2019). IMS-speech: A speech to text tool. Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019: 170-177.

[10] Hamp, B., & Feldweg, H. (1997) GermaNet - a lexical-semantic net for German. Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications.

[11] MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. Aphasiology, 25, 1286-1307.

Vokaalien kesto- ja laatuerojen tuotto namibialaisilla puhujilla: alustavia huomioita

Antti Saloranta[1], Katja Haapanen[1], Kimmo U. Peltola[1], Henna Tamminen[1], Paavo Alku[2], and Maija S. Peltola[1]

1) Fonetiikka ja Learning, Age & Bilingualism –laboratorio, Turun yliopisto
2) Akustiikan ja signaalinkäsittelyn laitos, Aalto-yliopisto

Namibiassa puhutaan noin 30:tä kieltä, joista englanti, afrikaans ja saksa ovat indoeurooppalaisia, ja muut paikallisia bantu- ja khoisankieliä. Kymmenellä näistä paikallisista kielistä tarjotaan opetusta kolmannen luokan loppuun asti, mutta neljännestä luokasta eteenpäin opetuskielenä on englanti. Englannintaito on maassa kuitenkin heikko, ja monille oppilaille kielitaitovaatimus nousee hankaluudeksi (Norro, 2021). Turun yliopiston fonetiikan oppiaineen kolmivuotinen **Tanssi uhanalaisten kielten ja foneettisen maailman tulkkina (T&T&F)** -hanke pyrkii vastaamaan tähän ongelmaan muun muassa kehittämällä kielikohtaisia harjoitusmenetelmiä puhutun englannin harjoitteluun sekä nostamalla paikallisten kielten arvostusta.

Osana T&T&F -hanketta kerätään myös verrokkiaineistoa fonetiikan oppiaineen Learning, Age & Bilingualism -laboratorion aiemmille, eri kielitaustaisilla ihmisillä toteutetuille harjoitustutkimuksille. Niissä lyhyttä foneettista harjoittelua on käytetty muun muassa vokaalien laadun ja pituuden tuoton opetteluun (esim. Peltola et al., 2017; Saloranta et al., 2020). Tässä abstraktissa esitellään alustavia huomioita useiden Namibian paikallisten kielten puhujilta kerätystä harjoitusaineistosta sekä vokaalien laadun että pituuden osalta.

Osallistujia oli kuusi (kolme miestä, kolme naista), ja he puhuivat viittä eri äidinkieltä (khoekhoegowab, subia, oshiwambo, englanti ja setswana). Kaikki osallistujat puhuivat hyvää englantia, ja muita puhuttuja kieliä ryhmässä olivat yleisyysjärjestyksessä afrikaans, (otji)herero, saksa, oshiwambo sekä shona. Osallistujat olivat iältään 21-42-vuotiaita (ka 26,7), ja joko Namibian yliopiston opiskelijoita tai työntekijöitä. Osallistujia ohjeistettiin ja heidän kanssaan keskusteltiin englanniksi.

Kokeen laatuero-osuudessa käytettiin ärsykeparia /tʉ:ti/ - /ty:ti/, ja pituusosuudessa paria /tite/ - /ti:te/. Kummatkin parit olivat keskenään akustisesti identtisiä lukuun ottamatta kohdevokaalin laatua tai pituutta. Sekä pituus- että laatukokeet toteutettiin samalla protokollalla, jossa osallistujat kuulivat vuorotellen ärsykeparien sanoja, jotka heidän tuli toistaa niin hyvin kuin osaavat. Kokeessa oli kolme nauhoitettua osuutta, joissa toistoja oli 10+10, ja kaksi nauhoittamatonta harjoitteluosuutta, joissa toistoja oli 30+30. Kokeet noudattivat nauhoitus-harjoitus-nauhoitus-harjoitus-nauhoitus-rakennetta, eli osallistujat toistivat kumpaakin ärsykesanaa yhteensä 90 kertaa.

Kaikista nauhoitetuista tuotoista analysoidaan akustisesti vokaalien F1- ja F2 –formantit ja kestot. Pituusosuudessa verrataan lisäksi lyhyiden ja pitkien toistettujen vokaalien kestoja toisiinsa. Fonetiikan päivillä esitellään näiden analyysien alustavia tuloksia ja löydöksiä. Jatkossa, kun aineistoa on saatu kerättyä lisää, analyysien pohjalta tehdään myös tilastollisia alku- ja loppuvertailuja harjoittelun mahdollisesti aikaansaamien muutosten kartoittamiseksi.

Norro, S. (2021). Namibian Teachers' Beliefs about Medium of Instruction and Language Education Policy Implementation. Language Matters, 52(3), 45–71. https://doi.org/10.1080/10228195.2021.1951334

Peltola, K. U., Rautaoja, T., Alku, P., & Peltola, M. S. (2017). Adult Learners and a One-day Production Training – Small Changes but the Native Language Sound System Prevails. Journal of Language Teaching and Research, 8(1), 1–7.

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. International Journal of Psychophysiology, 147(November 2019), 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

A Hierarchical Predictive Processing Approach to Modelling Prosody

Juraj Šimko, Adaeze Adigwe, Antti Suni, Martti Vainio
University of Helsinki

Prosodic patterns, and linguistic structures in general, are hierarchical in nature, providing for efficient means for encoding information in temporally constrained situations where communicative events occur [1]. However, there are no theoretical frameworks that are capable of representing the full extent of linguistic behaviour in a cohesive way that could capture the paradigmatic and syntagmatic links between the organizational levels present in everyday speech.

We introduce a novel theoretical and modelling account of perception and production of prosodic patterns in speech communication, derived from the influential Predictive Processing theory of neural implementation of perception and action based on a hierarchical system of generative models producing progressively more detailed probabilistic predictions of future events [2-4]. The presented framework provides a conceptualization of the hierarchical organization of speech prosody as well as a principled way of unifying speech perception and production by postulating a single predictive processing hierarchy shared by both modalities.

In this mostly theoretical contribution, we will discuss possible implications of the theory for prosodic analysis of speech communication, including conversational setting. In addition, we outline a viable computational implementation in the form of a machine learning architecture, based on existing speech synthesis and recognition architectures [5,6], that can be used as a testbed for generating and evaluating predictions brought forth by the theory.

References

[1] A. Suni, J. Šimko, D. Aalto, and M. Vainio, "Hierarchical representation and estimation of prosody using continuous wavelet transform," *Computer Speech & Language*, vol. 45, pp. 123–136, 2017. [2] K. Friston, "Hierarchical models in the brain," PLoS computational biology, vol. 4, no. 11, 2008. [3] A.Clark,"Whatever next? Predictive brains,situated agents,and the future of cognitive science," *Behavioral and brain sciences*, vol. 36, no. 3, pp. 181–204, 2013. [4] —, *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press, 2015. [5] A. v. d. Oord, Y. Li, and O.Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint* arXiv:1807.03748, 2018.
[6] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint* arXiv:1609.03499, 2016.

Puheen muuntaminen ja puhujan varmennus

Lauri Tavi[a], *Tomi Kinnunen*[b], *Tuija Niemi*[a], *Rosa Hautamäki González*[c]

[a]Rikostekninen laboratorio, Keskusrikospoliisi, Vantaa
[b]Tietojenkäsittelytieteen laitos, Itä-Suomen yliopisto, Joensuu
[c]Electrical and Computer Engineering, National University of Singapore, Singapore

Tässä puheenvuorossa esittelemme puhujan varmennukseen (automatic speaker verification, ASV) liittyviä haasteita perustuen aiempiin tutkimuksiimme. Puhujan varmennuksella tarkoitetaan tunnistamattoman henkilön puhenäytteen vertaamista tunnistetun puhujan puhenäytteeseen. Tällöin tavoitteena on selvittää, onko puhenäytteissä sama tunnistettu puhuja. Vertailu perustuu ääntöväylän muodon ja puhetapojen eroihin puhujien välillä. Vaikka 2010luvulla alkaneella syväoppimisen aikakaudella automaattisen puhujan varmennuksen suorituskyky on parantunut merkittävästi, puheen tarkoituksellinen muuntelu heikentää myös uusinta teknologiaa hyödyntävien x-vektori-pohjaisten ASV-järjestelmien tuloksia (Tavi et al. 2022). Puheen tarkoituksellinen muuntelu voidaan toteuttaa niin koneellisesti (esim. formantteja manipuloimalla) kuin luonnollisesti puhetapaa vaihtamalla (Tavi et al. 2022; Tavi et al. 2021; González Hautamäki et al. 2019). Tällaisen tarkoituksellisen muuntelun tavoitteena voi olla joko tulla tunnistetuksi toiseksi henkilöksi tai *olla tunnistumatta* samaksi henkilöksi. Biometrisen turvallisuuden yhteydessä ensimmäinen tunnetaan huijaushyökkäyksenä (engl. *spoofing attack*). Puheenvuorossa käsittelemme myös puhujan de-identifikaatiota ja yksityisyyden suojaa. Lisäksi tuomme lyhyesti esille edellä mainittuihin aiheisiin liittyviä haasteita rikostutkinnan näkökulmasta.

Viitteet

Tavi, L., Kinnunen, T., & Hautamäki González, R. (2022). Improving speaker de-identification with functional data analysis of f0 trajectories. *Speech Communication*, 140, 1—10.

Tavi, L., Kinnunen, T., Meister, E., González-Hautamäki, R., & Malmi, A. (2021). Articulation During Voice Disguise: A Pilot Study. *In International Conference on Speech and Computer* (pp. 680—691). Springer, Cham.

González Hautamäki, R., Hautamäki, V., & Kinnunen, T. (2019). On Limits of Automatic Speaker Verification: Explaining Degraded Recognizer Score Through Acoustic Changes Resulting from Voice Disguise, *Journal of the Acoustic Society of America*, 146(1), 693—704.

Analysis of a Latent Prosody Space for Controlling Speaking Styles in Finnish End-to-End Speech Synthesis

Tuukka Törö, Helsingin yliopisto

In recent years, advances in deep learning have made it possible to develop neural speech synthesizers that not only generate near natural speech but also enable us to control its acoustic features. This means it is possible to synthesize expressive speech with different speaking styles that fit a given context. One way of achieving this control is by adding a reference encoder on the synthesizer that models a prosody related latent space (Skerry-Ryan et al., 2018).

The aim of this study was to analyze how the latent space of a reference encoder models diverse and realistic speaking styles, and, basing on previous research on acoustic correlates of speaking styles (Wagner et al, 2015), what correlation there is between the phonetic features of encoded utterances and their latent space representations. Another aim was to analyze how the synthesizer output could be controlled in terms of speaking styles. For the synthesizer output, two evaluations were conducted: an objective evaluation assessing acoustic features and a subjective evaluation assessing appropriateness of synthesized speech in regard to the uttered sentence. The model used in the study was a Tacotron 2 speech synthesizer with a reference encoder that was trained with read speech uttered in various styles by one female speaker.

The results showed that the reference encoder modeled stylistic differences well, but the styles were complex with major internal variation within the styles. Acoustic features were somewhat disentangled on the latent space, and there was a correlation between it and prosodic features of the utterances. The objective evaluation suggested that the synthesizer did not produce all of the acoustic features of the styles, but the subjective evaluation showed that it did enough to affect judgments of appropriateness: speech synthesized in an informal style was deemed more appropriate than formal speech for informal sentences and vice versa.

References

Skerry-Ryan, R.J., Battenberg, E., and Xiao, Y., Wang, Y., Stanton, D., Shor, J., Weiss, R.J., Clark, R. and Saurous, R.A. (2018) Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron.

Wagner, P., Trouvain, J., Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics 48: 1-12.*

Puheen emootiotunnistimen kehittäminen laajamittaiselle lapsikeskeiselle ääniaineistolle sairaalaympäristöstä

Einari Vaaras[1] *& Okko Räsänen*[1]
[1]Tietotekniikan yksikkö, Tampereen yliopisto

Puhe sisältää lingvistisen sisällön lisäksi myös valtavan määrän paralingvistista sisältöä, kuten esimerkiksi tietoa puhujan terveydentilasta, asenteista, emootioista ja persoonallisuudesta [1]. Puheen tunteiden tunnistuksessa (SER; speech emotion recognition) tavoitteena onkin tunnistaa automaattisesti puhujan emotionaalinen tila puhesignaalista [2]. Puheen emootioanalyysi on erityisen kiinnostavaa lasten ääniympäristötutkimuksessa, sillä varhaisen kehitysvaiheen kokemukset voivat vaikuttaa lasten myöhempään kognitiiviseen ja sosio-emotionaaliseen kehitykseen merkittävästi. Keskoset voivat joutua viettämään sairaalassa jopa useita kuukausia, jonka aikana heidän kuulemansa puheen laatu ja määrä voi olla hyvinkin erilainen tavalliseen kotiympäristöön verrattuna. Keskosvauvoilla onkin kohonnut riski epätavalliseen kielenkehitykseen [3] tai emotionaalisiin ongelmiin kuten masennukseen [4]. Keskosvauvojen kuuleman puheen emootiosisällön vaikutusta vauvojen kehitykseen ei kuitenkaan ole vielä tutkittu. Tähän tarkoitukseen Turun yliopistollisen keskussairaalan vastasyntyneiden teho-osastolla on kerätty satoja tunteja lapsikeskeisiä ääninauhoitteita osana APPLE-tutkimusta [5]. Tämän aineiston massiivisen koon vuoksi tarvitaan olennaisesti automaattinen emootiotunnistin nauhoitteiden emotionaalisen sisällön analysointiin. Koska vastaavalle aineistolle soveltuvaa emootiotunnistinta ei ollut valmiina, lähdettiin tässä työssä kehittämään puheääneen pohjautuvaa emootiotunnistinta tutkimalla vaihtoehtoisia tapoja luoda tällainen tunnistin mahdollisimman pienellä vaivalla.

Tyypillisesti vastaavanlaista ongelmaa voitaisiin lähestyä ohjatun koneoppimisen menetelmin, mikäli riittävä määrä annotoitua opetusdataa on saatavilla. Tutkimuksen pääaineiston kokonaan annotoiminen manuaalisesti olisi kuitenkin aivan liian aikaavievää ja kallista aineiston suuren määrän vuoksi, ja täten tutkimuksessa vertailtiinkin vaihtoehtoisia koneoppimisen metodeja emootiotunnistimen kehityksessä: mallien yleistyskykyä korpuksesta toiseen (ns. ristiinopetus), *k*medoids -klusterointialgoritmiin perustuvaa aktiivista oppimista (AL, Active Learning) sekä Wasserstein-generatiiviseen kilpailevaan verkostoon perustuvaa tunnistusmallin mukauttamista (DA, Domain Adaptation). Näistä metodeista ristiinopetus sekä DA hyödyntävät useaa korpusta siten että metodin opetuskorpus on eri kuin käyttökorpus, kun taas AL:ssä opetus- ja käyttökorpus ovat samat. Menetelmiä kehitettiin ja verrattiin ensin simulaatioiden avulla hyödyntäen jo olemassa olevia emootiopuhekorpuksia, jotta saataisiin selville mikä olisi parhain lähestymistapa kehittää emootiotunnistin annotoimattomalle korpukselle. Tämän jälkeen osa teho-osastonauhoitteista annotoitiin ja näitä annotoituja nauhoitteita käytettiin arvioimaan simulaatioiden löydösten toimivuutta käytännössä.

Tulosten perusteella AL-metodi oli vähiten riippuva käytettävästä opetus- ja käyttökorpuksesta sekä mallissa käytetyistä akustisista piirteistä, ja oli täten myös testatuista metodeista johdonmukaisin. Toisaalta DA-metodi tuotti parhaimmat tulokset, kun annotoitua dataa ei ollut lainkaan saatavilla. Huomasimme myös, että kehittämämme emootiotunnistin saavutti aiempaa kirjallisuutta vastaavan luokittelutarkkuuden jo varsin maltillisella määrällä ihmisen tekemää annotaatiotyötä.

Viitteet

[1] A. Batliner and B. Schuller, Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. New York: John Wiley & Sons, Incorporated, 2013.

[2] A. Batliner, B. Schuller, D. Seppi, S. Steidl, L. Devillers, L. Vidrascu, T. Vogt, V. Aharonson, and N. Amir, "The Automatic Recognition of Emotions in Speech," in Emotion-Oriented Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 71–99.

[3] A. Nyman, T. Korhonen, P. Munck, R. Parkkola, L. Lehtonen, L. Haataja, and PIPARI Study Group, "Factors affecting the cognitive profile of 11-year-old children born very preterm," Pediatric Research, vol. 82, no. 2, pp. 324–332, 2017.

[4] S. Upadhyaya, A. Sourander, T. Luntamo, H.-M. Matinolli, R. Chudal, S. Hinkka-Yli-Salom¨aki, S. Filatova, K. Cheslack-Postava, M. Sucksdorff, M. Gissler, A. Brown, and L. Lehtonen, "Preterm Birth Is Associated With Depression From Childhood to Early Adulthood," Journal of the American Academy of Child and Adolescent Psychiatry, vol. 60, no. 9, pp. 1127–1136, 2021.

[5] E. Ståhlberg-Forsen, A. Aija, B. Kaasik, R. Latva, S. Ahlqvist-Björkroth, L. Toome, L. Lehtonen, and S. Stolt, "The validity of the Language Environment Analysis system in two neonatal intensive care units," Acta Paediatrica, 2021.

Prosodia perättömissä hätäpuheluissa

Anne Väisänen, Itä-Suomen yliopisto

Valehtelu on universaali ilmiö ja onkin arvioitu, että ihminen valehtelee keskimäärin 1–2 kertaa päivässä erilaisissa arkisissa sosiaalisissa tilanteissa (DePaulo et al. 1996; Vrij 2008, 11). Valheet voivat pienimmillään olla niin sanottuja hyvää tarkoittavia ”valkoisia valheita”, joilla voidaan pyrkiä esimerkiksi hienotunteisuuteen toista henkilöä kohtaan. Toisessa ääripäässä valheella voidaan myös tarkoituksellisesti pyrkiä ajamaan omaa etua ja aiheuttamaan vahinkoa muille. (Gneezy 2005.)

Hätäkeskuslaitos vastaanottaa vuosittain miljoonia hätäpuheluita, joista sadat tuhannet ovat virheellisiä tai ilkivaltaisesti tehtyjä. Erityisen haitallisia ovat kuitenkin sellaiset puhelut, joissa apua pyydetään lähettämään (ja saadaan) paikalle tekaistulla verukkeella. Tällaiset perättömät ilmoitukset ovat ongelma, sillä ne sitouttavat ja kuormittavat viranomaisia ja pelastustoimea tarpeettomasti, minkä vuoksi avunsaanti saattaa viivästyä toisaalla. Koska hätäkeskuspäivystäjällä ei ole visuaalista yhteyttä hätäilmoituksen tehneeseen ilmoittajaan tai onnettomuuspaikkaan, on hänen arvioitava avun tarve ilmoittajan kertomuksen pohjalta. Samalla päivystäjän on tehtävä arvio siitä, miten totuudenmukainen ilmoitus on.

Mielikuva stereotyyppisestä valehtelijasta on suurimmalta osin visuaalinen ja tutkimusten mukaan ihmiset kiinnittävät huomiota enimmäkseen nonverbaaleihin merkkeihin arvioidessaan toisten uskottavuutta (Bogaard & Meijer 2020). Visuaalisista vihjeistä, kuten eleistä tai ilmeistä, ei kuitenkaan ole hyötyä hätäkeskuksissa, joissa hätäkeskuspäivystäjien on tehtävä arvio ilmoittajan rehellisyydestä kuulonvaraisesti. Tutkimukselle on tarve, sillä aiemmat kansainväliset tutkimukset ovat keskittyneet pääosin englanninkieliseen puheaineistoon (Spence et al. 2012) ja esimerkiksi DePaulo et al. (2003) ovat nostaneet esille autenttisen, laboratorio-olosuhteiden ulkopuolelta kerätyn puheaineiston merkityksen valehteluun liittyvissä tutkimuksissa.

Väitöskirjatutkimuksen keskeisenä tavoitteena on tarkastella valehdellun puheen prosodisia ominaisuuksia suomenkielisestä puheaineistosta ja sitä, millaisia tyypillisiä piirteitä perättömiin hätäilmoituksiin mahdollisesti liittyy. Tutkimuksessa hyödynnetään Hätäkeskuslaitokselta saatuja autenttisia hätäilmoituksia, jotka on todettu perättömiksi poliisin rekisteröimien rikosilmoitusten perusteella. Aineistoa tarkastellaan akustisin puheentutkimusmenetelmin sekä myöhemmin järjestettävällä kuuntelukokeella. Väitöskirjatutkimuksessa on jo aiemmin muun muassa selvitetty suomen kieltä äidinkielenään puhuvien viranomaisten ja maallikoiden uskomuksia valheen paljastavista verbaalisista ja nonverbaalisista merkeistä. Tulevassa posteriesitelmässä on tarkoitus esitellä väitöskirjatutkimuksen eri vaiheita sekä jo saatuja alustavia tuloksia.

Lähteet:

Bogaard G & Meijer EH. Self-Reported Beliefs About Verbal Cues Correlate with Deception-Detection Performance. Applied cognitive psychology 2018;32:129-137.

DePaulo BM, Kashy DA, Kirkendol SE, Wyer MM, Epstein JA. Lying in everyday life. Journal of Personality and Social Psychology 1996;70(5):979–995.

DePaulo BM, Lindsay JJ, Malone BE, Muhlenbruck L, Charlton K, Cooper H. Cues to deception. *Psychological Bulletin* 2003;129(1):74–118.

Gneezy U. Deception: The Role of Consequences. The American Economic Review 2005:95(1);384–394.

Spence K, Villar G, Arciuli J. Markers of deception in Italian speech. *Frontiers in Psychology* 2012;3:453.

Vrij A. Detecting Lies and Deceit : Pitfalls and Opportunities. (2nd ed). Chichester, England: John Wiley & Sons, Inc., 2008.

Puheenkierratystutkimus suomen rytmistä

Joonas Vakkilainen
Tampereen yliopisto

Puheenkierrätys eli *speech cycling* on puheen toistoon perustuva menetelmä rytmin tutkimiseksi (Cummins & Port, 1998; Tajima & Port, 2003). Puheenkierrätyksessä puhuja synkronoi puheensa ulkoiseen ärsykkeeseen. Tämän tarkoituksena on saada puhe säännönmukaiseen muottiin, jotta voidaan luoda ympäristö, jossa rytmiin vaikuttavat tekijät nousevat esiin, kun niiden säännönmukainen hierarkkisuus ei painu esim. tempovaihtelun ja epäsujuvuuksien alle. Puheenkierrätys toisaalta haastaa perinteistä rytmitypologiaa ja toisaalta täydentää rytmitypologian menetelmiä.

Tamä tutkimus on puheenkierrätyskoe suomen rytmistä. Tarkoituksena oli selvittää, mitä rytmityyppiä suomi eniten edustaa ja millainen rooli moralla suomessa on. Kehyslauseessa olevaa koesanaa varioitiin kvantiteettihahmoltaan ja tavuluvultaan. Ajoitusta tarkkailtiin koesanan suhteellisen osuuden kannalta sekakoesanan ja sitä seuraavan sanan suhteellisen alkuajankohdan kannalta suhteessa sykliin. Koehenkilöt osoittivat erilaisia ajoitusstrategioita, joista osassa moralla oli enemmän merkitystä ja osassa tavulla. Keskimäärin moraluku vaikutti ajoitukseen, mutta tavu vaikutti enemmän. Lisäksi saman moraluvun ja segmenttien mutta eri kvantiteettihahmon omaavilla koesanoilla oli keskenään erilainen ajoitus, mitä mikään rytmitypologinen teoria ei esitä.

Puheenkierrätystutkimuksen tulosten perusteella suomessa voidaan katsoa olevan piirteitä sekä tavu- että mora-ajoituksesta. Koe ei nostanut esiin puhtaasti mitään yhtä¨ rytmityyppiä, mikä on linjassa muista kielistä tehtyjen puheenkierrätystutkimusten kanssa.

Viitteet

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171.

Tajima, K., & Port, R. F. (2003). Speech rhythm in English and Japanese. *Phonetic interpretation: Papers in laboratory phonology VI*, 317–334.

Foneettisen ohjeistuksen vaikutus vieraan kielen äänteen omaksumiseen

[1]Eveliina Viitala, [2]Kimmo U. Peltola, [3]Paavo Alku, [2]Maija S. Peltola

[1]Logopedia, Psykologian ja logopedian laitos, Turun yliopisto

[2]Fonetiikka ja Learning, Age & Bilingualism –laboratorio (LAB-lab), Tietotekniikan laitos, Turun yliopisto

[3]Signaalin käsittelyn ja akustiikan laitos, Aalto yliopisto

Vieraan kielen ääntämisen oppiminen on tärkeä osa kielitaidon rakentumista. Puheterapiassa puolestaan fokus on lähinnä siinä, miten artikulaatio muovautuu harjoittelun myötä. Aiempi tutkimus on osoittanut, että kuuntele ja toista –harjoittelu toimii tehokkaasti myös teoreettisesti hyvin vaikeiden äänteiden oppimisen tukena (esim. Tamminen et al. 2021, Peltola, K.U. et al. 2017) ja että eksplisiittisillä ääntämisohjeilla oppimista voidaan nopeuttaa (Saloranta et al. 2015, Kissling et al. 2015). Tämän tutkimuksen päätavoitteena oli tarkastella sitä, miten eksplisiittiset ääntämisohjeet vaikuttavat artikulaation oppimiseen audiovisuaalisen harjoitteen tukena. Lisäksi tutkimuksessa testattiin etäoppimisalustan toimivuutta sekä teknisestä että puheterapian ja vieraan kielen oppimisen näkökulmasta.

Tutkimukseen osallistui 21 vapaaehtoista nuorta aikuista ja heidät jaettiin kahteen ryhmään. Ryhmälle 1 (N=10) ärsykkeet esitettiin audiovisuaalisessa muodossa siten, että koehenkilöt sekä kuulivat ärsykkeet että näkivät videolta mallin huulion asennosta ääntämisen aikana. Ryhmälle 2 (N=11) ärsykkeet esitettiin niin ikään audiovisuaalisesti, minkä lisäksi ryhmä sai ohjeistuksen koskien äänteiden tuottoa. Koeasetelma seuraili aiempien tutkimusten mallia (Peltola K.U. et al 2017), jossa koehenkilöt osallistuivat neljään mittaukseen ja niiden välissä kolmeen harjoitusosioon. Mittauksissa koehenkilöt toistivat mallin mukaisesti epäsanojen /ty:ti/ (kontrollisana) ja /tʉ:ti/ (kohdesana) 10 kertaa, harjoitusosioissa sanat toistettiin 30 kertaa. Kontrolli- ja kohdesanojen akustinen analyysi kohdistui ensimmäisen tavun vokaalien F1- ja F2-arvoihin. Oletuksena oli, että harjaantuminen olisi nopeaa, sillä aiempiin tutkimuksiin verrattuna tarjolla oli nyt myös visuaalinen tuki. Lisäksi hypoteesina oli, että eksplisiittiset kohdesanan ääntämisohjeet nopeuttaisivat muutosta. Korona-pandemian vuoksi tutkimusta varten kehitettiin Sanakon alustalle uusi etäharjoittelualusta, jossa koehenkilöt pystyivät suorittamaan nauhoitukset kotioloissa.

Analyysi osoitti, että etänauhoituksien avulla kerättävä materiaali oli laadullisesti niin tasokasta, että siitä pystyttiin helposti erottelemaan tutkimuksen kannalta keskeiset akustiset parametrit. Toistettujen mittausten ANOVA paljasti selkeän sana pää vaikutuksen ($F(1,19)=60.550$, $p<0.001$) ja sana x formantti –interaktion ($F(1,19)=81.467$, $p<0.001$), mikä osoittaa, että aineistosta voitiin selkeästi erottaa molempien ryhmien tuottamat sanojen erilaiset akustisen ominaisuudet. Tarkempi analyysi paljasti, ettei Ryhmällä 1 tapahtunut muutoksia harjoittelun edetessä, mutta Ryhmän 2 kohdesanan F2-arvot muuttuivat sessioiden välillä ($F(3,8)=5.289$, $p=0.027$). Analyysi osoitti myös, että merkitsevä muutos /ʉ/-vokaalin F2-arvoissa tapahtui harjoitusosioiden 2 ja 3 välissä ($t(10)=-3.009$, $p=0.013$). Formanttiarvojen tarkastelu paljasti, että tämä muutos oli kohti suurempia F2-arvoja sen jälkeen, kun ohjeistuksessa oli mainittu, että kyseessä oli vokaalien /y/ ja /u/ välimuoto. Tämä muutos vei arvot kauemmas kohdesanan akustiikasta. Tulokset viittaavat siihen, että eksplisiittiset ääntämisohjeet ovat keskeinen tekijät ääntämisen oppimisessa ja siten niiden antamisessa on syytä olla huolellinen. Tässä tapauksessa kohdesanan artikulaatio lähti kehittymään pois mallista, kun ääntämisohjeet toivat esiin äidinkielen vokaalien kautta tutun ääntämismallin. Lisäksi tutkimus osoitti, että etäharjoittelulla voidaan saavuttaa vastaavat oppimistulokset kuin laboratorio-olosuhteissa.

LÄHTEET

Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research,* 19, 254–275.

Peltola, K. U., Alku, P. & Peltola, M. S. (2017). Non-native speech sound production changes even with passive listening training. *Linguistica Lettica, 25,* 158–172.

Saloranta, A., Tamminen, H., Alku, P. & Peltola M. S. (2015). Learning of a non-native vowel through instructed production training. *18th International Congress of Phonetic Sciences*.

Tamminen, H., Kujala, T., Näätänen, R. & Peltola, M. S. (2021). Aging and non-native speech perception: A phonetic training study. *Neuroscience Letters*, Volume 740, 1 January 2021.

Eksplisiittisen ääntämiskurssin vaikutus S2-puhujien suomen ääntämiseen

Päivi Virkkunen & Minnaleena Toivola

Helsingin yliopisto

Vieraan kielen ääntämisen oppiminen edellyttää lähes poikkeuksetta runsasta motorista harjoittelua. Kognitiivinen kielitieto voi auttaa harjoittelussa, mutta oppijalla ei useinkaan ole käsitystä siitä, miten puhetta tai vieraan kielen äänteitä ja prosodiaa tuotetaan. Motorinen harjottelu ilman ymmärrystä siitä, mitä ollaan tekemässä, voi olla tehotonta ja johtaa pahimmillaan väärien ääntämismallien oppimiseen. Vieraan kielen ääntämisen vaikeudet johtuvatkin yleisemmin kognitiivisen kuin motorisen kielitaidon puutteesta (Fraser, 2000).

S2-opiskelijat kuulevat jatkuvasti suomea opettajan puhumana ja ympäristössään, mutta ääntämistä tulee tutkimusten mukaan opettaa myös eksplisiittisesti, erityisesti siihen keskittyen (Derwing, Munro & Wiebe, 1998; Lintunen, 2014). Eksplisiittisistä ääntämiskursseista on saatu hyviä kokemuksia (ks.
esim. Couper, 2003; Zhang & Yuan, 2020), joten halusimme testata kurssin vaikuttavuutta myös suomen kielen kontekstissa.

Meneillään olevassa tutkimuksessa selvitämme eksplisiittisen ääntämiskurssin vaikutusta viidentoista noin B2-tasoisen suomenopiskelijan ääntämiseen. Syksyllä 2021 järjestetty seitsemän viikon kurssi koostui teorialuennoista, runsaista puheen analyyttisen kuuntelemisen ja tuottamisen harjoituksista sekä opettajan antamasta palautteesta. Kurssin osallistujien puhetta (viivästettyä toistoa, lukupuhetta ja spontaania kerrontaa) äänitettiin ennen ja jälkeen kurssin. Puheaineiston ja osallistujille tehtävän kyselyn perusteella selvitetään, millainen vaikutus kurssilla oli osallistujien ääntämiseen sekä heidän asenteisiinsa ääntämistä ja eksplisiittistä opetusta kohtaan.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect, 18(*3), s. 53-70.

Derwing, T.M., Munro, M. & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning 48(3)*, s. 393–410.

Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. ANTA Innovative Project. Canberra: DETYA.

Lintunen, P. (2014). Ääntämisen oppiminen ja opettaminen. Teoksessa P. Pietilä & P. Lintunen (toim.) *Kuinka kielta opitaan* (s. 165–187). Helsinki: Gaudeamus.

Zhang, R., & Yuan, Z. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition,42*(4), 905-918.

Improving the intelligibility of sung text: project introduction and some preliminary results

Allan Vurma[1], Jaan Ross[1], Marju Raju[1], Tuuri Dede[1], Einar Meister[2], and Lya Meister[2]

[1]Estonian Academy of Music and Theater

[2]Tallinn University of Technology

This year, a joint research project of the Estonian Academy of Music and Theater and Tallinn University of Technology titled "Improving the intelligibility of sung text: the problems and the scientific basis" funded by the Estonian Research Council, was launched. The project aims to create a scientific basis for the development of strategies to achieve a good balance between text intelligibility and the requirements of the music when singing in various acoustics and with the presence of accompaniment. The topic combines both practical and scientific issues with the need to bring more science-based methods to classical singing.

Listeners expect vocalists to sing with intelligible text, but singers also have to obey the constraints which are dictated by the music. Why are the lyrics of pop songs easier to understand than in the classical singing style? And why can some opera singers still sing in a way that the lyrics are intelligible but others can't? The fundamental difference between the vocal techniques of the Western classical singing style and the commercial styles may be one of the determining factors in the intelligibility of the text. Due to the constraints dictated by the music, techniques that improve speech intelligibility may not work in singing. The singers' understanding of how to ensure good lyric clarity is controversial, and investigations on the subject are scarce.

The research methods to be applied within the project include: (1) the acoustic analysis of the vocal performances, (2) perception tests of vocal stimuli with systematically modified phonetic and musical characteristics, and (3) qualitative methods – interviews with singers and singing teachers. The results will be applied to voice training and could help text writers and composers to reduce problems with text intelligibility.

In the presentation, we will introduce sung and read study material recorded by several opera singers and report some preliminary results on the variability of durations and intensities of vowels and consonants.

Asymmetrical Lombard Effect – Conversating in Loud and Quiet Environments Simultaneously

Alexandra Wikström, Juraj Šimko
University of Helsinki

Humans increase their vocal efforts in a noisy environment in a reflex-like manner. This phenomenon is called the Lombard effect (Lombard, 1911). The effect causes the speaker to produce Lombard speech, which has been researched for over a century from different standpoints. Research on the effect of asymmetrical noise conditions on speech production however has been lacking.

The goal of this experiment was to examine speech production in a conversational situation where simultaneously one of the interlocutors engaged in a conversation is subjected to noise and is thus producing Lombard speech, while the other interlocutor is communicating in silence without the direct effects of background noise, and to discover, whether there are differences in the acoustics or the intelligibility of speech in such an asymmetrical speech situation compared to a symmetrical situation where the noise environment of the interlocutors is the same. Two pairs of Finnish speakers (4 participants altogether, all female) were recorded doing sudoku-based tasks in three different background noise conditions: (1) in quiet, (2) with both interlocutors in noise (symmetrical), and (3) with only one of the interlocutors subjected to noise (asymmetrical). Intensity and f_0 were measured.

All participants increased their intensity level and f_0 from the quiet to the symmetrical condition, where both interlocutors produced Lombard speech, an expected outcome of the Lombard effect (Lu & Cooke, 2009). The participants who during the asymmetrical condition were in silence and communicated to the interlocutor who was in noise increased both their intensity and f_0 in the asymmetrical condition compared to the quiet condition. In addition, one of these participants increased both measures to nearly the levels that were measured from her Lombard speech in the symmetrical condition. The participants who were subjected to noise during the asymmetrical condition on average used lower intensity levels in the asymmetrical condition than in the symmetrical condition, even though they produced Lombard speech during both.

This experiment demonstrated that when the sound environments of two interlocutors are different, neither of the interlocutors produces speech that would be completely suitable for their respective environments but instead are indirectly affected by the sound environments of their conversational partners. In addition, it was shown that while communicativeness can increase the effects of the Lombard effect (Garnier et al., 2010), it can also decrease them.

References

Garnier, M., Henrich, N. & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research, 53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx, 37*(2), 101–119.

Lu, Y. & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America, 126*(3), 1495–1499. https://doi.org/10.1121/1.3179668

Prosodiset ja akustisesti määritellyt rajat vs. havaitut rajat spontaanin puheen aineistossa

Tiia Winther-Jensen
Helsingin yliopisto

Puhutun kielen segmentoinnin merkitystä perustellaan usein paitsi sen tarpeellisuudella foneettisen tutkimuksen välineenä (Machač & Skarnitzl 2009, 11) myös sillä, että se on puheen tuottamisen ja ymmärtämisen kannalta olennaista (Aho 2010, 11). Segmentointia voi lähestyä myös puhtaasti psykolingvistisestä näkökulmasta, jolloin lähtökohta on lingvisesti määriteltyjen segmenttirajojen sijaan kielenkäyttäjän havainto ja kokemus. Mutta miten tällaiset havaitut rajat suhteutuvat lingvististen kategorioiden mukaan määriteltyihin rajoihin? Esitelmässäni osoitan, mitkä prosodis-foneettiset piirteet osallistuvat rajojen havaitsemiseen suomenkielisessä spontaanissa puheessa.

SEGMENT-hankkeessa kerätty aineisto sisältää 51 äidinkielisen suomen puhujan (ei-lingvistin) spontaanissa puheessa havaitsemia rajoja. Aineisto kerättiin tätä varten kehitetyllä ChunkitApp-sovelluksella (Vetchinnikova ym. 2017). Koehenkilöt kuuntelivat sovelluksen avulla spontaania puhetta sisältäviä katkelmia ja seurasivat samalla tabletin ruudulta katkelman litteraattia, jossa ortografiset sanat oli erotettu toisistaan interaktiivisella tilde-merkillä (~). Painamalla merkkejä koehenkilöt merkitsivät litteraattiin rajoja valitsemiinsa kohtiin. Tehtävässä pyydettiin toimivaan vaistonvaraisesti eikä rajan käsitettä määritelty.

Esitelmässäni tarkastelen koehenkilöiden merkitsemiä rajoja suhteessa prosodisiin rajoihin. Keskeisenä vertailukohtana on Continuous Wavelet Transform -tekniikkaa hyödyntävän analyysityökalun (Suni ym. 2016) mittaamat arviot prosodisen rajan todennäköisyydestä, joita arvioin myös suhteessa manuaalisesti määriteltyihin intonaatiojaksojen rajoihin. Osoitan esityksessäni, millaisissa kohdissa CWT:n avulla määritellyt rajat eroavat koehenkilöiden havaitsemista rajoista sekä korvakuulolta määritellyistä intonaatiojaksojen rajoista.

Lähteet

Aho, Eija 2010. *Spontaanin puheen prosodinen jaksottelu.* Helsingin yliopisto.

Machač, Pavel & Skarnitzl, Radek 2009. *Principles of Phonetic Segmentation.* Epocha Publishing House, Praha.

Suni, A., Šimko, J. & Vainio, M. 2016. *Boundary detection using Continuous Wavelet Analysis.* Proceedings of Speech prosody 2016.

Vetchinnikova, S., Mauranen, A., Mikusova N. 2017. *ChunkitApp: Investigating the relevant units of online speech processing.* INTERSPEECH 2017: Show & Tell Contribution.