universität wien

Institut für Romanistik

Book of Abstracts

14. Tagung Phonetik und Phonologie im deutschsprachigen Raum

6.–7. September 2018, Wien, Österreich

AMBASSADE DE FRANCE
EN SLOVAQUIE

BOOK OF ABSTRACTS

Phonetik und Phonologie im deutschsprachigen Raum

14. P&P 2018

6. - 7. September 2018

Wien, Österreich

14. Tagung Phonetik und Phonologie im deutschsprachigen Raum

Publiziert von:
Institut für Romanistik
Universität Wien

Institut für Schallforschung
Österreichische Akademie der Wissenschaften

Tagungsinformationen:
http://pundp2018.univie.ac.at/home/

Organisationsteam
Univ.-Doz. Dr. Sylvia Moosmüller † (Institut für Schallforschung, ÖAW)
Univ.-Prof. Dr. habil. Elissa Pustka, Dipl.-Journ.Univ. M.A (Institut für Romanistik)

Barbara Tiefenbacher (Sekretariat, Arbeitsbereich Prof. Pustka)

Marc Chalier, M.A. (Institut für Romanistik)
Mag. Johanna Fanta-Jende (Institut für Germanistik)
Fabian Fleißner, BA MA (Institut für Germanistik)
Mag. Julia Forster, BA (Institut für Romanistik)
Luise Jansen (Institut für Romanistik)
Mag. Julia Kamerhuber, BA (Institut für Romanistik)
Nicola Klingler, M.A. (Institut für Schallforschung, ÖAW)
Univ.-Prof. Dr. Alexandra Lenz (Institut für Germanistik)
Mag. Hannah Leykum (Institut für Schallforschung, ÖAW)
Mag. Dr. Markus Pöchtrager (Institut für Sprachwissenschaft)
Mag. Thomas Skarits (Institut für Romanistik)

Redaktion des Book of Abstracts
Barbara Tiefenbacher

VORWORT

Die 14. P&P findet am 6. und 7. September 2018 an der Universität Wien und somit erstmalig in Österreich statt. Die P&P kommt damit in die Geburtsstadt der 1939 postum erschienenen *Grundzüge der Phonologie* von Nikolai Trubetzkoy (1890-1938) mit seinen berühmten Definitionen der *Phonetik* und *Phonologie*. In diesem Geist organisieren wir in Wien die P&P in Kooperation zwischen dem Institut für Schallforschung der Österreichischen Akademie der Wissenschaften, an dem die phonetische Forschung angesiedelt ist, und mehreren Instituten der philologisch-kulturwissenschaftlichen Fakultät der Universität Wien (Romanistik, Germanistik, Sprachwissenschaft), an denen Phonologie gelehrt und geforscht wird. Entsprechend führen auch die *Keynote*-Vorträge unserer *invited speakers* Chiara Celata (Pisa, Italien) und Jean-Pierre Chevrot (Grenoble, Frankreich) an der psycho- und soziolinguistischen Schnittstelle phonetische und phonologische Aspekte zusammen.

Seit ihrer Gründung 2004 in Potsdam hat sich die P&P als interdisziplinäres Diskussionsforum ein ausgezeichnetes Renommee erworben. Im Zentrum des Treffens steht der Austausch zwischen fortgeschrittenen Studierenden, DoktorandInnen, PostDocs und erfahrenen WissenschaftlerInnen. In diesem Jahr laden neben den zwei 60-minütigen *invited talks* und 15 weiteren 25-minütigen Vorträgen im Plenum (inklusive Diskussion) drei Poster-Sessions mit insgesamt 46 Postern zu intensivem Austausch ein. Ein Poster-Slam im Plenum, in dem jedes einzelne Poster in einer 1-minütigen Blitzpräsentation kurz vorgestellt wird, ermöglicht es den TeilnehmerInnen, sich einen Überblick über die Gesamtheit der Poster der P&P zu verschaffen.

Alle Beiträge sind im Rahmen eines anonymen *peer review*-Verfahrens durch jeweils zwei GutachterInnen mit Blick auf die Kriterien Passung, Innovativität/Relevanz, Stand der Forschung, Methodik und Klarheit beurteilt worden. Wir möchten daher herzlichen allen Mitgliedern unseres wissenschaftlichen Beirats danken, die diesen Qualitätssicherungsprozess ermöglicht haben! Die Beiträge können im Anschluss an die Konferenz in den Online-Tagungsakten auf Deutsch und Englisch publiziert werden.

Ein großes Dankeschön auch an unsere Sponsoren: der *Österreichischen HochschülerInnenschaft* (ÖH), dem Institut für Schallforschung, der deutschen, französischen und italienischen Botschaft sowie den Verlagen Facultas, Erich Schmidt, Narr, Franck & Timme und Langenscheidt. Ohne diese Drittmittel wäre die weiterhin kostenfreie Teilnahme an der P&P in Österreich finanziell nicht möglich gewesen.

Wir bedanken uns schließlich bei allen AutorInnen für ihre spannenden Einreichungen und freuen uns schon sehr auf die P&P14!

Elissa Pustka

Wien, im September 2018 *(im Namen des Organisationsteams)*

NACHRUF Sylvia Moosmüller (1954 – 2018)

Mit großer Betroffenheit geben wir Nachricht vom Tod unserer langjährigen Kollegin Doz. Dr. Sylvia Moosmüller. Als Leiterin der Forschungsgruppe für Akustische Phonetik war sie maßgeblich an der erfolgreichen Forschung unseres Instituts und dem Aufbau seiner internationalen Reputation beteiligt.

Sylvia Moosmüller studierte Anglistik, Romanistik sowie Allgemeine und Angewandte Sprachwissenschaften an der Universität Wien. 1984 erlangte sie mit der Dissertation "Soziale und psychosoziale Sprachvariation: eine quantitative und qualitative Untersuchung zum gegenwärtigen Wiener Deutsch" ihr Doktorat in Angewandter Sprachwissenschaft. Anschließend arbeitete sie in verschiedenen Projekten zum Wiener Deutsch vor allem in Zusammenarbeit mit Univ. Prof. Dr. Wolfgang Dressler. Zeitgleich forschte sie auch immer wieder im Bereich der feministischen Linguistik und gender studies. Ihre Arbeit am Institut für Schallforschung (damals noch Kommission für Schallforschung) nahm sie im Jahr 1992 unter dem damaligen Leiter Doz. Dr. Werner Deutsch im Rahmen eines Projektes zum Thema forensische Phonetik auf. Mit der Approbation ihrer Habilitation "Vowels in Standard Austrian German. An Acoustic-Phonetic and Phonological Analysis" erlangte sie 2008 die Venia legendi für Angewandte Sprachwissenschaft zuzüglich Phonetik und Phonologie. Neben ihrer Lehre an den Universitäten Wien und Graz unterrichtete sie auch an der Fachhochschule „Logopädie-Phoniatrie und Audiologie“.

Über die Jahre hinweg baute sie am Institut die Arbeitsgruppe für Akustische Phonetik auf, die unter ihrer Leitung zahlreiche national und international erfolgreiche und beachtete Projekte durchführte. In den Jahren 2008 bis 2015 übte sie zusätzlich noch die Funktion der stellvertretenden Direktorin des Instituts für Schallforschung aus.

Ihre Forschungsschwerpunkte lagen in den Bereichen der phonetischen und phonologischen Variation des Österreichischen Deutsch sowie in der akustisch-phonetischen Beschreibung der Vokale ausgewählter, bislang unzureichend beschriebener Sprachen. Daneben betrieb sie Forschung im Bereich der forensischen Phonetik, der Soziophonetik und klinischen Phonetik. Als Generalsekretärin der IAFPA (International Association for Forensic Phonetics and Acoustics) sowie als Mitglied der AG für forensische Sprach- und Audioanalyse des ENFSI war sie auch international in diesem Forschungsbereich sehr anerkannt.

Am 17. April 2018 verstarb sie nach schwerer Krankheit viel zu früh. Das Institut für Schallforschung, seine Mitarbeiterinnen und Mitarbeiter, trauern um eine hervorragende Wissenschaftlerin, die in ihrer feinen und zurückhaltenden Art eine sehr geschätzte und wertvolle Kollegin war. Unsere Gedanken sind bei ihrer Familie, der wir von Herzen unser Mitgefühl ausdrücken.

Konrad Antonicek, Christiane Herzog und Michael Pucher (Institut für Schallforschung, ÖAW)

WISSENSCHAFTLICHER BEIRAT

- Univ.-Prof. Dr. Daniel Büring (Universität Wien)
- Assoz. Prof. Mag. Dr. Dina El Zarka (Universität Graz)
- Dr. Matthias Fingerhuth, M.A. (Universität Wien)
- PD Mag. Dr. Manfred Glauninger (Universität Wien)
- Univ.-Prof. Dr. Matthias Heinz (Universität Salzburg)
- O.Univ.-Prof. Dr. Bernhard Hurch (Universität Graz)
- Mag. Dr. Nadja Kerschhofer-Puhalo (Universität Wien)
- Ao.Univ.-Prof. Mag. Dr. Hugo Kubarth (Universität Graz)
- Univ.-Prof. Dr. Alexandra Lenz (Universität Wien)
- Ass.-Prof. Mag. Dr. Hans Christian Luschützky (Universität Wien)
- Prof. em. Dr. Utz Maas (Universität Graz)
- Univ.-Prof. PhD Ineke Mennen (Universität Graz)
- Mag. Dr. Friedrich Neubarth (Universität Wien)
- Mag. Dr. Markus Pöchtrager (Universität Wien)
- Doz. Dr. Michael Pucher (ISF, ÖAW)
- Mag. Dr. Sandra Reitbrecht (PH Wien)
- Univ.-Prof. Mag. Dr. Nikolaus Ritt (Universität Wien)
- Mag. Dr. Hannes Scheutz (Universität Salzburg)
- Jessica Siddins, M.A. (Universität Graz)
- Mag. Dr. Barbara Soukup-Kirchner, MSc (Universität Wien)
- Doz. Dr. Lorenzo Spreafico (Universität Bozen)
- Ao.Univ.-Prof. Mag. Dr. Ralf Vollmann (Universität Graz)

ORGANISATIONSTEAM

- Univ.-Doz. Dr. Sylvia Moosmüller † (Institut für Schallforschung, ÖAW)
- Univ.-Prof. Dr. habil. Elissa Pustka, Dipl.-Journ.Univ. M.A (Institut für Romanistik)

- Barbara Tiefenbacher (Sekretariat, Arbeitsbereich Prof. Pustka)

- Marc Chalier, M.A. (Institut für Romanistik)
- Mag. Johanna Fanta-Jende (Institut für Germanistik)
- Fabian Fleißner, BA MA (Institut für Germanistik)
- Mag. Julia Forster, BA (Institut für Romanistik)
- Luise Jansen (Institut für Romanistik)
- Mag. Julia Kamerhuber, BA (Institut für Romanistik)
- Nicola Klingler, M.A. (Institut für Schallforschung, ÖAW)
- Univ.-Prof. Dr. Alexandra Lenz (Institut für Germanistik)
- Mag. Hannah Leykum (Institut für Schallforschung, ÖAW)
- Mag. Dr. Markus Pöchtrager (Institut für Sprachwissenschaft)
- Mag. Thomas Skarits (Institut für Romanistik)

TAGUNGSPROGRAMM

Mittwoch, 05.09.2018

9.15 – 16.30 Uhr	CLARIN-D Workshop
19.00 Uhr	*Conference warm-up* im Universitätsbräuhaus (Alser Straße 4, Hof 1)

Donnerstag 06.09.2018

8.00 – 9.00 Uhr	**Registrierung**
9.00 – 9.15 Uhr	**Begrüßung** Elissa Pustka
9.15 – 10.15 Uhr	**Plenarvortrag** *Jean-Pierre Chevrot (Grenoble)* "Interfaces of sociolinguistics: Cognition, acquisition and massive data"
10.15 – 11.05 Uhr	Kaffeepause
11.05 – 12.45 Uhr	**Sektion I** *(1) Weirich, Simpson* "Phonetic characteristics of motherese and fatherese" *(2) Eger, Mitterer, Reinisch* "Processing of German /h/ and /ʔ/ by Italian learners" *(3) Kerschhofer-Puhalo* "Phonetische und phonologische Ähnlichkeit in der Vokalperzeption – Zur Operationalisierung eines Konzeptes der L2-Forschung" *(4) Kornder, Mennen* "Arnold Schwarzenegger Now and Then: A longitudinal pilot investigation into Schwarzenegger's production of plosives in German and English"
12.45 – 14.00 Uhr	Mittagspause
14.00 – 15.15 Uhr	**Sektion II** *(5) Belz* Vowel quality of filler particles differs as a function of dialogue move *(6) Jannedy, Weirich, Leeman* "The ecological validity of crowd sourced data" *(7) Vagnini-Holbl, Draxler* "Comparing acoustic measurements from manual and automatic segmentations"

15.15 – 15.45 Uhr **Poster Slam der Postersession I**

Asano, Mitterer
Rich intonational patterns are more advantageous than a lexical pitch accent in storing (non-)linguistic pitch

Betz, Carlmeyer, Wagner, Wrede
Adaptive Hesitation Insertion Model for Incremental Speech Synthesis (Demo)

Brackhane
/'zama/ = <sag mal>? Perzeption phonetisch ambiger Reduktionsformen

Draxler, Trouvain
The new kid on the block - Zur Rolle des Datenarchivars bei Sprachdatensammlungen

Endes
F0 adaptation in conversation: the effect of sex

Eppensteiner
Lieben oder leben? Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Sprecher/innen mit L1 Japanisch

Gessinger, Raveh, Möbius, Steiner
Phonetic Accommodation in HCI: Introducing a Wizard-of-Oz Experiment

Grünke
Zur Intonation des Spanischen junger bilingualer Sprecher aus Girona

Klingler
Homonym timing of vowel + consonant sequences in three Bavarian varieties

Otto, Simpson
Comparing the lingual correlates of coda /r/ in North and East Central German

Pistor, Keil
VJ.PEAT: Automated measurement and classification of prosodic features

Sennema, Hamann
Konsonantencluster in der Wortverarbeitung vietnamesischsprachiger Deutschlernender

Thon, Weirich, Simpson
Sprechgeschwindigkeit von Müttern und Vätern in kind- und erwachsenengerichteter Sprache

	Tomaschek, Leemann How do anatomical differences affect individual speech rhythm? *Wolfswinkler, Kleber, Cunha* On the effects of state borders and standard varieties on the spread of the Bavarian /e-ɛ/ merger
15.45 – 16.45 Uhr	**Postersession mit Kaffeepause**
16.45 – 17.15 Uhr	**P&P-Besprechung**
17.30 – 19.45 Uhr	**Rahmenprogramm** Stadtführung (Treffpunkt: Info-Point) Führung durch das Phonetik-Labor (Treffpunkt: Info-Point)
20.00 Uhr	**Conference Dinner am Institut für Romanistik**

Freitag 07.09.2018

9.00 – 10.00 Uhr	**Plenarvortrag** *Chiara Celata (Pisa)* "Plasticity of the native dialect in adulthood"
10.00 – 10.30 Uhr	**Poster Slam der Postersession II** *Adam, Henneberg, Jördens, Krämer, Pankratz, Wesolek, Yan, Mooshammer* The perception of the /ʃ - ç/contrast in younger and older German listeners *Bross* The good, the bad, the bouba, and the kiki. Cross-modal correspondences between evaluative meanings, speech-sounds, and object shapes *Carignan, Hoole, Kunay, Pouplier, Harrington, Frahm* Examining the temporal extent of co-articulatory nasalization in German using real-time MRI *Cronenberg, Haller, Patáková, Draxler,* Devoicing of /z/ in the Bavarian dialect: a large scale study *Carlsen, Doberaß, Huttner, Radtke* How to Measure a Pleasant Voice *Ebel* Sprechgeschwindigkeit in Lehrvideos auf YouTube *Funk, Voigt-Zimmermann, Simpson* Junge oder Mädchen? Zur Geschlechtsidentifikation präpubertärer Stimmen

	Feindt Realisations of /ɹ/ in Learner English – The Role of Bilingualism *Genzel, Kügler* The prosody of Yes-No questions in Pulaar *Leykum* Voice quality of ironic utterances *Linke, Torres-Tamarit* German dorsal fricative assimilation (DFA) revisited *Tomaschek, Baayen, Tucker* Changes in vowel stem articulation due to suffixation -- Investigating the effects of practice *Trouvain, Möbius* Zu phonetischen Partikeln in "stillen" Pausen *Voße, Wagner* Automatic Acoustic Prominence Annotation: Re-Implementing a Tool for Model Exploration and Refinement *Wokurek, Schweitzer* Adaptation of Harmonic Spectrum-based Voice Quality Parameters in Spontaneous Dialogues
10.30 – 11.30 Uhr	**Postersession mit Kaffeepause**
11.30 – 12.45 Uhr	**Sektion III** (8) *Jochim, Kleber, Klingler, Pucher, Schmid, Zihlmann* Measuring the Role of Hypoarticulation in a Sound Change in Progress in Southern German (9) *Hödl* Kategoriale Wahrnehmung von Voice Onset Time durch österreichische Hörer (10) *Oschkinat, Reinisch, Hoole* Temporal Perturbation of German Quantity Contrasts
12.45 – 14.00 Uhr	Mittagspause
14.00 – 14.50 Uhr	(11) *Mittelhammer, Kleber* Implicit dialect association and accommodation (12) *Riccabona* Faktoren für die Variation des tonal alignment in Nord- und Südtiroler Dialekten
14.50 – 15.20 Uhr	**Poster Slam der Postersession III** *Arnold, Tomaschek* Methodische Betrachtungen zur Erschließung großer Datenmengen in der Phonetik

Brandt, Andreeva, Möbius
Voice quality as a function of information density and prosodic factors

Buchmüller, Feng, Knott, Krämer, Palmer, Pankratz, Reule, Rizaeva, Zöllner, Mooshammer
The influence of phonemic similarity on speech errors and variability in CVC syllables

Chasiotaki, Patáková, Slavik, Kleber
Aspiration und Akkommodation im Slowakischen

Conrad, Manzoni
Die Aussprache des Deutschen in Luxemburg

Fanta
Intersituative language dynamics among the dialect-standard-axis in rural Austria

Fey, Lewandowski
Uptalk bei weiblichen Sprecherinnen im Deutschen – eine Korpusstudie

Frank
The Merger of [eː] and [ɛː] in Standard German

Gube
Normalisierung von Vokalen im DaF-Kontext: Vergleich von Normalisierungsmethoden

Jaekel, Wagner, Betz
Speaking Style Alignment Initiated by Virtual Agents

Kalkhoff
Annotation of Haitian Creole prosody and intonation

Link, Baayen, Tomaschek
Quantitative analysis of Glossolalia – A case of phonetic signals without semantics

Neder, Weirich, Simpson
Der Zusammenhang zwischen selbsteingeschätzter und wahrgenommener Maskulinität in männlichen Stimmen

Wilke, Duran, Lewandowski
Wanted! – Assessing effects of distraction on working memory in speech perception

Wehrle, Cangemi, Vogeley, Grice
The timing of turn-taking in highfunctioning autism

15.20 – 16.20 Uhr **Postersession mit Kaffeepause**

16.20 – 17.35 Uhr	**Sektion IV**
	(13) Geumann, Arnold Vowel movement revisted – Simple or complex metrics?
	(14) Röhr, Niemann, Baumann, Grice Prominence marking depends on expectations
	(15) Harris, Tang, Neasom Do phonologists know too much? A regular, simple, unnatural pattern in English
17.35 – 17.50 Uhr	**Abschluss** Alexandra Lenz

EINGELADENE VORTRÄGE

(Invited Speakers)

Interfaces of sociolinguistics: Cognition, acquisition and massive data

Jean-Pierre Chevrot
(Grenoble)

For fifty years, Sociolinguistics has explored the interactions between language and society. Intrinsically interdisciplinary, since its foundation, it has maintained relations with other fields of social science, such as anthropology, dialectology and sociology. For a decade, certain areas of sociolinguistics have come closer to disciplines further away from the initial epistemic framework of the domain (Chevrot & Nardy, to appear). This new interdisciplinary research involves first several subfields of cognitive science, such as neuroscience, language acquisition and psycholinguistics, which are more or less rooted in the Life sciences (Chevrot, Drager & Foulkes, in progress). Second, the new interdisciplinary research involves several sectors of Computer science, such as Network science, Data sciences and Data modeling (Nguyen, Doğruöz, Rosé, & de Jong, 2016).

The first type of collaboration will be illustrated with studies on the acquisition of sociolinguistic variables in children. Because these studies consider simultaneously the acquisition of linguistic cues and the acquisition of their socio-indexical values, they are grounded both in the psycholinguistic and the sociolinguistic frameworks (Chevrot & Foulkes, 2013; De Vogelaer & Katerbow, 2017). We will focus on French sociolinguistic variables from the phonological level (Barbu, Nardy, Chevrot, & Juhel, 2013; Chevrot, Nardy, & Barbu, 2011; Nardy, Chevrot, & Barbu, 2014) and the results will be put in perspective with results involving other languages (Nardy, Chevrot, & Barbu, 2013).

The second type of collaboration will be documented with studies from the nascent field of Computational sociolinguistics (Nguyen et al., 2017). As a part of Computational social science (Lazer et al., 2009), Computational sociolinguistics results from our new-found ability to collect and analyze vast amounts of data from the use of sensors (proximity sensors, wearable audio recorders, etc.) or from the digital communication (real-time and unsupervised recording of digital traces left on the blogosphere, the social media, the peer-to-peer services, etc.). We will illustrate this second trend with new results on the usage on Twitter of a well-known sociolinguistic variable of French – the optional deletion of the preverbal negative "ne" (Levy Abitbol et al., 2018).

Finally, the conclusion will emphasize two points. First, a very promising avenue of research is to combine these two kinds of collaboration. It is the case of the *DyLNet* Project *(Language Dynamics, Linguistic Learning, and Sociability at Preschool: Benefits of Wireless Proximity Sensors in Collecting Big Data)* (Nardy et al., 2016) that addresses psycholinguist and sociolinguistic issues using the

empirical and modeling tools of Network science (Barabási, 2011). Second, the idea will be defended that - whatever the risks for the autonomy of Language science - these interdisciplinary connections place sociolinguistics in a strategic position for integrating the linguistic, social and cognitive aspects of language at the individual and collective levels.

Placticity of the native dialect in adulthood

Chiara Celata
(Pisa)

Analyses of the speech of adults who have experienced social changes in their lives such as migration to a non-native language setting, an international adoption or even less extreme forms of contact with an L2 (such as full immersion in an L2 course) reveal that the use of one's native language can be challenged by the competing demands of a second language acquired later in life. Phonetic-phonological systems do in fact retain some degree of plasticity even after the completion of the critical period for language acquisition. This talk discusses if socio-indexical values that can be associated to speech patterns are involved in such changes occurring in adulthood. In particular, we have investigated the retention and inter-generational transmission of inherent (as opposed to contact-induced) sociophonetic variables in heritage Italian varieties. Heritage Italian varieties can be strongly influenced by the Romance dialect spoken in the region of origin of the migrants; local pronunciation features are used to convey socio-indexical attributes of regional, more than national, identity. The ultimate aim of our studies is therefore to understand if socio-indexical features, which are rooted in the social dynamics of a linguistic community, change when the speakers move to another community, in which they represent a minority group, and how such changes affect the transmission to subsequent generations of speakers.

VORTRÄGE

Phonetic characteristics of motherese and fatherese

Melanie Weirich, Adrian P. Simpson
Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena
melanie.weirich@uni-jena.de, adrian.simpson@uni-jena.de

Infant-directed speech (IDS) or *motherese* has been found to be characterized by various phonetic modifications such as increased average fundamental frequency, larger f0 excursions (Fernald et al. 1989, Shute & Wheldall 1999) and enhanced vowel spaces (Kuhl et al. 1997, Liu et al. 2003, Uther et al. 2007). Recently, the question arose if the vowel space expansion is indeed an inherent characteristic of IDS with the purpose of enhancing acoustic contrasts between sounds (Kuhl et al. 1997, Uther et al. 2007), or if it rather results from formant raising as a byproduct of increased smiling behavior in IDS (Englund & Behne 2005, Benders 2013, Cristia & Seidel 2013, Burnham et al. 2015). Furthermore, IDS has been investigated primarily in mothers, who traditionally are the main caregivers. Studies including fathers show contradictory results regarding f0 modifications (cf. Shute & Wheldall 1999). A possible interacting factor might be the involvement in childcare as was found in Sheehan (2004) regarding increased f0 in IDS in fathers. We are not aware of any study on vowel space size in IDS that also includes fathers.

The study presented here is part of a larger project (*AVarE*, Weirich & Simpson 2017) investigating IDS and ADS in German and Swedish mothers and fathers during the child's first year. Here, German data from 15 mothers and 18 fathers of the second recording is presented (babies' age: 5m0w4d - 6m3w2d). The data consists of semi-spontaneous speech of participants describing 15 pictures (with recurrent objects and animals carrying target sounds) to their child (IDS) and to the experimenter (ADS). F1 and F2 were estimated in /iː ɛ a uː/ and used to calculate the acoustic vowel space (AVS) for each speaker and register. Also, mean fundamental frequency (f0) and variation in f0 (SD) was calculated. Furthermore, DCT coefficients of the sibilants /s/ and /ʃ/ were estimated and the Euclidean Distances between the sibilants in DCT1xDCT2xDCT3 space was measured. To normalize between speakers and genders, all parameters (vowel space size, f0 measures and acoustic contrast between sibilants) were expressed as percentages of values measured at an earlier recording of ADS arbitrarily chosen as a reference point (for each speaker respectively). In addition, a questionnaire assessed the involvement in childcare for each participant.

Linear mixed models were run with speaker as random effect and the fixed factors gender, speech register and involvement in childcare. Significance testing was done using Likelihood Ratio Tests comparing the model with and without the factor/the interaction in question. Results show a significant effect of register with IDS revealing larger AVS (χ^2 (1) = 5.4, $p < .05$), higher mean f0 ($\chi^2(1)=27.37$, $p < .0001$) and larger f0 variation ($\chi^2(1) = 11.81$, $p < .001$) than ADS (see Figures 1 and 2). All effects were independent of gender or involvement in childcare. No effect of speech register in sibilant contrast was found. However, IDS and ADS did differ in individual coefficients, e.g. DCT3 varied between the registers for both sibilants and sexes.

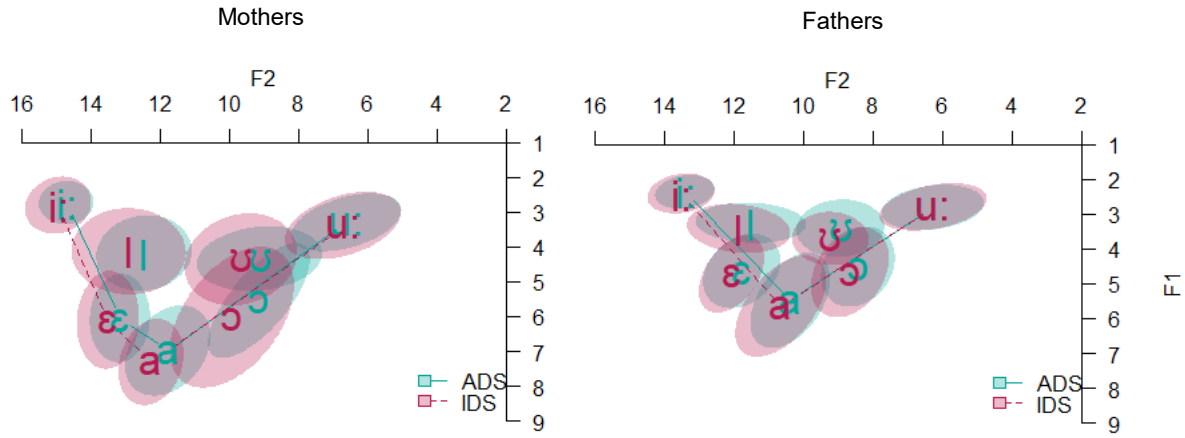


Figure 1: F1xF2 dimension of vowels in IDS and ADS separated by speaker gender

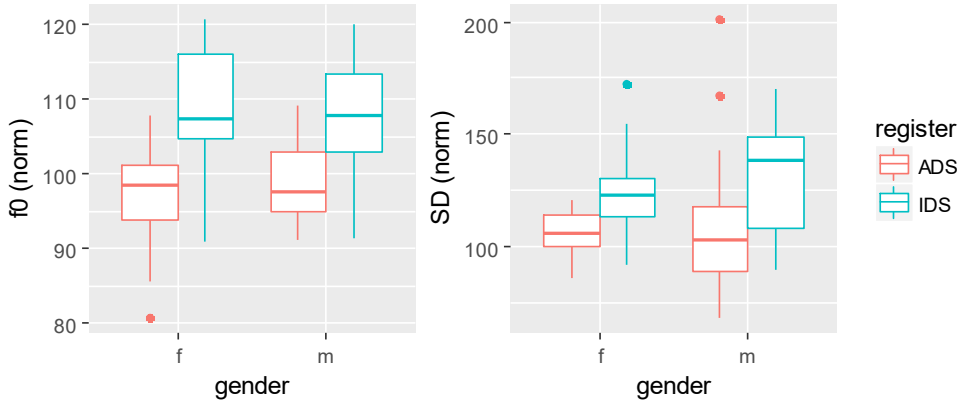


Figure 2: Mean fundamental frequency (f0) and standard deviation (SD) in % for IDS and ADS separated by speaker gender

References:

Benders, T. (2013) Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. Infant Behavior & Development 36, 847-862.

Burnham, E.B., Wieland,E.A., Kondaurova, M.V., Devin McAuley, J., Bergeson, R.R & Dilley, L.C. (2015) Phonetic Modification of Vowel Space in Storybook Speech to Infants up to 2 Years of Age. J Speech Lang Hear Res. 58(2): 241–253.

Cristia, A. & Seidl, A. (2013) The hyperarticulation hypothesis of infant directed speech. Journal of Child Language 40 (2), 1-22.

Englund, K. & Behne, D. (2005) Infant directed speech in natural interaction: Norwegian vowel quantity and quality. Journal of Psycholinguistic Research 34, 259–80.

Fernald, A., Taeschner, T., Dunn, J. & Papousek, M. (1989) A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. Journal of Child Language 16(3) 477-501.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A. & Chistovich, L.A. (1997) Cross-language analysis of phonetic units in language addressed to infants. Science 277 (5326), 684-686.

Liu, H. M., Kuhl, P. K. & Tsao, F. M. (2003) An association between mothers' speech clarity and infants' speech discrimination skills. Developmental Science 6, F1-F10.

Sheehan E.A. (2004) Influence of paternal involvement on fathers' infant-directed speech and infants' brain activity to male and female speech. PhD dissertation. Emory University.

Shute, B. & Wheldall, K. (1999) Fundamental frequency and temporal modifications in the speech of british fathers to their children. Educational Psychology 19(2), 221-233.

Uther, M., Knoll, M. A. & Burnham, D. (2007) Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. Speech Communication 49, 2–7.

Weirich, M. & Simpson, A.P. (2017) Acoustic correlates of parental role and gender identity in the speech of expecting parents, Interspeech 2017, Stockholm, 924-928.

Processing of German /h/ and /ʔ/ by Italian learners

Nikola Anna Eger[1], Holger Mitterer[2], Eva Reinisch[3]
[1,3]Institute of Phonetics and Speech Processing, LMU Munich
[2]Department of Cognitive Science, University of Malta
eger@phonetik.uni-muenchen.de, *holger.mitterer@um.edu.mt*, *evarei@phonetik.uni-muenchen.de*

While most research on second language (L2) sound learning has focused on sounds that are difficult because they are close but not identical to first language (L1) categories [1, 2] relatively less is known about the acquisition of L2 sounds that occur in the L1 only as paralinguistic sounds or in hyperarticulated speech. A case in point are German /h/ and /ʔ/ for Italian learners. Importantly, for Italians German /h/ is a well-known problem. In contrast, the presence of /ʔ/ in words that are spelled with an initial vowel is often not even known to native speakers of German. Yet, both /h/ and /ʔ/ are important for spoken word recognition for native speakers of German [3]. Here we asked whether the different status of awareness of the presence and difficulty of these sounds – including the orthographic coding of /h/ but not /ʔ/ – plays a role in L2 word recognition and is reflected in L2 lexical representations.

Forty /h/- and 40 /ʔ/-initial picturable German words were recorded by a native speaker of German in three versions: correctly, with the two critical sounds substituted (e.g., */ʔ/andschuh* for *Handschuh*), or the critical sound deleted. To highlight the presence vs. absence of the critical sounds words were embedded in carrier sentences following a nasal, such as *Er gab ihr seine**n H**andschuh* "He gave her his glove". Thirty-six Italian medium-to-good proficiency learners of German were asked to listen to the sentences and choose the target from three pictures on a screen while their eye-movements were monitored. Half of the participants were assigned the correct vs. substituted condition, half the correct vs. deleted condition.

Analyses of target fixations revealed that word recognition in the correct vs. substituted condition was not different for either target sound. However, in the deleted vs. correct condition learners fixated on the target less if the initial sound was deleted, but even this effect was subtle. The deletion effect was also larger in size for /h/ than for /ʔ/, but given that the main effect was subtle, the interaction (deletion by target sound) was far from significant. Assuming that eye-tracking reveals the degree of match between acoustic input and lexical representations that are being accessed [4], these results suggest that the Italian learners have established a representation for a fuzzy glottal sound for (otherwise) vowel-initial words, which might be stronger for /h/-initial words. An additional explicit goodness rating task of the targets in the different conditions showed effects of substitutions and deletions for both, /ʔ/ and /h/, suggesting that learners can acoustically differentiate between them, presumably because they are known as paralinguistic sounds in the L1. We conclude that the awareness of a difficult L2 sound may help forming a new category. Yet, the actual representation may not be fine-grained enough to be used in processing small differences.

Literature

[1] J. E. Flege, “Assessing constraints on second-language segmental production and perception.” in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, N. Shiller & A. Meyer (Hrsg.), Berlin: Mouton de Gruyter, 2003, pp. 319–358.

[2] M. Broersma, “Increased lexical activation and reduced competition in second-language listening“. *Language and Cognitive Processes*, 27, 1205-1224, 2012.

[3] H. Mitterer & E. Reinisch, “Letters don’t matter: No effect of orthography on the perception of conversational speech“. *Journal of Memory and Language*, 85, 116-134, 2015.

[4] D. Dahan, J. Magnuson, & M. Tanenhaus, “Time course of frequency effects in spoken-word recognition: Evidence from eye-movements“. *Cognitive Psychology*, 42, 317-367, 2001.

Phonetische und phonologische Ähnlichkeit in der Vokalperzeption – Zur Operationalisierung eines Konzepts der L2-Forschung

Ähnlichkeit ist eines der zentralen Konzepte in Modellen fremdsprachlicher Perzeption (Best 1995; Flege 1987, 1995; Kuhl 1992, 1993). Arbeiten, die eine Operationalisierung dieses Begriffs anstreben, sind jedoch rar. Auf Basis von Daten eines Vokalidentifikationsexperiment mit 173 L2-Deutsch-Lernenden (Kerschhofer-Puhalo 2014) werden hier verschiedene Formen der Operationalisierung und Visualisierung von phonetischer und phonologischer Ähnlichkeit von L2-Vokalen diskutiert. Das Sample bestand aus zehn L1-Subsamples von Deutschlernenden, denen in einem *forced choice*-Identifikationstest alle 15 Vokalphoneme des Deutschen in wechselndem prä- und postvokalischen Kontext präsentiert wurden. Die Responses der L2-Lernenden, Fehlerhäufigkeiten und Präferenzen für bestimmte Response-Kategorien wurden mit deskriptiv-statistischen Methoden sprachspezifisch und sprachübergreifend ausgewertet. Die Identifikationen wurden in Konfusionsmatrizen zusammengefasst und daraus Werte perzipierter Ähnlichkeit und Distanz zwischen L2-Vokalkategorien errechnet (Johnson 2012). Mittels Hierarchischer Clusteranalyse und Multidimensionaler Skalierung wurden Aussagen zu akustischer und sprachspezifischer perzeptueller Ähnlichkeit und Distanz, der Fehleranfälligkeit bestimmter Kategorien und der Richtung perzeptueller Substitutionsprozesse (*Bias*) abgeleitet. Mittels Multidimensionaler Skalierung (Shepard 1972; Terbeek 1977; Kewley-Port & Atal 1989; Iverson & Kuhl 1995; Fox, Flege & Munro 1995; Francis & Nusbaum 2002) werden L1-spezifische räumliche Repräsentationen des perzeptuellen Vokalraums (*perceptual vowel maps*) abgeleitet, die eine *Visualisierung* sprachspezifischer *perzeptueller Ähnlichkeiten und Distanzen* zwischen L2-Vokalkategorien durch Lernende verschiedener Erstsprachen ermöglichen.

Die Modellierung perzeptueller Ähnlichkeiten zwischen L2-Vokalkategorien basiert auf akustischen Eigenschaften des Lautsignals und sprachspezifischen wie auch allgemeineren kognitiven Präferenzen (*biases*) sowie der relativen Gewichtung (*attentional weight*) von Dimensionen im perzeptuellen Raum. Wahrgenommene perzeptuelle Ähnlichkeit s_{ij} zwischen Kategorien i und j wird als Resultat der Interaktion von phonetischer Nähe p_{ij} und der unterschiedlichen Gewichtung von Stimuli-Biases b_i und Response-Biases b_j modelliert ($s_{ij} = p_{ij} * b_i * b_j$). *Biases* sind von phonetischen Eigenschaften des Input-Signals, dem Set von Stimuli- und Response-Kategorien und individuellem sprachlichen Wissen, Erfahrungen und Repräsentationen des Zielsprachen-Systems bestimmt.

References

[1] Best, C. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, pp. 171–204.

[2] Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, pp. 47–65.

[3] Flege, J. 1995. Second-language speech learning: Theory, findings, and problems. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, pp. 233–277.

[4] Fox, R., Flege, J., Munro, M. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97, pp. 2540–2551.

[5] Francis, A. & Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2), pp. 349–366.

[6] Hillenbrand, J. 2013. Static and dynamic approaches to vowel perception. In: Morrison, G. & Assmann, P. (eds.). *Vowel inherent Spectral Change. Modern Acoustics and Signal Processing*. Berlin, Heidelberg: Springer, p. 9–30.

[7] Iverson, P. & Kuhl, P. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *JASA*, 97, pp. 553–562.

[8] Johnson, K. 2012. *Acoustic and Audotory phonetics*. 3rd ed. Wiley-Blackwell.

[9] Kerschhofer-Puhalo, N. 2014. *Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception. An experimental cross-language comparison of German vowel identification by non-native listeners*. Dissertation, University of Vienna.

[10] Kewley-Port, D. & Atal, B. 1989. Perceptual distances between vowels located in a limited phonetic space. *JASA*, 85, pp. 1726–1740.

[11] Kuhl, P. 1993. An examination of the 'perceptual magnet' effect. *JASA*, 93, p. 2423.

[12] Kuhl. P. 1992. Speech prototypes: studies in the nature, function, ontogeny and phylogeny of the 'centers' of speech categories. In: Tohkura, Y., Vatikiotis-Bateson, W., Sagisaka, Y. (eds.). *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, pp. 197–219.

[13] Shepard, R. 1972. Psychological representation of speech sounds. In: David, E. & Denes, P. (ed.). *Human Communication: A Unified View*. New York: McGraw-Hill, pp. 67–113. [http://courses.cit.cornell.edu/ling4476/shepard-on-speech-sounds.pdf]

[14] Shepard, R. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, 210, pp. 390–398. [http://serendip.brynmawr.edu/local/scisoc/emergence/shepard80.pdf]

[15] Terbeek, D. 1977. *A Cross-Language Multi-Dimensional Scaling Study of Vowel Perception*. PhD Dissertation. University of California, Los Angeles. [http://escholarship.org/uc/item/3nx4g138?display=all#page-1]

Arnold Schwarzenegger Now and Then: A longitudinal pilot investigation into Schwarzenegger‘s production of plosives in German and English

Lisa Kornder & Ineke Mennen
Department of English Linguistics, University of Graz

Arnold Alois Schwarzenegger (AS henceforth) was born into an Austrian family in 1947 in Thal, a small village near Graz (Austria). At the age of 21, AS left his home country and moved to the United States where he lived permanently ever since. Against the background of his language learning history, AS can be described as a late consecutive bilingual who acquired the second language (L2; English) in adulthood when his first language (L1; Austrian German) had already been fully developed.

The present pilot study is part of a larger research project which aims at examining AS’s L1 and L2 segmental speech production over a period of 49 years, i.e. from the time he moved to the US in 1968 up to the 2010s. Based on acoustic investigations of plosives and vowel space, the main objectives are to find out (1) whether speech production in his L2 has improved since he moved to an L2-speaking environment, (2) whether an acoustic analysis of his segmental speech production provides evidence for L1 attrition, i.e. changes in L1 phonetic categories due to L2 learning experience [1] [2] [3], and (3) if it is possible to establish a relationship between the development of AS’s L2 and potential modifications of phonetic categories in his L1.

For this purpose, speech data were taken from various interviews with AS which were conducted in either Austrian German (AG) or American English (AE) and were broadcasted on different US and Austrian TV and radio channels. The AE and AG corpora were divided into early (1977-1989) and late (2010-2017) speech samples, respectively. The pilot investigation presents preliminary results of AS’s realization of voice onset time (VOT) contrast in word-initial plosives in his L1 and L2. While English distinguishes between short-lag and long-lag VOT [4], speakers of AG varieties have been observed to neutralize VOT contrast in word-initial bilabial and alveolar plosives in conversational speech [5], [6]. Based on the differences between AG and English plosive contrasts, the aim was to determine if and to what extent AS’s production of VOT in word-initial plosives in his L1 and L2 has changed over the past 49 years, i.e. if L2 VOT production has moved towards more native-like norms and if the realization of VOT contrast in his current L1 speech has changed compared to early L1 speech samples. Preliminary results of an investigation into AS’s current speech show that he realizes a significant VOT contrast in his L2 which is in fact quite unlike typical production of the same contrast in AG.

References

[1] De Bot, K. (1998). The psycholinguistics of language loss. In G. Extra & L. Verhoeven (Eds.), *Bilingualism and migration* (pp. 345–361). Berlin: Mouton de Gruyter.

[2] Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, M. Keijzer, & L. Weilemar (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issues* (pp. 1–46). Amsterdam: John Benjamins.

[3] Köpke, B., & Schmid, M. (2017). *First Language Attrition*. John Benjamins.

[4] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word 20* (3), 384-422.

[5] Moosmüller, S., Schmid, C., & Brandstätter, J. (2015). Standard Austrian German. *Journal of the International Phonetic Association* 45 (3), 339-348.

[6] Wiesinger, P. (2008). *Das österreichische Deutsch in Gegenwart und Geschichte*. Wien: LIT.

Vowel quality of filler particles differs as a function of dialogue move

Malte Belz
Humboldt-Universität zu Berlin

Filler particles (FP) take different forms, such as [ɛː ɛːm əː əːm ??? !], and many more. Research on their acoustic form can be grouped into two main trends. On the one hand, forms are categorized in a 'scriptistic' way, meaning their forms are reduced to a vocalic and a vocalic-nasal form, and then analyzed in various contexts (Clark & Fox Tree 2002). On the other hand, there are studies looking closer at their actual acoustic features, such as f0 and duration, but in the context of human-machine-interaction (Shriberg & Lickley 1993), monologues (Swerts 1998), or speaker identification (Hughes et al. 2016). Here, I will focus on the link between FP form and their discourse-specific functions in dialogue, investigating whether form and function "are learned pairings" (Goldberg 2006: 5).

Dialogic interaction is realized by a variety of dialogue moves, e. g. narrative sequences, questions, replies, backchanneling, etc. I conduct a corpus-based study on spontaneous, task-free German dialogues using and enhancing the GECO corpus (Schweitzer & Lewandowski, 2013) with new layers of annotation. FPs are identified and categorized into vocalic, vocalic-nasal, glottal and click forms. Dialogue moves are annotated following the dialogue coding system of Carletta et al. (1997). So far, 8 speakers are annotated, using a total of 76 clicks, 105 glottal FPs, and 391 vocalic/vocalic-nasal FPs, representing 1.3 % of all words in the corpus. Figure 1 shows the formant trajectories of *äh* and *ähm* vowels in the dialogue moves *explain* (which contains narratives and explanations) and *reply-wh* (which contains replies to wh-questions).

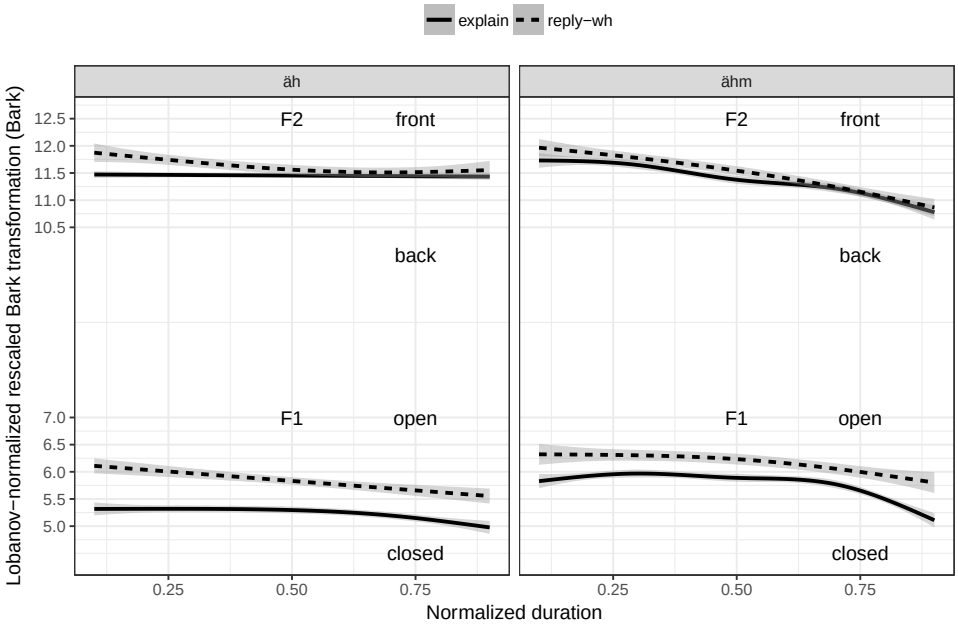


Figure 1: *Mean time normalized formant trajectories for* äh *and* ähm *occurring within the dialogue move categories* explain *and* reply-wh.

FP vowels produced in reply to a wh-question show a significantly higher formant frequency F1 for both *äh* and *ähm* as opposed to FP vowels in narrative sequences. Thus, FPs in wh-replies are uttered in a more open position in the vowel chart (see Figure 2). The difference amounts to ≈ 50 Hz. These results provide new insights into the *prima facie* arbitrary form variation of filler particles. It seems that their acoustic form is (although showing idiosyncratic tendencies as well) more sensitive to dialogue context than previously assumed.

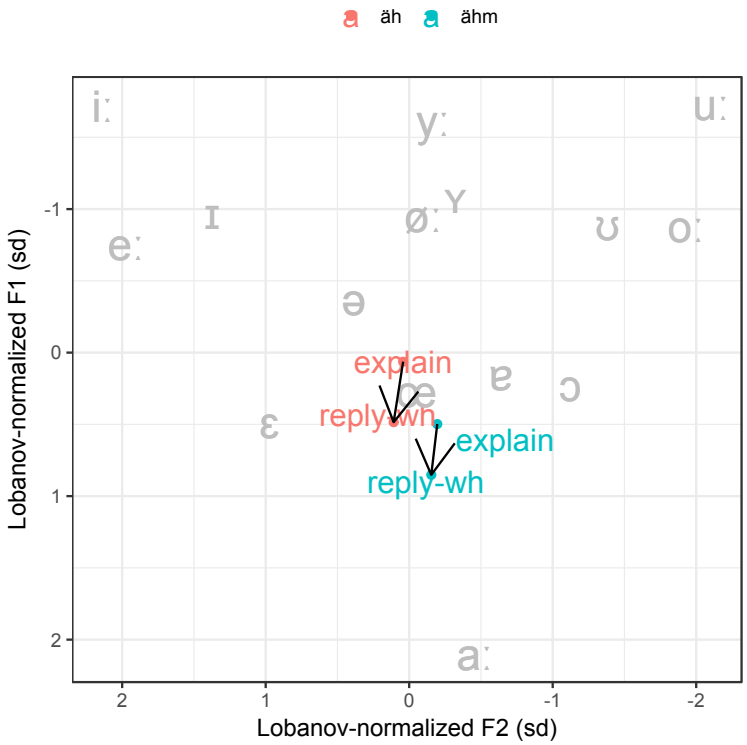


Figure 2: *Mean values of F1 and F2 of FP vowels over all speakers, showing a shift in FP vowel closeness for* explain *moves as compared to* reply-wh *moves. Vowels in* ähm *are uttered more to the back.*

References

[1] J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon & A. H. Anderson, "The reliability of a dialogue structure coding scheme," *Computational Linguistics*, 23(1), pp. 13–31, 1997.

[2] H. H. Clark, & J. E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, pp. 73–111, 84(1). 2002.

[3] A. E. Goldberg, *Constructions at Work. The Nature of Generalization in Language*, New York: Oxford University Press, 2006.

[4] V. Hughes, S. Wood, & P Foulkes, "Strength of forensic voice comparison evidence from the acoustics of filled pauses," *The International Journal of Speech, Language and the Law*, 23(1), pp. 99–132, 2016.

[5] A. Schweitzer, N. Lewandowski, *Convergence of Articulation Rate in Spontaneous Speech*, Proceedings of Interspeech, pp. 525–529, 2013.

[6] M. Swerts, "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, pp. 485–496, 30. 1998.

The ecological validity of *crowd sourced* data

Stefanie Jannedy[1], Melanie Weirich[2], Adrian Leemann[3]
[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin
[2]Institut für Germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena
[3]Department of Linguistics and English Language, Lancaster University
jannedy@leibniz-zas.de, melanie.weirich@uni-jena.de, a.leemann@lancaster.ac.uk

A comparison of the acoustic analysis of the five voiceless German fricatives from lab recorded non-words with fricatives extracted from words recorded with the smartphone app *Deutschklang* [1] was inconclusive with regard to the usability of app recorded speech for phonetic analyses [2]. By removing the confounding effects of having different speech material between lab recorded non-words and ecologically-valid smartphone recordings of the Wenker sentences, we have now redone a comparison: We have now compared the voiceless fricatives /f s ʃ ç χ/ from *Deutschklang* recorded Wenker sentences (11 speakers) to the fricatives taken from the same Wenker sentences in lab recorded speech (6 speakers), essentially removing confounding effects of differences in coarticulation between different recording materials. Acoustic analyses comprise Center of Gravity (COG), standard deviation (SD), skewness and kurtosis. Furthermore, Discrete Cosine Transformation (DCT, [3] [4]) was used to quantify the shape of the spectra.

Several linear mixed models were run with the different acoustic measurements as dependent variables, the potential fixed effects *fricative* and *recording condition* (lab vs. app) and the random effects *subject* and *word*. Since the aim of the study was to test the reliability of the app data with respect to analyzing spectral contrasts between fricatives, we were mostly interested in a potential interaction between fricative and recording condition pointing to differences between the lab and app data. For COG a significant interaction between fricative and recording condition was found ($\chi^2(5)=14.4$ $p < .05$), However, post hoc tests for multiple comparisons between fricatives for each recording condition revealed very similar results with the comparisons /ç-f/, /ç-ʃ/ and /f-ʃ/ to be not significant. For SD also a significant interaction between fricative and recording condition was found ($\chi^2(5)=18.6$, $p< .01$). While /f/ differed significantly from all other fricatives in both recording conditions ($p < .001$), the comparison /s-χ/ was significant only in the lab condition ($p<.05$). For skewness, kurtosis, DCT1 and DCT2 no interaction between fricative and recording condition was found, neither a main effect of recording condition. For DCT3 the interaction turned out significant ($\chi^2(5)=12.3$, $p < .05$). Here, differences were found for /f-ʃ/ and /x-ʃ/ only in the app data ($p < .05$).

Our results suggests that data collected with the aid of an app provides usable data, since the distribution of the data in the app and lab condition is very similar. However, we also had to prune our database to select speakers who recorded themselves in a quiet environment without much background noise, factually eliminating about half of the participants. Also, the Discrete Cosine plots show more overlap of the categories than clean lab based data. Nevertheless, we conclude that the distribution of data is fairly similar so that the advantage of having a larger sample size outweighs the disadvantage of having more overlap between the categories. We are currently adding more data to both groups.

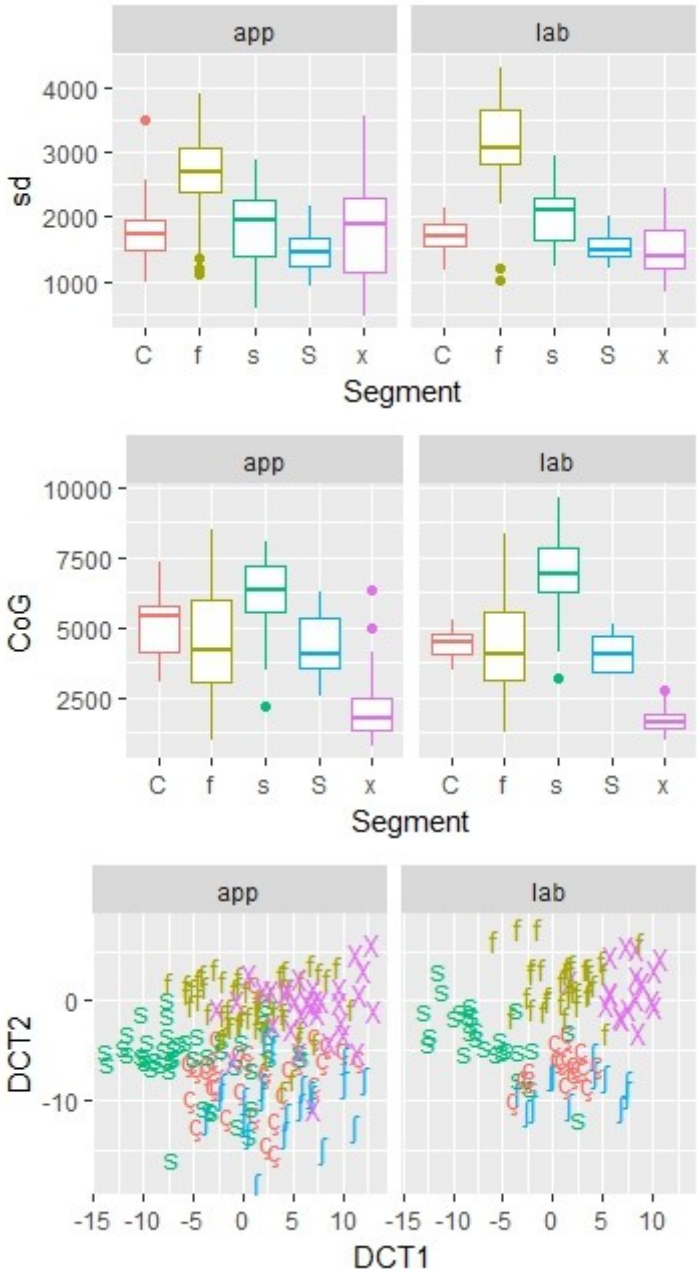


Abbildung 1: *Spectral characteristics of German fricatives in two recording conditions*

Literatur

[1]Leemann, A. (2016) Deutschklang. Smartphone App. https://play.google.com/store/apps/details?id=ch.deutschklang&hl=de

[2]Jannedy, S., Weirich, M. & Leemann, A. (2017) Voiceless Fricatives in Deutschklang. P&P13, Berlin.

[3]Jannedy, S. & Weirich, M. (2017) Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. Journal of the Acoustical Society of America 142 (1), 395-405.

[4]Watson, C. I. & Harrington, J. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. Journal of the Acoustical Society of America 106, 458-468.

Comparing acoustic measurements from manual and automatic segmentations

Constanza Vagnini-Holbl[1], Christoph Draxler[2]
[1]Institut für Romanistik, LMU München
[2]Institut für Phonetik und Sprachverarbeitung, LMU München
constanza@holbl.de, draxler@phonetik.uni-muenchen.de

Manual transcription and segmentation of speech is a time-consuming (> 500 times real time) and thus very expensive task. Automatic segmentation is faster (< 5 times real time), and its quality for phonemic segmentation has reached 95% of human performance in inter-rater comparisons [1]. It is now common practice to perform an automatic segmentation first, and then manually correct it – which again is time-consuming (> 300 times real time). The research question is whether this costly manual correction is necessary if one is only interested in acoustic measurements. This pilot study presents a comparison of pitch and formant measurements for manually and automatically segmented speech.

The corpus contains read speech recordings of Spanish native speakers and of non-native speakers of all proficiency levels learning Spanish. The items recorded were mostly isolated words. A total of 8500 recordings of 94 speakers (33m, 61f) were made. These recordings were automatically segmented by WebMAUS [2] using the Castilian Spanish settings. 1200 of these recordings were then corrected manually by adding an interval tier using Praat [3] (figure 1). All segmentations were imported into a multi-corpus relational database system, along with pitch and formant measurements [4]. The acoustic measures were computed by Praat, with the default formant settings for male and female speakers.

For this study, we restricted the comparison to the vowels /a, e, i, o, u/. We compared duration, f0, and f1 and f2 for those segments that have the same label in both the automatic and the manual segmentation. Pitch is computed as the average of individual pitch values for the entire length of the segment. For the formant measurements we took the average of the mid 60% of the segment. Tables 1 to 3 show the results for the segment durations, f0, and f1 and f2.

Conclusions

Most differences in vowel segment durations are below the segmentation precision of MAUS (default 10ms) and hence negligible. Although f0 is a speaker-dependent measure, the f0 values for both male and female may vary up to approx. 10% between the different vowels (intrinsic pitch [5]). The measured difference in f0 between automatic and manual segmentation is virtually non-existent at max. 0.9Hz – an expected result. The same is true for formant measurements. They vary by max. 2%.

Although these results were computed for vowels only and only for a subset of the corpus, they indicate that acoustic measurements such as pitch and formants in WebMAUS automatic segmentations are very close to those of manual segmentations. If this result holds true for a more thorough analysis, then there is no longer a need for manual verification of automatic segmentations for two important acoustical measurements. This will greatly reduce the transcription effort both in terms of time and cost.

Table 1: *Segment durations for automatic segmentations and the average difference in segment durations between automatically and manually segmented vowel segments ($d_{auto} - d_{man}$). Positive values indicate longer durations for automatic segmentation*

Duration (ms)	a	e	i	o	u
duration	172	131	131	145	103
Δ duration	7.1	0.5	13.3	2.2	8.5

Table 2: *f0 values for automatic segmentations and the average difference in f0 by sex between automatically and manually segmented vowel segments ($f0_{auto} - f0_{man}$)*

Pitch (Hz)	a	e	i	o	u
male	136	127	138	134	140
Δ male	0.4	0.4	0.3	0.1	0.8
female	232	237	246	249	252
Δ female	0.4	0.3	0.1	0.9	0.1

Table 3: *Formant values for automatic segmentations and the average difference in f1 and f2 between automatically and manually segmented vowel segments ($fx_{auto} - fx_{man}$)*

Formants (Hz)	a	e	i	o	u
f1	840	477	332	498	364
Δ f1	3	1	3	2	6
f2	1489	2230	2349	1023	1119
Δ f2	3	9	8	4	10

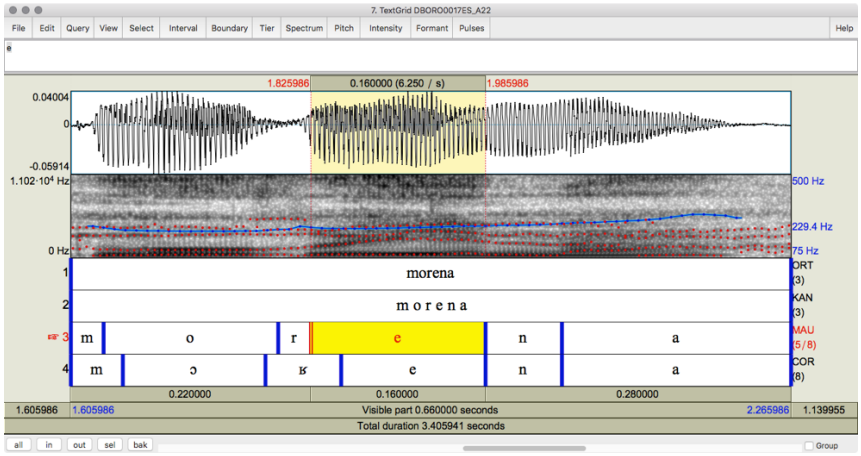


Figure 1: *Sample audio file with tiers for automatic (MAU) and manual (COR) segmentations*

References

[1] F. Schiel, *MAUS goes iterative*, in Proc. LREC, pp. 1015-1018, Lisbon, 2004

[2] Th. Kisler, F. Schiel, H. Sloetjes, *Signal processing via web services: the use case WebMAUS.* Proc. Digital Humanities, pp. 30-34, Hamburg, 2012

[3] P. Boersma, *Praat – a system for doing phonetics by computer*, in *Glot International,* Vol. 5 No 9/19, pp. 341-345, 2001.

[4] Chr. Draxler, St. Kleiner, *A cross-database comparison of two large German speech databases*, Proc. *ICPhS*, Glasgow, 2015

[5] W. Ewan, *Explaining the intrinsic pitch of vowels*, in *The Journal of the Acoustical Society of America* 58, S40, 1975; doi: 10.1121/1.2002115

Measuring the Role of Hypoarticulation in a Sound Change in Progress in Southern German

Markus Jochim[1], Felicitas Kleber[1], Nicola Klingler[2], Michael Pucher[2], Stephan Schmid[3], Urban Zihlmann[3]
[1]IPS, Ludwig-Maximilians-Universität München, [2]ISF, Österreichische Akademie der Wissenschaften, [3]Phonetisches Laboratorium, Universität Zürich
{markusjochim|kleber}@phonetik.uni-muenchen.de,
{nicola.klingler|michael.pucher}@oeaw.ac.at, {stephan.schmid|urban.zihlmann}@uzh.ch

We present results from perception experiments designed to investigate the role of speech-rate-induced hypoarticulation in diachronic sound change. Hypoarticulation has been identified as one potential source for such change [1, 2]. However, this account is challenged by the fact that listeners generally compensate for the effects of fast speech: In minimal pairs differing in vowel length (e. g. German *Stadt*, 'city', vs. *Staat*, 'state'), they tend to classify identical word tokens as containing long vowels when they occur in a fast carrier phrase (i. e. shorter duration) and vice versa [3]. The aim of the present experiments is to test whether compensation is diminished in listeners partaking in a sound change in progress.

The experiments are part of a larger-scale study on the evolution of quantity contrasts in consonants and vowels in Southern German (including Austrian and Swiss) varieties. Standard German features a phonemic vowel length (cf. example above) and a consonantal fortis-lenis contrast (e. g. *Hagen*, a given name, vs. *Haken*, 'hook'), both of which are cued by duration (the latter in particular in the absence of aspiration [4, 5, 6]). The two contrasts can be freely combined in standard German vowel-consonant sequences. Central Bavarian (CB) varieties spoken in the south of Germany and in Austria differ from standard German in that combinations of long+fortis and short+lenis are phonotactically illegal [7]. Previous studies [8, 9], however, have presented evidence for a sound change in progress by which these combinations emerge in CB, too.

Our experiments test whether listeners of CB compensate less for speech rate than listeners from other German varieties (e.g. German Standard German, Swiss German) in which no such changes are expected. We are obtaining forced-choice judgments to continua spanning resynthesized versions of minimal pair words (e.g. *bitter-Bieter*, *baden-baten*) embedded in fast and slower carrier phrases spoken in the respective variety. The only cue to differentiate the words within any of these continua is the duration of the vowel or consonant, respectively. We will test a total of 136 listeners from three countries.

So far, we have tested a small number of listeners of German (N=7) and Austrian (N=5) CB varieties. Commensurate with Fig. 1, German CB listeners classify stimuli in fast carrier phrases more often as long vowels than in slower carrier phrases, indicating that they compensate for speech rate. Austrian CB listeners, on the other hand, show no such difference, implying that compensation is diminished in this group. These preliminary results suggest that the expected sound change is more active in Austria, where the standard is influenced by CB. The complete set of results will be discussed in light of usage-based theories [10] by which dialect leveling is the result of increased experience with the standard language.

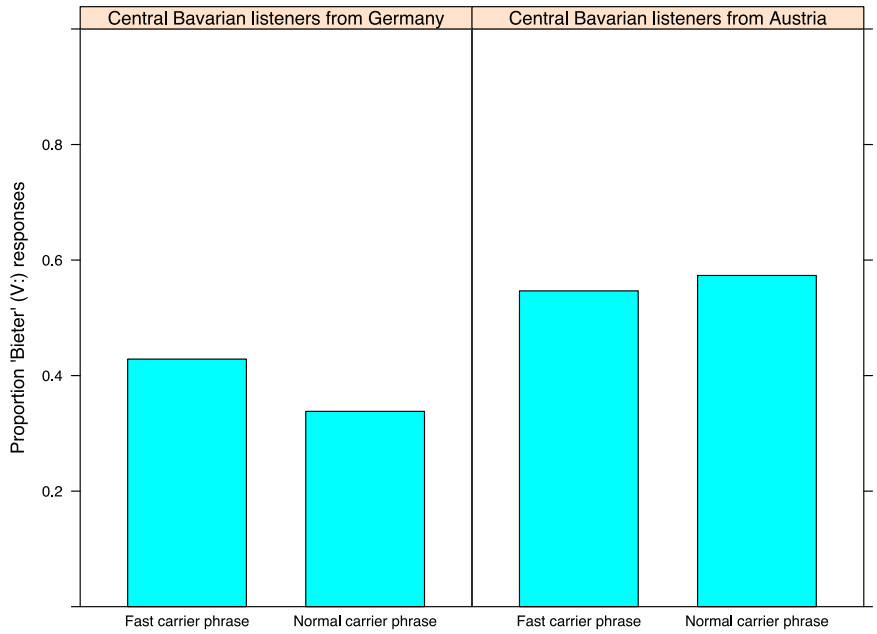


Figure 1: *Responses in categorization tasks (30 trials per listener). Central Bavarian listeners from Germany (N=7) and Austria (N=5) had to categorize tokens as* Bieter *(long vowel) or* bitter *(short vowel). Tokens were embedded in either a fast or a slower carrier phrase. Germans compensated for speech rate, as indicated by the increased proportion of V: responses in the fast condition. Austrians showed no difference between conditions, suggesting that compensation for speech rate was diminished.*

References

[1] K. J. Kohler, "Explanation in phonology: The feature fortis/lenis," *Phonetica*, 41, pp. 150–174, 1984.

[2] J. Harrington, F. Kleber, U. Reubold & J. Siddins, "The implications for sound change of prosodic weakening: evidence from polysyllabic shortening," *Laboratory Phonology*, 6(1), pp. 87–117, 2015.

[3] E. Reinisch, A. Jesse & J. M. McQueen, "Speaking Rate from Proximal and Distal Contexts Is Used during Word Segmentation," *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), pp. 978–996, 2011.

[4] K. J. Kohler, "Dimensions in the perception of fortis and lenis plosives," *Phonetica*, 36, pp. 332–343, 1979.

[5] R. Wiese, *The Phonology of German*. Oxford: Oxford University Press, 1996.

[6] M. Jessen, *Phonetics and Phonology of Tense and Lax Obstruents in German.* Amsterdam & Philadelphia: John Benjamins Publishing, 1998.

[7] R. Bannert, *Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage*. Ph.D. dissertation, Lund University, 1976.

[8] F. Kleber, "Complementary length in vowel-consonant sequences: acoustic and perceptual evidence for a sound change in progress in Bavarian German," *Journal of the International Phonetics Association*, 17(5), pp. 1–22, 2017.

[9] S. Moosmüller & J. Brandstätter, "Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect," *Language Sciences*, 46, pp. 84–95, 2014.

[10] J. Pierrehumbert, "Word-specific phonetics," in *Laboratory phonology VII*, C. Gussenhoven & N. Warner (Eds.), Berlin: Mouton de Gruyter, 2002, pp. 101–140.

Kategoriale Wahrnehmung von Voice Onset Time durch österreichische Hörer

Petra Hödl
Institut für Sprachwissenschaft (Karl-Franzens-Universität Graz)
petra.hoedl@uni-graz.at

Im Standarddeutschen gibt es eine phonologische Opposition zwischen sogenannten stimmhaften (/b, d, g/) und stimmlosen (/p, t, k/) Plosiven. Diese phonologische Opposition wird allerdings vor allem im Anlaut zumeist als Aspirationskontrast und nicht als „echter" Stimmtonkontrast realisiert (vgl. [1], [2]). Somit haben wir es phonetisch gesehen in beiden Fällen mit physiologisch stimmlosen Lauten zu tun, die durch unterschiedlich lange positive Voice Onset Times (VOT) [3] gekennzeichnet sind. Diese phonetische Unterscheidung ist für die Aussprache norddeutscher Sprecher empirisch bestätigt worden (vgl. [1]). Bei Sprechern einiger (südlicher) Varietäten des Deutschen – wie auch des österreichischen Deutsch –kommt es laut Literatur jedoch häufig zu einer Reduzierung der Aspiration (vgl. [2], [4], [5]).

Das Ziel der hier vorgeschlagenen Studie ist es, zu untersuchen, ob diese Tendenz zur reduzierten Aspiration Auswirkungen auf die Wahrnehmung des phonetischen Parameters VOT durch österreichische Hörer hat. Im Speziellen soll die Position und der Schärfegrad der kategorialen Grenze zwischen /b, d, g/ und /p, t, k/ hinsichtlich VOT bei dieser Hörergruppe erhoben werden. Die Annahme lautet, dass, wenn VOT in der Produktion von österreichischen Plosiven ein unzuverlässiges akustisch-auditives Unterscheidungskriterium ist, sich dies auch in der Perzeption widerspiegeln könnte.

Es wurde ein Perzeptionsexperiment durchgeführt, bei dem 33 österreichische Hörer als Testgruppe und 47 deutsche Hörer als Kontrollgruppe semi-manipulierte Wortäußerungen einer österreichischen Muttersprachlerin identifizieren mussten. Als Stimuli wurden die Wörter „backen", „packen", „danken", „tanken", „Gasse" und „Kasse" verwendet, wobei die VOT der Äußerungen in Stufen von 5 Millisekunden akustisch manipuliert wurde. Die 6 so entstandenen Kontinua bestanden aus jeweils 13 Stimuli, die VOT-Werte von 0 bis 60 Millisekunden aufwiesen.

Die vorläufige Auswertung der Daten ergab, dass die Kategoriengrenze bei velaren Stimuli und Stimuli, die auf Basis einer originalen Lenes-Artikulation manipuliert wurden, statistisch signifikant nach rechts verschoben ist. D.h. höhere VOT-Werte sind notwendig, damit Hörer diese Stimuli als stimmlos identifizieren. Die Muttersprache der Hörer (Österreichisch vs. Deutsch) erwies sich im Gegensatz dazu als statistisch nicht-signifikant. Interessanterweise fanden sich allerdings v.a. in der Testgruppe Hörer, die bei einzelnen Kontinua überhaupt keinen kategorialen Sprung in der Identifikationskurve zeigten. Für diese Hörer scheint VOT keine entscheidende Rolle bei der Identifikation der präsentierten Stimuli gespielt zu haben. Dies könnte unter Umständen als perzeptueller Reflex auf die reduzierte Zuverlässigkeit dieses Parameters in der Aussprache gedeutet werden.

Literatur

[1] M. Jessen, *Phonetics and phonology of tense and lax obstruents in German*, Amsterdam/Philadelphia: Benjamins, 1998.

[2] S. Moosmüller & C. Ringen, "Voice and aspiration in Austrian German plosives," *Folia Linguistica*, 38, pp. 43–62, 2004.

[3] L. Lisker & A. S. Abramson, "A cross-language study of voicing in initial stops: acoustical measurements," *Word*, 20, pp. 384–422, 1964.

[4] R. Muhr, "Varietäten des Österreichischen Deutsch," *Revue belge de philologie et d'histoire*, 79, pp. 779–803, 2001.

[5] E.-M. Krech, E. Stock, U. Hirschfeld & L. C. Anders (Hrsg.), *Deutsches Aussprachewörterbuch*, Berlin/New York: de Gruyter, 2009.

Temporal Perturbation of German Quantity Contrasts

Miriam Oschkinat[1], Eva Reinisch[1], Phil Hoole[1]
[1]Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München
[miriamo | evarei | hoole] @phonetik.uni-muenchen.de

People are sensitive to the interplay between acoustic and sensorimotor information during speech production. A tool that has been used to demonstrate this connection is online-altered auditory feedback (OAF). As participants produce speech, their speech signal is filtered and manipulated before being fed back via headphones in close-to-real time. The mismatch between the intended production and perception triggers articulatory compensation. While this phenomenon has been established as a reliable effect for spectral perturbation (e.g., vowel formants [1]), relatively less is known about its effect and underlying mechanism in the temporal domain (i.e., lengthening or shortening of segments).

The present study set out to fill this gap by testing German speakers' reaction to temporal OAF during the production of words containing the vowels /a/ or /a:/, a phoneme contrast that is realized as a quantity contrast without strong additional spectral cues [2]. In a paradigm developed by Cai et al. [3,4] participants were asked to either produce the words *Stab* (/ʃtaːp/) and *Staat* (/ʃtaːt/) in which the vowel was compressed, or *Stamm* (/ʃtam/) and *Stadt* (/ʃtat/) in which the vowel was lengthened. The production of the words was implemented in 4 blocks of 30 trials, with a baseline phase, a ramp phase in which the magnitude of OAF was increased on every trial, a hold phase in which OAF was kept at a maximum of 80% change relative to the produced duration (intended to cross the perceptual /a/-/a:/ boundary), and a reset phase without OAF. Previous studies on OAF in the spectral domain have shown that most participants react by adjusting their production in the "opposite" direction to the manipulation [e.g., 5,6]. However, it is commonly reported that subsets of participants do not react or even follow the direction of the manipulation [e.g., 7]. If the sensorimotor coupling in the temporal domain works similarly, we might expect similar reactions and grouping patterns with temporal OAF.

Data of the first six participants (out of 30; 3 per condition) show clear reactions to the temporal manipulation. Reactions were greater for the compression of /a:/ than lengthening of /a/, with a compensatory lengthening of 40ms and 100ms for two participants and a shortening of 40ms for another participant (following the perturbation). In short-vowel manipulation participants compensated for *Stamm* by shortening the vowel by ~27ms, which indicates a compensatory behavior, but not for *Stadt*. This could be due to the already shorter vowel duration of /a/ in *Stadt* than in *Stamm* in the baseline condition and reflect physiological limits to compression.

This suggests that the auditory manipulation of temporal cues has similar impact on sensorimotor planning and execution of quantity contrasts as the manipulation of spectral cues for phonemic quality contrasts. As such this study contributes to research on how temporal cues are built up and stored.

Literatur

[1] Houde, J.F. & Jordan, M. I. (1998). “Sensorimotor adaptation in speech production”. *Science, 279*, 1213–1216.

[2] Pätzold, M., & Simpson, A. P. (1997). “Acoustic analysis of German vowels in read speech”. In A. P. Simpson, K. J. Kohler, & T. Rettstadt (Eds.), The Kiel corpus of read/spontaneous speech–Acoustic database, processing tools and analysis results, AIPUK 32 (pp. 215–247). Germany: IPDS.

[3] Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). “Weak responses to auditory feedback perturbation during articulation in persons who stutter: evidence for abnormal auditory-motor transformation”. *PloS one, 7*(7), e41830.

[4] Cai, S., Beal, D. S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2014). “Impaired timing adjustments in response to time-varying auditory perturbation during connected speech production in persons who stutter.” *Brain and language, 129*, 24-29.

[5] Brunner, J. (2009). "Perturbed speech".

[6] Mitsuya, T., et al. (2011). “A cross-language study of compensation in response to real-time formant perturbation”. *The Journal of the Acoustical Society of America, 130*(5), 2978-2986.

[7] Klein, E., Brunner, J., & Hoole, P. (2017). “Real-time auditory perturbation of fricative spectra”. *P&P13. Proceedings of the Conference on Phonetics & Phonology in German-speaking countries*, 105-108.

Implicit dialect association and accommodation

Katharina Mittelhammer, Felicitas Kleber[2]
[1,2]LMU München
k.mittelhammer@campus.lmu.de, kleber@phonetik.uni-muenchen.de

The goals of this study were threefold: the first was to quantify acoustically selected dialectal differences between the Austrian Kitzbühel (Tyrol) and the German Tegernsee (Bavaria) region which belong to the same dialect area (Southern-central Bavarian). Following a study by [1], the second and third goals were to measure phonetic accommodation between speakers of the two regions and to combine these results with measurement of the same speakers' bias towards one of the two regions. The intentions of the second and third goals were thus to further our understanding of how speech varies as a result of (simulated) language contact, on the one hand, and regional identity, on the other, both of which have been claimed to play a major role in dialect mixture and the emergence of new varieties (cf. [4]).

Eight female speakers each from the Tegernsee area and the Kitzbühel region participated in a two-part study comprising a speech production and an implicit association task [2]. The speech production task involved in a first step a picture naming task where speakers had to label target words such as /katʂə/ 'cat' and in a second step a shadowing task where participants were asked to identify and repeat the same target words uttered by one female model talker from the other dialect region. The target words used in the speech production task were selected based on the description of dialect differences between the two regions in [5]. The implicit association task measures reaction times in judgments to combined primes of concepts (here representing Austria or Germany) and attributes (e.g. good vs. bad). Two dialect features were examined by means of different acoustic measures: (1) duration of spectral noise between the release and vowel onset to quantify the degree of affrication of the velar plosive (i.e. German Bavarian /kʰ/ vs. Tyrolean /kχ/ and (2) analyses of the spectrum and preceding formant transitions to determine velarization of the palatal fricative (e.g. German Bavarian /bɛçɐ/ vs. Tyrolean /bɛxɐ/).

Region had a significant effect on the production of the examined dialect features – both in the model talkers and participants. For example, and commensurate with Figure 1, speakers from the Kitzbühel area produced longer aspiration than speakers from the Tegernsee area indicating affrication of velar stops in the Tyrolean variety of Kitzbühel. The extent and direction of phonetic accommodation (in form of the difference in distance) in these two features were, however, speaker- and word-specific with a general tendency towards less accommodation than found in [1]. Moreover, implicit attitudes towards dialect regions (which were all pro home region, cf. Fig. 2) did not correlate with the degree of phonetic accommodation. For instance, Tirol speakers T3 and T4 diverged and converged, respectively, despite the similar IAT score. These results suggest that dialect mixture is not necessarily a direct consequence of phonetic accommodation as proposed by [4] since accommodation is variable in that both divergence and convergence can be found on different phonetic levels within the same speaker.

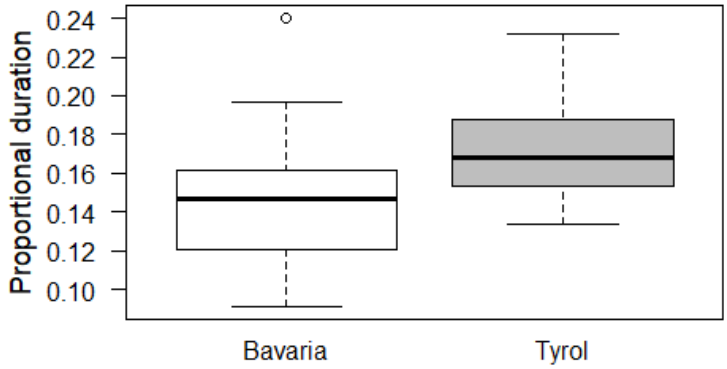


Figure 1: *Proportional duration of aspiration in tokens with an initial velar stop.*

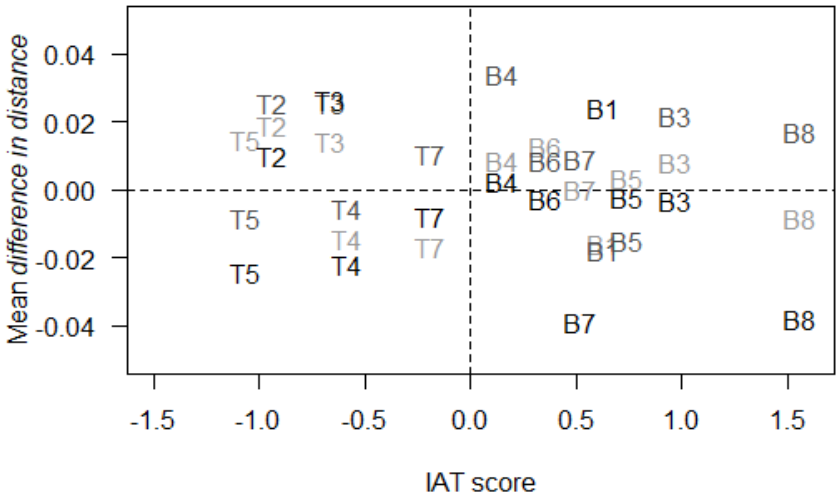


Figure 2: *Correlation between the IAT score and the mean difference in distance for initial velar stops separately for Tyrol (T) and German Bavarian (B) speakers (1-8) and for words (dark = 'Koch', medium grey = 'Katze', and light grey = 'Kasse'). Positive IAT scores indicate a bias towards Germany, negative scores a bias towards Austria. Negative difference-in-distance-values indicate convergence (reduction of distance to the model talker in the acoustic space) and positive values divergence (increasing distance to the model talker).*

Bibliography

[1] Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39* (4), 437-456. doi:10.1017/S0047404510000400

[2] Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74* (6), 1464-1480.

[3] Pardo, J. S. (2013). Measuring phonetic convergence in speech production. Frontiers in psychology, 4.

[4] Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. Language in Society, 37, S. 241–254.

[5] Wiesinger, P. (1990). The Central and Southern Bavarian Dialects in Bavaria and Austria. In Russ, C. V. J. (Hg.), The Dialects of modern German: a linguistic survey, S. 438–519. Routledge, London.

Faktoren für die Variation des *tonal alignment* in Nord- und Südtiroler Dialekten

Laut Erkenntnissen der Dialektologie sind die deutschen Dialekte des österreichischen Bundeslandes Tirol und der angrenzenden Provinz Südtirol in Italien sehr ähnlich (vgl. [1]). Unterschiede wurden vor allem in der Prosodie vermutet (z. B. [2]), wobei diese Hypothese bislang noch nicht empirisch überprüft wurde.

In einem aktuellen Forschungsprojekt wird nun die Intonation dieser Dialekte untersucht. Der Fokus wird dabei auf mögliche Unterschiede in der Realisierung von nuklearen fallenden Konturen – laut autosegmental-metrischer Theorie aus einem steigenden Pitch-Akzent (H* bzw. L+H*) und einem tiefen finalen Grenzton (L%) zusammengesetzt – gelegt. Wie bisherige Untersuchungen gezeigt haben, haben sich für die intonatorische Unterscheidung eng verwandter Varietäten nämlich insbesondere phonetische Differenzen als relevant herausgestellt (vgl. [3], [4]). In diesem Projekt wird unter anderem ermittelt, welche dialektalen Differenzen beim *tonal alignment* (= zeitliche Ausrichtung) von Pitch-Akzenten bestehen und wie das *alignment* durch verschiedene Faktoren auf segmentaler Ebene sowie unterschiedlichen Fokus in der Intonationsphrase beeinflusst wird.

Zu diesem Zweck wurden zwei Experimente durchgeführt. Im ersten Experiment wurden mithilfe im Dialekt verschrifteter Sätze kontrollierte Sprachdaten gewonnen (vgl. [5]), während im zweiten Experiment halbspontane Sprachdaten mittels Bildern evoziert wurden.

Im Beitrag werden Zwischenergebnisse des Projektes präsentiert: Diese zeigen, dass beträchtliche regionale Unterschiede beim *alignment* von Pitch-Akzenten bestehen. In Nordtirol treten Grundfrequenzgipfel spät, in Südtirol teilweise früh in der betonten Silbe auf, woraus deutlich wird, dass die Annahme, dass Grundfrequenzgipfel im Süden des deutschen Sprachraumes spätes *alignment* zeigen (vgl. [6]), nicht für den äußersten Süden gelten kann. Weiters besteht ein signifikanter Zusammenhang zwischen der Silbenstruktur und dem *alignment* des F_0-Gipfels (später in CVC-Silben gegenüber CV:-Silben) sowie zwischen dem Abstand des Pitch-Akzents von der Phrasengrenze und dem *alignment* (je näher der Pitch-Akzent an der Phrasengrenze liegt, desto weiter wird der F_0-Gipfel nach vorne verschoben). Dies deckt sich mit Erkenntnissen früherer Studien (vgl. [7], [8]). Hingegen zeigt die Gegenüberstellung von weitem und kontrastivem Fokus, dass kontrastiver Fokus kaum zu einer Verschiebung des *alignment* nach hinten führt, was den Ergebnissen von Grice et al. [9] für mittel- bzw. norddeutsche SprecherInnen entgegensteht. In Tiroler Dialekten scheint kontrastiver Fokus stattdessen mithilfe anderer Strategien, wie etwa eines größeren Scalings des Pitch-Akzents, gekennzeichnet zu werden.

Literatur

[1] E. Kühebacher, "Zur Lautgeographie von Tirol: Aus den Vorarbeiten zu einem Tiroler Sprachatlas," *Zeitschrift für Mundartforschung*, 29 (2), 1962, pp. 150–168.

[2] F. Lanthaler, "The German Language in South Tyrol – some Sociolinguistic Aspects," in *Aspects of multilingualism in European border regions*, A. Abel, M. Stuflesser & L. Voltmer (Hrsg.), Bozen: Europäische Akademie Bozen, 2007, pp. 220–235.

[3] P. Gilles & B. Siebenhaar, "Areal variation in prosody," in *Language and space*: *Theories and methods*, P. Auer & J. E. Schmidt (Hrsg.), Berlin, New York: De Gruyter Mouton, 2010, pp. 786–804.

[4] J. Peters, "Intonation des Niederdeutschen: Eine Untersuchung zu Weener (Rheiderland)," *Jahrbuch des Vereins für niederdeutsche Sprachforschung*, 133, 2010, pp. 105–140.

[5] J. Peters, J. Hanssen & C. Gussenhoven, "The timing of nuclear falls: Evidence from Dutch, West Frisian, Dutch Low Saxon, German Low Saxon, and High German," *Laboratory Phonology*, 6 (1), 2015, pp. 1–52.

[6] M. Atterer & D. R. Ladd, "On the phonetics and phonology of 'segmental anchoring' of *F0*: evidence from German," *Journal of phonetics*, 32, 2004, pp. 177–197.

[7] D. Mücke, M. Grice, A. Hermes & J. Becker, "Sources of variation in tonal alignment: Evidence from acoustic and kinematic data," *Journal of phonetics*, 37 (3), 2009, pp. 321–338.

[8] T. Rathcke & J. Harrington, "The phonetics and phonology of high and low tones in two falling f0-contours in Standard German," in *Proceedings of INTERSPEECH*, 2007, pp. 982–985.

[9] M. Grice, S. Ritter, H. Niemann & T. B. Roettger, "Integrating the discreteness and continuity of intonational categories," *Journal of phonetics*, 64, 2017, pp. 90–107.

Vowel movement revisited – Simple or complex metrics?

Anja Geumann, Denis Arnold
Institut für Deutsche Sprache, Mannheim
geumann@ids-mannheim.de, arnold@ids-mannheim.de

A number of recent studies, e.g. [1], [3], [6], [7] have suggested that dynamic aspects play a role for differentiating English vowel qualities that are traditionally considered monophthongal in nature. Other studies (e.g. [8]) have argued that, while there are dynamic differences they are not that important for differentiating vowel qualities. Usually, in static phonetic descriptions, the formant values around 50% of vowel duration are considered most representative of target quality. In the dynamic studies mentioned, the movement over time of the vowel formants is studied. As for the beginning and end of a vowel, those are more indicative of the consonantal context than of the identity of the vowel. An investigation over more data revealed 25-75% of the vowel duration as a fairly safe approximation of the actual vowel trajectory. In the following, metrics are listed that were used in an earlier unpublished study to describe the characteristic movement in a vowel, viz. distance between start and end of the actual vowel movement (25-75%) and distance between start and mid of vowel (50%) and distance between mid and end of vowel movement, movement to the front/back: F2 slope, movement upwards/downwards: F1 slope. In this earlier study results for English and German high tense and lax vowels were presented with these metrics. Parts of these metrics were used in [5]. We are using two well established corpora of read speech for American English [2] and Standard German [4] to investigate the vowel dynamics of English and German vowels.

In this presentation we want to broaden the investigation to all nonlow vowels. Results indicate that the dynamic metrics for all nonlow English tense vowels are quite different from English lax vowels and different from German tense vowels.

Additionally, for modeling the dynamic behavior of the vowels a more complex analysis of the dynamic formant data using generalized additive (mixed) models (GA(M)Ms) [9], [10] is used.

References

[1] K. Chládková, S. Hamann, D. Williams & S. Hellmuth, “F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English”, *Language and Speech,* First published date: May-30-2016, https://doi.org/10.1177/0023830916650991, 2016.

[2] J. Garofolo et al., TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

[3] J. M. Hillenbrand, “Static and dynamic approaches to vowel perception”, in *Vowel Inherent Spectral Change*, G. S. Morrison & P. F. Assmann, (Eds.) Berlin, Heidelberg: Springer, 2013.

[4] K. J. Kohler, Ed., Phonetisch-Akustische Datenbasis des Hochdeutschen. Kieler Arbeiten zu den PHONDAT-Projekten 1989-1992. (=Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 26), 1992.

[5] R. Raffelsiefen & A. Geumann, “Diphthongs versus monophthongs in English”, *Proceedings of the conference on phonetics & phonology in German-speaking countries (P&P 13), M.* Belz, S. Fuchs, S. Jannedy, C. Mooshammer, O. Rasskazova & M. Żygis, (Eds.), Berlin: ZAS; Humboldt-Universität zu Berlin, 2018, pp. 157-160.
[6] J. Slifka, “Tense/lax vowel classification using dynamic spectral cues”, *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 2003, pp. 921-924.
[7] W. Strange, W. & O.S. Bohn, “Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies”, *JASA* 1004(1), 1998, pp. 488-504.
[8] C. I. Watson & J. Harrington, “Acoustic evidence for dynamic formant trajectories in Australian English vowels”, *JASA* 106(1), 1999, pp. 458-468.
[9] S. Wood, *Generalized additive models: an introduction with R*. CRC press, 2006.
[10] S. N. Wood, Y. Goude & S. Shaw, “Generalized additive models for large data sets”, *Journal of the Royal Statistical Society: Series C (Applied Statistics), 2014*.

Prominence marking depends on expectations

Christine T. Röhr, Henrik Niemann, Stefan Baumann, Martine Grice
IfL Phonetik, University of Cologne
{christine.roehr; henrik.niemann; stefan.baumann; martine.grice}@uni-koeln.de

In many languages speakers employ prosody to highlight new information, making it more prominent. They also use prosody to attenuate given information, making this information less prominent. Thus, a listener should be able to interpret the level of prominence of a word as indicative of the information status that the speaker is intending to convey. However, both the speaker's marking and the listener's perception of prominence can also be affected by expectations evoked by the context.

For West Germanic languages it has been shown that information status (i.e. newness/givenness in discourse) is marked not only by accent *placement* but also by the level of prominence expressed by different pitch accent *types*. Studies on German ([1], [2], [3], [4]) suggest an inverse relation between discourse givenness and prosodic prominence, i.e. the more accessible a concept in the listener's mind, the lower its prosodic prominence. Results reveal that the pitch movement leading towards the target on the accented syllable ('onglide' [5]) is the most important tonal cue for prominence (rising onglides being perceptually more prominent than falling ones [6]).

This study aims to find out how different types of expectation influence a speaker's choice of prosodic cues. Two discrete pre-contexts for each test sentence (60 items) were designed to trigger expectations about appropriate upcoming information. For example, the pre-context in (1a) builds up an expectation for new (unpredictable) information, whereas the pre-context in (2a) establishes that nothing new is going to follow (predictable). We hypothesise that a prominent accent is appropriate on the noun in (1b), whereas a less prominent accent should be appropriate on the noun in (2b).

Preliminary results generally confirm our hypothesis (Fig.1). In 80% of all test sentences, subjects (10f, 4m) realize the nuclear accent on the noun. After context (1a) they use accents with a rising onglide 91.8% of the time, rarely using ones with a falling onglide (only 8.2%). After context (2a), there is a reduced percentage of accents with a rising onglide (65%), and an increase of accents with a falling onglide (35%). Thus, accent types are distributed differently, depending on the prior context.

Whilst an ERP study [7] has already shown that information status cued by pitch accent type is processed in real-time, the effect of context has not yet been explored. The production results will serve as a basis for a follow-up ERP study to disentangle expectation-based vs. signal-driven aspects of perceptual prominence in neurocognitive information processing.

(1) (a) *Rate mal, was uns heute passiert ist!*
'Guess what happened to us today!

(b) *Wir haben Milena getroffen.*
We met Milena.'

(2) (a) *Heute ist nichts Besonderes passiert.*
'Today, nothing special happened.

(b) *Wir haben Milena getroffen.*
We met Milena.'

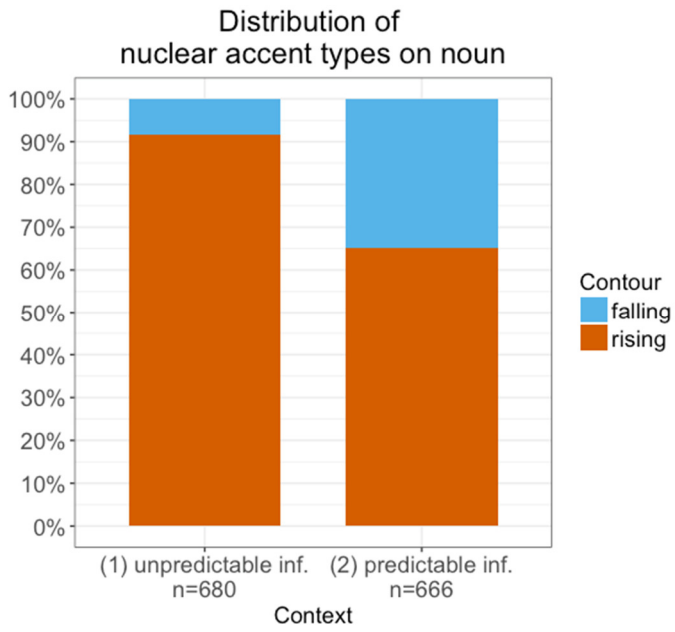


Figure 1: *Relative distribution of nuclear accent types on the nouns in the test sentences plotted against their respective pre-context.*

References

[1] S. Baumann & M. Grice, "The intonation of accessibility," *Journal of Pragmatics*, 38(10), pp. 1636–1657, 2006.

[2] S. Baumann, C. T. Röhr & M. Grice, "Prosodische (De-)Kodierung des Informationsstatus im Deutschen," *Zeitschrift für Sprachwissenschaft*, 34(1), pp. 1-42, 2015.

[3] C. T. Röhr & S. Baumann, "Decoding Information Status by Type and Position of Accent in German," *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, pp. 1706-1709, 2011.

[4] C. T. Röhr & S. Baumann, "Prosodic Marking of Information Status in German," *Proceedings of the 5th International Conference on Speech Prosody*, 100019, Chicago, USA, pp. 1-4, 2010.

[5] S. Ritter & M. Grice, "The Role of Tonal Onglides in German Nuclear Pitch Accents," *Language and Speech*, 58(1), pp. 114-128, 2015.

[6] S. Baumann & C. T. Röhr, "The Perceptual Prominence of Pitch Accent Types in German," in *Proceedings of the 18th International Congress of Phonetic Sciences*, The Scottish Consortium for ICPhS 2015 (ed.), Glasgow, UK: the University of Glasgow, Paper No. 298, 2015, pp. 1-5.

[7] P. B. Schumacher & S. Baumann, "Pitch accent type affects the N400 during referential processing," *NeuroReport*, 21(9), pp. 618–622, 2010.

Do phonologists know too much?
A regular, simple, unnatural pattern in English

John Harris[1], Kevin Tang[2], Nick Neasom[1]
[1]Department of Linguistics, University College London
[2]Department of Linguistics, Zhejiang University
john.harris@ucl.ac.uk, linguist@kevintang.org, nicholas.neasom.10@gmail.com

There is a well-known phonotactic pattern in English that restricts consonants following /aw/ to coronals (hence **mouth, sound, powder**, but nothing along the lines of */awb/, */awk/). Phonologists might know about the pattern (call it 'awT'), but is it internalised by native speakers? We investigate this question with an auditory acceptability task with nonwords containing /aw/ followed by different consonants. The awT pattern is phonologically regular, structurally simple, and lexically general. All of these factors are known to promote the ability of speakers to learn phonotactic patterns [2]. It has been claimed that awT is also phonetically natural [1, 4], another factor known to enhance learnability [3]. We present comparative and historical evidence showing that the pattern is in fact not natural but is rather the synchronically accidental outcome of a series of largely unrelated sound changes.

One indication that awT is not natural is that it has no parallel in other Germanic languages. The cognate vowel is readily followed by non-coronals in German (e.g. **taub, Baum, schaukeln, Auge**) and Scots (e.g. **cowp** 'tip over', **bowk** 'vomit'). Moreover, proper names from these languages and other sources can be adapted into English with awT-defying results (e.g. **Baum, Smaug**). Another sign of awT's unnaturalness is that various recent sound changes in English have ridden roughshod over it. For example, *l*-vocalisation produces /awk/ in **Malcom, calculate**, while TH-fronting produces /awf/ in **mouth**. The historical origins of awT lie mainly in a series of sound changes that shortened historically long /u:/ before it shifted to /aw/. The result was to syphon off words that would otherwise have contained /aw/ plus non-coronal in the present day, e.g. **plum** (cf. German **Pflaume**), **shovel** (cf. **Schaufel**), **suck** (cf. **saugen**). Furthermore, we found quantifiable differences between German and English VV+C phonotactics and they undermine any attempt to provide natural/formal explanations.

Naturalness in phonological patterns is often observed to correlate with regularity and structural simplicity. To tease out the relative contribution of these factors to phonological learnability, we need to investigate patterns where they can be decoupled. Our non-word acceptability study of awT is a step in that direction, allowing us to investigate a pattern that is at once regular, simple, but not natural.

The results of the study indicate that, if speakers have any tacit awareness of awT at all, it is not encapsulated in anything like a phonologist's rule or constraint. Where a coronal preference can be detected with /aw/, it is no different than what can be observed with other long vowels, which are not restricted to coronals, investigated in the study. Moreover, the preference is influenced by lexical neighbourhood factors, which suggests that participants were making on-the-fly judgements of how much the non-words resemble real words. We conclude that awT is a case where phonologists know more about a phonotactic pattern than speakers know.

References

[1] Kubozono, H. (2001). On the markedness of diphthongs. *Kobe Papers in Linguistics* 3. 60-73.

[2] Moreton, E. & J. Pater (2012a). Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass* 6. 686-701.

[3] Moreton, E. & J. Pater (2012b). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass* 6. 702-718.

[4] Nevins, A. (2012). Vowel lenition and fortition in Brazilian Portuguese. *Letras de Hoje* 47. 228-233.

POSTER

Rich intonational patterns are more advantageous than a lexical pitch accent in storing (non-)linguistic pitch

Autor Yuki Asano[1], Holger Mitterer[2]
[1]University of Tübingen
[2]University of Malta
yuki.asano@uni-tuebingen.de, holger.mitterer@um.edu.mt

The ability to store prosodic information is modulated by the use of linguistic prosodic contrast in one's first language (L1) [1, 2]. The present work investigates whether this L1-dependent ability can be transferred to non-linguistic pitch contrast. Given the findings that the brain mechanisms governing language and music processing interact with each other and share an important link with respect to their underlying processing [3, 4], we expect not to find differences between linguistic and non-linguistic pitch within L1, instead to only find L1-differences across the pitch conditions.

We tested 24 of each Chinese, Japanese and Germans, who differ in the use of linguistic pitch contrasts (Chinese exhibits four types of lexical tones, to which each syllable is obligatory assigned; Japanese has a culminative falling lexical pitch accent based on the accentual system [5]; German primarily uses rich patterns of pitch contrasts at the intonational level [6, 7, 8]). An online adaptive version of the working memory recall task [9, 10, 11] presented sequences of between 2-9 stimuli (high-low or low-high pitch). The task was to reproduce each sequence by typing the associated keys in the correct order. In the *Word* condition, Chinese and Japanese listened to an L1 lexical contrast and Germans to a German intonation contrast. In the *Tone* condition, all participants listened to a contrast of non-linguistic complex tones. The *Segment* condition presenting *numi* vs. *numu* served as a control.

The mean achieved length of the last three stimuli was extracted for each condition and participant, see Figure1. The between-condition analysis using LMER showed that the three groups did not differ from each other in the *Segment* condition (overall $p > 0.3$). In the *Word* and *Tone* conditions, Chinese outperformed Germans, followed by Japanese ($p < 0.05$ in the *Word* condition between all comparisons; Chinese vs. Japanese/Chinese vs. Germans, both $p < 0.01$, Japanese vs. Germans $p = 0.05$ in the *Tone* condition). The between-language analysis showed that Chinese reached higher stimulus numbers in the *Word* and *Tone* conditions than in the *Segment* condition (*Segment* vs. *Word*/*Tone*, $p < 0.01$, but *Word* vs. *Tone*, $p = 1.0$). Japanese reached lower numbers in the *Word* and *Tone* conditions than in the *Segment* (*Segment* vs. *Word*/*Tone*, $p < 0.01$, but *Word* vs. *Tone*, $p = 1.0$). Germans did not show any differences across conditions (overall $p > 0.1$). There were no differences between the linguistic and non-linguistic pitch conditions in any L1 groups. Interestingly, the use of lexical pitch contrasts in the tonal system was more advantageous than that in the accentual system [12] and rich patterns of intonational pitch contrasts was more advantageous than a limited pattern of lexical pitch contrast in storing pitch information [1, 2].

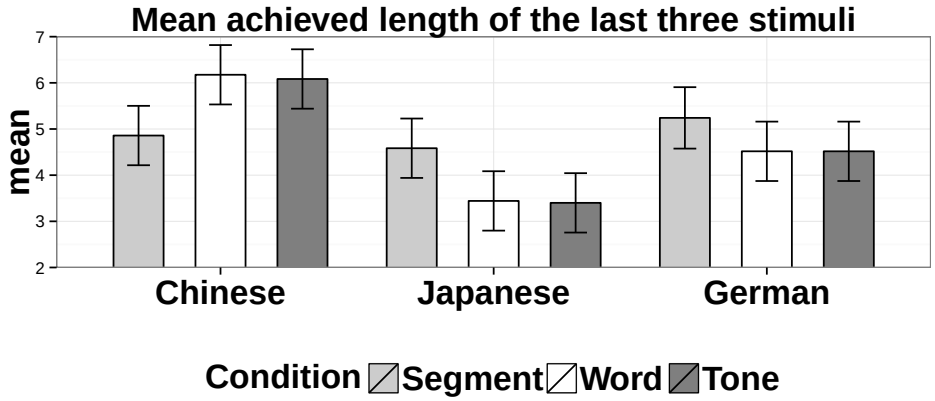


Abbildung 1: *Mean achieved lengths of the last three stimuli and 95% CI bars for each condition and L1.*

Literatur

[1] Y. Asano, "Localising foreign accents in speech perception, storage and production," Ph.D. dissertation, University of Konstanz, Konstanz, Germany, 2016.

[2] B. Braun, T. Galts, and B. Kabak, "Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation," *Second Language Research*, vol. 30, no. 3, pp. 323–350, 2014.

[3] G. M. Bidelman, S. Hutka, and S. Moreno, "Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music," *PLOS ONE*, vol. 8, no. 4, 04 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0060676

[4] J. R. Iversen, A. D. Patel, and K. Ohgushi, "Perception of rhythmic grouping depends on auditory experience," *Journal of Acoustical Society of America*, vol. 124, pp. 2263–2271, 2008.

[5] L. M. Hyman, "How (not) to do phonological typology: The case of pitch-accent," *Berkeley Phonology Lab Annual Report*, pp. 654–685, 2007.

[6] C. Féry, *German Intonational Patterns*. Tübingen: Niemeyer, 1993.

[7] D. Gibbon, "Intonation in German," in *Intonation systems: a survey of twenty languages*, D. Hirst and A. Cristo, Eds. Cambridge University Press, 1998, ch. 4, pp. 78–95.

[8] M. Grice and S. Baumann, *An introduction to intonation - functions and models*. Berlin: Mouton de Gruyter, 2007, pp. 25–32.

[9] P. C. S.-G. N. Dupoux, E. and J. Mehler, "A destressing deafness in French," *Journal of Memory and Language*, vol. 36, pp. 165–179, 1997.

[10] E. Dupoux, Y. Kakehi, K. Hirose, C. Pallier, and J. Mehler, "Epenthetic vowels in Japanese: A perceptual illusion?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, pp. 1568–1578, 1999.

[11] S. Peperkamp and E. Dupoux, "A typological study of stress 'deafness'," *Laboratory Phonology*, vol. 7, pp. 203–240, 2002.

[12] Y. Asano and S. Yuen, "Asymmetric representations of lexical pitch," in *Abstraction, Diversity, and Speech Dynamics*, 2017.

Adaptive Hesitation Insertion Model for Incremental Speech Synthesis (Demo)

Simon Betz[1,2,4], Birte Carlmeyer[1,3,4] , Petra Wagner[1,2], Britta Wrede[1,3]

[1]CITEC (Excellence Cluster Cognitive Interaction Technology), Universität Bielefeld
[2]AG Phonetik & Phonologie, Universität Bielefeld
[3]AG Angewandte Informatik, Universität Bielefeld
[4]Dialogue Systems Group, Universität Bielefeld

simon.betz@uni-bielefeld.de

Synthetic speech is applied in various fields, and it has entered the realm of everyday life, be it in public transportation announcements, telephone customer services, smartphone speech output, or smart-home environments. This synthetic speech output, as opposed to human speech, does not contain the typical markers of human conversation, such as hesitations or reformulations. Rather, the speech is planned and delivered within a single chunk containing all information the user requested in one utterance. We argue that hesitations serve an important role in human-machine communication in the same way they do in human-human communication, namely allowing extra time for the speaker and establishing an interaction between speaker and listener. Based on studies of human hesitation behaviour, we developed and tested a model for inserting hesitations into ongoing speech synthesis output [1].

During this conference we present a live demo of this model implemented in an incremental spoken dialogue system. As can be seen in Figure 1, the model is capable of inserting a cascade of hesitation phenomena for as long as an event of hesitation is active. What this event is and what to do if the extra time granted by the model does not suffice to resolve the event of hesitation is governed by the design of the system. In [1] the event was triggered when users shifted their gaze away from the virtual agent and the continuation strategy was to wait until users refocused. In this demo, the user can start and end hesitation mode by clicking a button. The continuation strategy is to wait for hesitation mode to end, but with a pre-defined threshold, after which the original speech plan will be resumed to avoid cases of ever-lasting awkward silence.

The core part and starting point of the hesitation cascade is lengthening. Previous studies suggest that lengthening is the most subtle of hesitations, that listeners (and even trained annotators) miss frequently [2]. It is thus a promising element to allow extra time without the user noticing, or, at least, without detrimenting perceived speech synthesis quality. The placement and duration of all hesitation elements follow the patterns observed in human communication. The model automatically searches for the best entry point to apply lengthening. If at any time during the cascade hesitation mode ends, the original speech plan will be resumed immediately. For examples of hesitation cascades of various extents see figure 2.

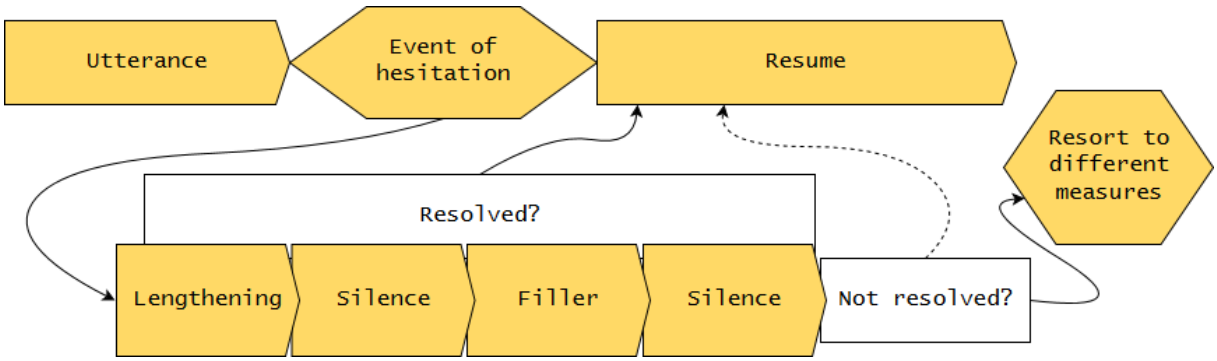


Figure 1: *Hesitation insertion model.*

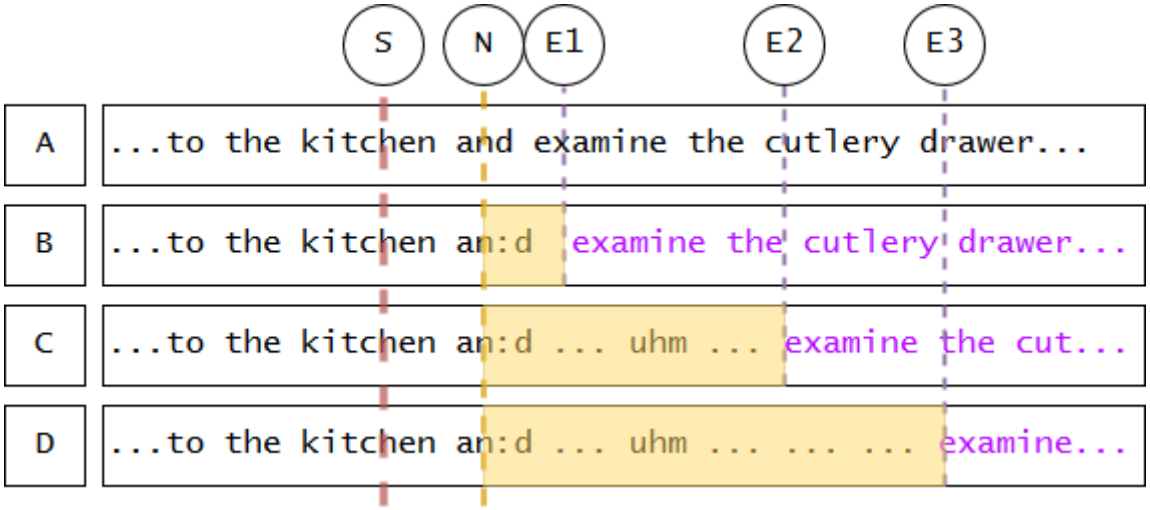


Figure 2: *Examples of insertable hesitations. S = Starting point (external event triggering hesitation mode); N = Entry point (best upcoming target segment, in this case nasal sound in function word); E1–E3 = End points (external events triggering end of hesitation mode, after which originally intended utterance A is resumed); B = utterance with short hesitation, resumption (in purple) starting at E1; C = longer hesitation with lengthening, silence, filler, silence; D = same as C, but with a long second silence (when the strategy is "after the loop, remain silent until an external event triggers the end of hesitation mode".) Hesitation mode intervals are highlighted in yellow.*

Literatur

[1] Betz, Simon, Carlmeyer, Birte, Wagner, Petra, and Wrede, Britta. "Interactive Hesitation Synthesis: Modelling and Evaluation". *Multimodal Technologies and Interaction* 2.1 (2018): 9.

[2] Betz, Simon, Voße, Jana, Zarrieß, Sina, and Wagner, Petra. "Increasing Recall of Lengthening Detection via Semi-Automatic Classification". *Proceedings of Interspeech*. 2017. 1084-1088.

/’zama/ = <sag mal>?
Perzeption phonetisch ambiger Reduktionsformen

Fabian Brackhane
Institut für Deutsche Sprache, Mannheim
brackhane@ids-mannheim.de

Noch stärker als geschriebene Sprache sind gesprochensprachliche Äußerungen geprägt von Reduktionsprozessen, die auf verschiedenen phonetisch-phonologischen Ebenen ablaufen. Allen solchen Mechanismen gemeinsam ist das Ziel eines möglichst ökonomischen Sprachgebrauchs. In diesem Zusammenhang entstehen auch Reduktionsformen, die bisweilen – für sich gestellt – segmentalphonetisch nicht mehr eindeutig einem Ursprungslexem zugeordnet werden können. Dies ist jedoch auch wortübergreifend zu beobachten, in besonderem Maße bei pragmatikalisierten Mehrworteinheiten (PME) wie <was weiß ich> oder <ich sage mal>, die phonetisch bis hin zu Realisierungen wie /zɪç/ resp. /’zama/ reduziert werden können, ohne dass ihr semantischer Gehalt nachweislich beeinträchtigt wird.

Das Korpus FOLK [1] enthält 91 eindeutige Belege für die PME <ich sage mal> und 154 für <sagen wir mal>. Hinzu kommen 46 Belege phonetisch ambig als /’zama/ realisierter Äußerungsbestandteile, von denen allerdings 28 während des Annotationsprozesses durch die Bearbeiter zu den mutmaßlichen Grundformen <ich sage mal> bzw. <sagen wir mal> ergänzt wurden, ohne dass für diese Entscheidung segmentalphonetische Indizien vorlägen. Die vorzustellende Studie untersucht diese Instanzen aus perzeptionsphonetischer Sicht und fragt nach den Gründen, die die Transkribenten veranlasst haben können, die akustisch ambige Formen der einen oder eben der anderen orthografischen Vollform zuzurechnen.

Als Hypothese wird angenommen, dass Hörer – vor allem – eines Mindestmaßes an gesprochenem Kontext bedürfen, um phonetisch stark reduzierte Lexeme oder gar phrasale Konstruktionen zuverlässig rekonstruieren zu können. Dabei steigt die Wahrscheinlichkeit einer sicheren Perzeption mit der Menge an Kontext.

Hierzu sollen unvoreingenommene Hörer in einem Perzeptionsexperiment die betreffenden /’zama/-Belege einer der beiden Grundformen <ich sage mal> oder <sagen wir mal> zuordnen (forced choice). Mithilfe des Frameworks »Percy« [2] wurde ein Online-Perzeptionsexperiment erstellt, das die Instanzen in randomisierter Reihenfolge und drei Kontexten enthält: Isoliert, mit einem Wort linkem Kontext und im Zusammenhang der vollständigen jeweiligen Intonationsphrase. Jeder Stimulus wird dreimal präsentiert; wiederholtes Anhören ist je Stimulus einmal möglich.

Literatur

[1] Th. Schmidt *DGD – Die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim.* Zeitschrift für Germanistische Linguistik 45(3), 2017, S. 451–463. (http://hdl.handle.net/10932/00-03BC-737F-0F33-A501-8)

[2] Chr. Draxler *Percy – an HTML 5 framework for media rich web experiments on mobile devices.* Proc. 12th Interspeech, Florence, S. 3339–3340.

The new kid on the block – Zur Rolle des Datenarchivars bei Sprachdatensammlungen

Christoph Draxler[2], Jürgen Trouvain[2]
[1]Institut für Phonetik und Sprachverarbeitung, LMU München
[2] Sprachwissenschaft und Sprachtechnologie, Universität des Saarlandes, Saarbrücken
trouvain@coli.uni-saarland.de, draxler@phonetik.uni-muenchen.de

Sprachdatensammlungen unterliegen mindestens zwei rechtlichen Regelungen: a) dem Urheberrecht [1] und b) dem Datenschutzrecht [2]. Bislang beschränkte sich die Diskussion dieser rechtlichen Aspekte im Wesentlichen auf den Datenspender, den Datenerheber und eventuelle weitere Nutzer der Daten [3]. Durch den Wunsch nach dauerhafter Speicherung und langfristiger Verfügbarkeit der Daten und Korpora, z. B. in Repositories, kommt ein weiterer Akteur oder Tätigkeitsbereich hinzu: der *Datenarchivar*, der sich spätestens nach Projekt- oder Studienende um Datensammlungen kümmert.

Der Datenerheber überlässt dem Datenarchivar eine Datensammlung, die in der Regel aus Primär- und Metadaten sowie Annotationen und Auswertungen besteht. Dazu vereinbaren beide einen Überlassungsvertrag. Der Datenarchivar implementiert die technische Verfügbarkeit der Datensammlung und verwaltet den weiteren Zugriff auf diese Daten, z. B. indem er interessierten Dritten die Nutzung der Daten für eigene wissenschaftliche Untersuchungen gestattet. Dazu schließen Datenerheber und -archivar idealerweise eine Nutzungsvereinbarung ab. Diese Nutzungsvereinbarung regelt, wer unter welchen Bedingungen und für welche Zeit Zugriff auf die Daten hat. Sie regelt auch den Übergang der Pflichten des Datenerhebers in Bezug auf etwaige Wünsche des Datenspenders nach Änderung oder Löschung der Daten.

Abbildung 1 ist eine schematische Darstellung der Akteure, der Beziehungen untereinander sowie der Dokumente und der beteiligten Institutionen. Dazu gehören außer dem Datenspender, Datenerheber, Datenarchivar und den Sekundärnutzern auch die Rechtsabteilungen und/oder Datenschutzbeauftragten der Forschungsinstitution sowie Ethikkommissionen der Fachverbände bzw. der Forschungs-einrichtungen.

Darüber hinaus kümmert sich der Datenarchivar um Aspekte der Benutzbarkeit der zu pflegenden Korpora, um verwaiste Datensammlungen sowie um die technische Aufbereitung und Konvertierung von nicht mehr aktuellen Formaten, z.B. die Digitalisierung von analogen Tonträger-Sammlungen.

Die Trennung von Datenerheber und Archivar sowie die dazugehörige Festlegung ihrer jeweiligen Aufgabenbereiche erfüllt die Anforderungen von Förderinstitutionen nach einem nachhaltigen Datenmanagement. Im Idealfall hat jede datenerhebende Einrichtung eigenes Personal, das die Rolle des Archivars wahrnehmen kann. Alternativ muss sichergestellt sein, dass Korpora nach Projektabschluss an eine Einrichtung mit Datenarchivar übergeben werden. Ansonsten droht die Gefahr, dass mit großem Aufwand erstellte Datensammlungen verloren gehen.

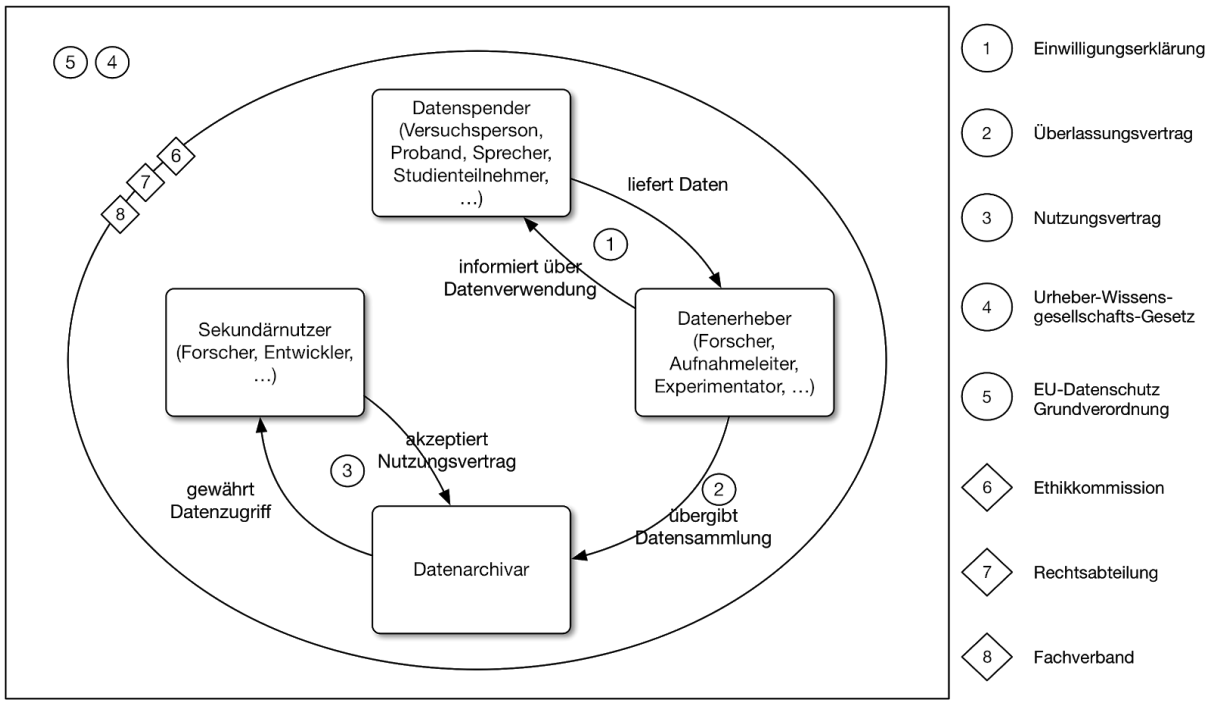


Abbildung 1: *Schematische Darstellung der Akteure und ihrer Beziehungen bei der Erstellung und Distribution von Sprachdatenbanken*

Literatur

[1] UrheberG (2018) https://de.wikipedia.org/wiki/Urheberrecht_(Deutschland)

[2] *DGSVO* (2018) http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.DEU&toc=OJ:L:2016:119:TOC

[3] Christoph Draxler, Jürgen Trouvain (2017) *Rechtliche Grundlagen von Sprach- und Videodatenbanken – zwei Ansätze*, Poster auf der Tagung “Phonetik und Phonologie”, Berlin

F_0 adaptation in conversation: the effect of sex

Kerstin Endes
English and American Studies
University of Graz

Monday, 28 May 2018

Previous studies on fundamental frequency (F_0) in conversation have found that a speaker´s F_0 changes depending on the sex of their conversation partner. For example, [1] investigated the perception of social behavior in dialogue, and the linkage between emotions and prosodic and acoustic cues in speakers has been found in several studies [2] [3]. Based on [1]´s findings, which reported significant differences between speakers depending on sex, for perceived attempts to be liked, likeability and speech planning, the present study aimed to establish whether female pitch range is correlated with the sex of their interlocutor. In [1]´s study, female speakers were found to raise their pitch level when perceived as attempting to be liked by their male interlocutor, while pitch was significantly lowered in conversation with other females, with an increase in intensity. We studied the F_0 of nine female native-speakers of English, measured in Hz, in two conditions, i.e. in a conversation with a male and a female interlocutor, respectively, in order to observe potential changes in F_0. The subjects were asked to engage in a conversation with a male and then a female speaker, previously unknown to them, using a set of questions related to a map task. The target words were then manually extracted from the speech samples, using PRAAT [4], with which the F_0 values were obtained. Counter to the expectations, female F_0 range was not significantly broader when communicating with a male speaker as compared to a female.

The scope of our study did not allow us to control strictly for compounds such as age, class and dialectal variation. Nevertheless, we believe that the results have important consequences for accommodation theory, illustrating the potential effect language competence can have on the interlocutor´s fundamental frequency.

key words: sociophonetics, dialectal variation, accommodation theory

Literature

[1] S. Benus, A. Gravano, J. Hirschberg, R. Levitan, A. Nenkova, L. Wilson, Acoustic and Prosodic Correlates of Social Behavior, Paper presented at the "Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies" (28-31 August 2011). Florence, Italy.

[2] S. Biersack, V Kempe, L. Knapton, Fine-Tuning Speech Register: A comparison of the prosodic features of child-directed and foreigner-directed speech, Paper presented at a conference "9th European Conference on Speech Communication and Technology" (5-11 May 2007). Lisbon, Portugal.

[3] M. Knoll and L. Scharrer, Acoustic and affective comparison of natural and imaginary infant,- foreigner- and adult-directed speech, Paper presented at a conference "9th European Conference on Speech Communication and Technology". (5-11 May 2007). Lisbon, Portugal.

[4] P. Boersma and D. Weenink (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/

Abstract für die P&P 14 – Wien, 6.-7. September 2018

David Eppensteiner[1]
[1]Universität Wien
David.Eppensteiner@outlook.com

Lieben oder leben? Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Sprecher/innen mit L1 Japanisch

Die Wahrnehmung neuer Phonemkontraste stellt mitunter ein Problem für L2-Lernende dar [1]. Ziel dieser Untersuchung ist es, herauszufinden, ob und wie gut erwachsene japanische DaF-Lernende den /i:/-/e:/-Vokalkontrast unterscheiden können, welcher im Japanischen nicht kontrastiv ist. Dazu werden zwei Gruppen japanischer Studierender getestet, die sich hinsichtlich Deutschkenntnissen bzw. Kontaktdauer mit der L2 Deutsch unterscheiden. Durch den Vergleich beider Gruppen soll die Relevanz des nur schwer quantifizierbaren Faktors Spracherfahrung für diesen Kontrast bestimmt werden. Die Vorhersagen zweier phonetischer Modelle – des Perceptual Assimilation Models (PAM bzw. die Weiterentwicklung PAM-L2) von Best und Tyler [2] und des Speech Learning Models (SLM) von Flege [3] – sollen dabei überprüft werden. Aufgrund der inter- und intralingualen Ähnlichkeit des Vokalkontrasts und der Ergebnisse vorangegangener Studien [1] wurde von einem Category-Goodness Difference ausgegangen, wonach Hörer/innen mit L1 Japanisch voraussichtlich leichte Schwierigkeiten haben, die beiden Vokale zu diskriminieren (Diskriminationsrate < 90%). Dem PAM-L2 zufolge ist die Diskrimination der Gruppe der japanischen Student/inn/en mit Spracherfahrung voraussichtlich signifikant besser als jene ohne sprachlicher Erfahrung in der L2 Deutsch und voraussichtlich annähernd „native-like“.

Die Diskrimination des deutschen /i:/-/e:/-Vokalkontrasts durch Sprecher/innen mit L1 Japanisch wird in einem AXB-Diskriminationsexperiment untersucht. Dazu werden zwei Gruppen japanischer Studierender zwischen 18 und 22 Jahren der Dokkyo Universität in Saitama getestet. Gruppe A besteht aus 10 Studierenden ohne Deutschkenntnisse, Gruppe B aus 17 japanischen Deutschstudent/inn/en, die seit mindestens 2 Jahren Deutsch lernen. Als Stimuli fungieren 24 Minimalpaare (gemischt mit anderen irrelevanten Vokalkontrasten), die bis auf den betreffenden Vokal identisch sind (z.B. *lieben* - *leben*). Um die Gleichheit der Silbenränder zu gewährleisten, wurde der „i-Vokal“ in das „e-Wort“ hineingeschnitten (sog. splicing).

Zur Untersuchung der Fragen wird ein MFC-Wahrnehmungsexperiment in Praat konzipiert. Außerdem werden insgesamt drei Schwierigkeitsstufen (mit und ohne Störgeräusch) eingefügt, um die Bedingungen des Hörens zu erschweren und möglicherweise größere Unterschiede herausarbeiten zu können. Als Störgeräusch wird ein Speech Shaped Noise (SSN) in unterschiedlicher Stärke (0dB SNR und -5dB SNR) verwendet, das mithilfe eines Praatskripts aus einer vielstimmigen Aufnahme eines Stimmengewirrs in einer Bar erzeugt wurde. Die Studie versteht sich als Beitrag zum Verständnis der Vokalwahrnehmung in der Zweit- bzw. Fremdsprache.

Literatur

[1] R. Mazuka, M. Hasegawa & S. Tsuji, „Development of Non-Native Vowel Discrimination: Improvement Without Exposure," *Developmental Psychobiology*, February 2014, Vol. 56(2), pp. 192-209, 2014.

[2] C. T. Best & M. D. Tyler, „Nonnative and second-language speech perception: Commonalities and complementarities," in Second language speech learning: The role of language experience in speech perception and production, M. J. Munro & O.-S. Bohn (Hrsg.), Amsterdam: John Benjamins, 2007, pp. 13-34.

[3] J. E. Flege, "Language contact in bilingualism: Phonetic system interactions," in *Laboratory Phonology 9*, J. Cole & J. I. Hualde (Hrsg.), Berlin, New York: Mouton de Gruyter, 2007, pp. 353–381.

Phonetic Accommodation in HCI: Introducing a Wizard-of-Oz Experiment

Iona Gessinger[1,2], Eran Raveh[1,2], Bernd Möbius[1], Ingmar Steiner[1-3]
[1]Language Science and Technology, Saarland University
[2]Multimodal Computing and Interaction, Saarland University
[3]German Research Center for Artificial Intelligence (DFKI GmbH), Saarbrücken
gessinger@coli.uni-saarland.de

The phenomenon of phonetic accommodation in human communication has been a topic of interest to the research community for a long time. It has been shown that human interlocutors converge to, or, under specific circumstances, diverge from each other with respect to suprasegmentalia and segmentalia. As spoken dialogue systems (SDSs) are developing and we are encountering an increasing number of non-human interlocutors in our daily life, the question arises whether phonetic accommodation occurs in human-computer interaction (HCI) as well.

Prior experimentation has shown that humans exhibit phonetically accommodative behavior when shadowing synthetic stimuli [1], [2]. This means that one precondition for accommodation in HCI is fulfilled, namely that humans are responsive to phonetic variation in synthetic voices. However, shadowing speech, i.e., repeating after hearing, is a far more static way to interact with a synthetic voice than an actual dialogue. Therefore, it is desirable to extend the previous experimental approach to a more dynamic exchange between system and user in order to better understand the behavior of the latter in HCI.

Up to now, SDSs themselves are not phonetically responsive to the user input. Suggestions for models that shall enable the computer to show phonetically accommodative behavior are being developed [3], [4], yet there is no working system which could be employed to study the user side. We hence apply the Wizard-of-Oz (WoZ) method to simulate an intelligent SDS as it is good practice in HCI research [5]–[8]. While the user believes to interact with an autonomous system, it is in fact the wizard, i.e., the experimenter, who decides about the system's responses.

The present study is embedded in a language learning scenario with target language German. This allows for the simulation of a realistic use case as it may occur in the context of computer-assisted language learning. Two user groups will participate in the experiment: (1) German L1; (2) French L1 and German L2. The dialogue is task-oriented and supported by visual feedback. The wizard chooses responses from a database of pre-synthesized utterances. During the dialogue, the wizard voice systematically varies certain phonetic features, such as intonation pattern, segmental pronunciation, and speaking rate. We are assuming that users will phonetically accommodate to the system to different degrees depending on their L1 and their general responsiveness which may be grounded in different personality traits (e.g., permeability of ego boundaries), the attitude toward the non-human interlocutor, and overall phonetic talent [9]–[11].

This is work in progress. Currently, we are implementing the WoZ framework that will be used in the experiment.

References

[1] I. Gessinger, E. Raveh, S. Le Maguer, B. Möbius, and I. Steiner, "Shadowing synthesized speech – Segmental analysis of phonetic convergence", in *Interspeech*, 2017, pp. 3797–3801.

[2] I. Gessinger, A. Schweitzer, B. Andreeva, E. Raveh, B. Möbius, and I. Steiner, "Convergence of pitch accents in a shadowing task", in *Speech Prosody*, 2018, in press.

[3] R. Levitan, Š. Beňuš, R. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, "Implementing acoustic-prosodic entrainment in a conversational avatar", in *Interspeech*, 2016, pp. 1166–1170.

[4] E. Raveh, I. Steiner, and B. Möbius, "A computational model for phonetically responsive spoken dialogue systems", in *Interspeech*, Stockholm, 2017, pp. 884–888.

[5] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz studies – why and how", *Knowledge-based systems*, vol. 6, no. 4, pp. 258–266, 1993.

[6] L. Bell, J. Gustafson, and M. Heldner, "Prosodic Adaptation in Human-Computer Interaction", in *ICPHS*, Barcelona, 2003, pp. 833–836.

[7] S. Oviatt, C. Darves, and R. Coulston, "Toward adaptive conversational interfaces: Modeling speech convergence with animated personas", *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 11, no. 3, pp. 300–328, 2004.

[8] D. DeVault, J. Mell, and J. Gratch, "Toward natural turn-taking in a virtual human negotiation agent", in *AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction*, 2015.

[9] A. Więckowska, "Ego boundaries as determinants of success in foreign language learning – A state-of-the-art perspective", in *Acta Universitatis Wratslaviensis*, ser. Anglica Wratslaviensia XLIX, 2011, pp. 199–208.

[10] A. Yu, C. Abrego-Collier, and M. Sonderegger, "Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and »autistic« traits", *PloS ONE*, vol. 8, no. 9, e74746, 2013.

[11] N. Lewandowski, "Sociolinguistic factors in language proficiency: Phonetic convergence as a signature of pronunciation talent", in *Language Talent & Brain Activity*, G. Dogil and S. Reiterer, Eds., Berlin: de Gruyter, 2009, pp. 257–278.

Zur Intonation des Spanischen junger bilingualer Sprecher aus Girona

Jonas Grünke
Johannes-Gutenberg-Universität Mainz
jguenke@uni-mainz.de

Während sich die linguistische Forschung zu Katalonien traditionell vor allem auf die Erforschung des Katalanischen beschränkt hat, ist in den letzten Jahrzehnten eine Reihe vom Arbeiten entstanden, die sich auch den spezifischen Merkmalen des dortigen Spanischen zugewendet haben. So konnte ein Einfluss des intensiven Sprachkontaktes mit dem Katalanischen auf den verschiedensten sprachlichen Ebenen gezeigt werden. Beispielhaft sei hier für die segmentale Phonologie die Verwendung des velarisierten [ɬ] als Allophon von /l/, für die Morphologie die verstärkte Verwendung „regularisierter" *indefinido*-Formen wie *(yo) *traducí* statt *(yo) traduje*, für die Syntax der Gebrauch der frageeinleitenden Partikel *que* in Sätzen wie *¿Que tienen mandarinas?* und für die Lexik der Gebrauch der Wendung *encontrar a faltar* ‚vermissen' genannt [1].

Einen Bereich, der bisher jedoch nahezu gänzlich unerforscht geblieben ist, stellt die suprasegmentale Phonologie dar. Zu den wenigen einschlägigen Studien zählen [2], wo besonders bei jüngeren Sprechern ein prosodischer Einfluss des mallorkinischen Katalanisch auf die Intonation des dortigen Spanisch gezeigt werden konnte, und [3], [4], wo im Rahmen des AMPER-Projektes u. a. eine Übernahme der katalanischen Intonationsmuster in den oben erwähnten *que*-Fragen festgestellt wird.

Ziel dieses Beitrages ist es, erste Ergebnisse aus meiner Dissertation zur Prosodie des katalanischen Spanisch vorzustellen. Dafür wurden 30 junge, spanisch-katalanisch-bilinguale Sprecher (Altersschnitt: 20,8), die in der Provinz Girona geboren und aufgewachsen sind und heute dort leben, in beiden Sprachen aufgenommen. Sowohl das Verhältnis zwischen den Geschlechtern als auch der Anteil von spanisch- bzw. katalanisch-dominanten Sprechern war dabei ausgeglichen. Die Analyse eines mithilfe eines *Discourse Completion Task* erhobenen Korpus im Rahmen des Autosegmental-Metrischen Modells ([5, 6, 7]) soll zur Beantwortung der folgenden Forschungsfragen beitragen.

(1) Welche Intonationskonturen werden im katalanischen Spanisch verwendet?
(2) Entsprechen diese den Konturen des Katalanischen?
(3) Gibt es im Spanischen von Girona eine einheitliche Intonation oder unterscheiden sich die Intonationsmuster in Abhängigkeit von der Sprachdominanz?

Eine Hypothese, die es dabei zu überprüfen gilt, ist, dass (zumindest in Girona) keine einheitliche Intonation für das Spanische existiert, sondern dass diese insofern von der Sprachdominanz abhängt, als katalanisch-dominante Sprecher Konturen aus dem Katalanischen übernehmen, während spanisch-dominante Sprecher in der Regel die Intonation ihrer Eltern übernehmen, die fast immer aus anderen Regionen Spaniens oder aus Lateinamerika stammen.

Literatur

[1] C. Sinner, *El castellano de Cataluña: estudio empírico de aspectos léxicos, morfosintácticos, pragmáticos y metalingüísticos*, Tübingen: Niemeyer, 2004.

[2] M. Simonet, *Language Contact in Majorca: An Experimental Sociophonetic Approach,* PhD dissertation, University of Illinois, 2008.

[3] L. Romera Barrios & A. Fernández Planas & V. Salcioli Guidi & J. Carrera Sabaté & D. Román Montes de Oca, "Una muestra del español de Barcelona en el marco AMPER," *Estudios de fonética experimental*, 16, pp. 147–184, 2007.

[4] L. Romera Barrios & V. Salcioli Guidi & A. Fernández Planas & J. Carrera Sabaté & D. Román Montes de Oca, "Prosody of Simple Sentences Spanish of Barcelona. A Spanish-Catalan Bilingual Context," in *Selected Proceedings of the 3rd Conference on Laboratory Approaches on Spanish Phonology*, L. Colantoni & J. Steele (Hrsg.), pp. 167–181, 2008.

[5] J. Pierrehumbert, *The Phonology and Phonetics of English Intonation*, Bloomington, Ind.: Indiana Univ. Linguistics Club, 1987.

[6] P. Prieto & P. Roseano (Hrsg.), *Transcription of Intonation of the Spanish Language*, München: Lincom Europa, 2010.

[7] J. I. Hualde & P. Prieto, "Intonational variation in Spanish: European and American varieties," in *Intonational Variation in Romance*, S. Frota & P. Prieto (Hrsg.), Oxford: Oxford University Press, pp. 350–391, 2015.

Homonym timing of vowel + consonant sequences in three Bavarian varieties

Nicola Klingler
Acoustics Research Institute – Austrian Academy of Sciences
nicola.klingler@oeaw.ac.at

The homonyms 'lauter$_1$' (comparative degree of the adjective 'laut', lit. 'louder') and 'lauter$_2$' (not comparable adjective of quantity, lit. 'many; nothing but ...') share their spelling since Early New High German (~1650 a. D.). They trace back to the (proto-) Germanic adjective stems **hluᵭa* (to sound) and **hluᵭ* (pure), displaying the then different pronunciation of the plosive in the phonemic rendering. The /d/ in **hluᵭa* was what we want to describe as lenis, a short plosive, with less to no aspiration, while the /t/ in **hluᵭ* is a fortis or geminate, a long plosive with a longer aspiration phase/voice onset time [1]. These assumptions are supported by the orthographic representations throughout the last centuries [2].

With the first and second German sound shift, and the transition from a syllable language (Old High and Middle High German) to a word language, some timing patterns in Standard German German (SGG) and Standard Austrian German (SAG), which includes the standard variety as spoken in Vienna (VS), coincide. While SAG still has features of both a word and a syllable language [3], SGG is very close to a prototypical word language [4]. Many studies have shown that SGG and SAG display a four-way timing pattern for vowel + consonant sequences: V+F; V:+L; V:+F; and V+L [5, 6]. As concerns the Central Bavarian dialects, which include the Viennese dialect (VD) and the dialect of Munich (MD), only two possibilities of vowel + consonant sequences are assumed in traditional dialectology [5, 7]: V+F and V:+L. However, [6] attested a further combination, namely V:+F, for VD.

The test words were recorded in carrier phrase which were produced five times, and in two speech rates by 17 speakers of VS, 13 speakers of MD, and 6 speakers of VD. The phonetic analysis includes measurements of the absolute and relative vowel and consonant durations, and the target word duration.

The results of this study indicate that both, MD and VD speakers display a clear distinction between lauter$_{1,2}$ via the duration of the plosive. In the Central Bavarian dialects, the plosive in 'lauter$_1$' is realised as a lenis plosive [ˈlɒːd̥ɐ] or [ˈlɒːdɐ], which is often spirantized to [ˈlɒːðɐ]. The plosive in 'lauter$_2$', which traces back to an old geminate, is realised as a fortis plosive [ˈlɒːtɐ]. However, VS speakers do not differentiate the homonyms due to the typological differences, in fact they treat lauter$_{1,2}$ as homonyms.

These results are in line with the presumed typological differences, and show the differences between the Central Bavarian varieties in Munich and Vienna: while MD speakers retain the complementary distinction between V+F and V:L, VD speakers seem to converse towards the Viennese standard variety timing pattern. These results will be discussed in regard of an assumed sound change in progress in Vienna [8].

References

[1] Kohler, K. 1977. Einführung in die Phonetik des Deutschen. Erich Schmidt Verlag, Berlin.

[2] Weigand, F. L. K. 1860. Deutsches Wörterbuch. Erster Band: A -L. Hrsg Hirt, H., Nachdruck, Berlin: De Gruyter.

[3] Moosmüller, S. and Brandstätter, J. 2014. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. Language Sciences 46, 84-95.

[4] Szczepaniak, R. 2007. Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache. Berlin, New York: de Gruyter (Studia Linguistica Germanica, 85).

[5] Ronneberger-Siebold, E. 1999. Ambisyllabic consonants in German: Evidence from dialectal pronunciation of lexical creations. In Rennison, J. and K. Kühnhammer (eds). Phonologica 1996. The Hague: Theseus. 247-271.

[6] Moosmüller, S. 2007. On some Timing Aspects of the Viennese Dialect. The Phonetician 95, 19-27.

[7] Pfalz, A. 1913. Die Mundart des Marchfeldes. Sitzungsberichte der kaiserlichen Akademie der Wissenschaften in Wien. Philosophisch-historische Klasse 170. Bd.6. Abhandlung. Wien: Hölder.

[8] Moosmüller, S. 2011. Aussprachevarianten im Österreichischen Standarddeutsch. In: I. Bose, B. Neuber (eds.), Interpersonelle Kommunikation - Analyse und Optimierung (pp. 193-199). Frankfurt: Lang.

Comparing the lingual correlates of coda /r/ in North and East Central German

Christina Otto, Adrian P. Simpson
Friedrich-Schiller-Universität Jena
otto.christina@uni-jena.de, adrian.simpson@uni-jena.de

The reduction of consonantal /r/ in coda position to a pharyngealized vowel has been described as a distinct and salient feature of East Central German (ECG) [2, 6]. As an earlier articulatory ultrasound tongue imaging (UTI) study on 7 male East Thuringian speakers shows, /r/ has an influence both on lingual configuration as well as on spectral measures (F1-F3) consistent with pharyngeal constriction. Articulatory and acoustic analysis of the dynamics of tongue configurations over the course of the word shows that the correlates of /r/ can extend over one or more syllables rather just a single segment [4, 5].

In an attempt to compare and contextualise the patterns produced by ECG speakers, we recorded a control sample comprising 10 speakers of broadly defined North German (NG). In common with ECG, /r/ at onset is realised as a uvular fricative. By contrast with ECG, however, we expect to find non-pharyngeal, vocalised correlates of /r/ in coda position [3]. Articulatory and acoustic data from the NG speakers were recorded using a typical UTI setup and AAA software [1]. Subjects produced an analogous sentence set to that used for the ECG speakers eliciting /r/ in onset and coda positions in short and long vowel contexts in a range of minimal pairs embedded in similar sentence frames (e.g. *Sie hat den* ***Metz/März*** *geliebt, Sie hat unter* ***Spott/Sport*** *gelitten*).

The present study examines the range and variation of articulatory realizations of /r/ in coda position after short vowels of 3 NG speakers. A first analysis at 2 different time points throughout the vocalic part of the word (after 25 % V2 and after 75% V4) shows individual variation in realizations of coda /r/ in short vowel contexts. Results of the NG speakers showed a clear change in tongue configuration over the course of the vocalic stretch which is also reflected in acoustic change in F1 and F2. This is particularly the case for front vowels such as /ɛ(r)/ in *Metz/März*. In stark contrast to this are the patterns produced by the ECG speakers who exhibit significant lingual differences throughout the whole of the syllable with correlates of /r/ being clearly present from syllable onset (right graphs in figure 1). Back vowels such as /ɔ(r)/ showed little or no tongue movement and reflect the auditory impression that the main correlate of /r/ in word pairs such as *Spott/Sport* is one of duration, whereas in ECG duration plays a more subordinate role (left graphs in figure 1), which is also supported by analysis and comparison of duration of the vocalic stretch.

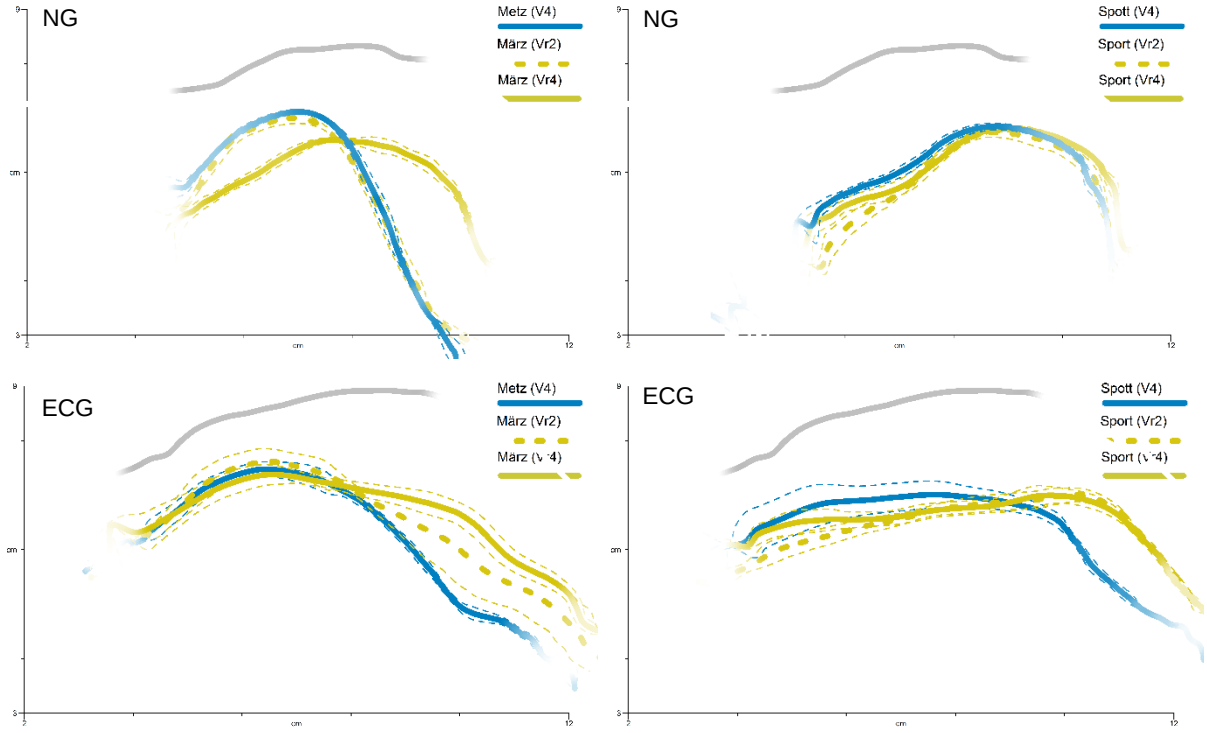


Figure 1: *Tongue configurations of word pairs without /r/ (blue) and with /r/ (yellow) at two different time points (after 25% dotted line, after 75% solid line) for a speaker of North German (NG, upper two graphs) and a speaker of East Central German (ECG, lower two graphs). Metz/März in the left column, Spott/Sport in the right column. Tongue tip is on the left, tongue root on the right.*

References

[1] Articulate Instruments Ltd, 2014. *Articulate Assistant Advanced User Guide: Version 2.16*, Articulate Instruments Ltd., Edinburgh, UK.

[2] S. u. D. Khan & C. Weise, "Upper Saxon (Chemnitz dialect). *JIPA*, 2013, *43*, 231–241.

[3] K. J. Kohler, 1995. *Einführung in die Phonetik des Deutschen*, Erich Schmidt Verlag.

[4] C. Otto & A. P. Simpson, "Pharyngealization of East Thuringian postvocalic /r/". In: The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow UK, 2015, pp. 1–5.

[5] C. Otto, M. Weirich & A. P. Simpson, "Spectral and lingual correlates of /r/ in East Thuringia". In: *Proceedings of the 10th ISSP*, Köln, 2014, pp. 293–296.

[6] K. Spangenberg, "Thuringian". In: C. Russ (Ed.), *The Dialects of Modern German: A Linguistic Survey*, Hoboken: Taylor and Francis, 2013, pp. 265–289.

VJ.PEAT: Automated measurement and classification of prosodic features

Tillmann Pistor, Carsten Keil
Research Center Deutscher Sprachatlas
Philipps-University of Marburg, Germany
tillmann.pistor@staff.uni-marburg.de, carsten.keil2@web.de

Reading and interpreting computer- or algorithm-based modellings of the acoustic correlates of prosodic structures can present several challenges (cf. [7], [8], [9], [10], [11], [12], [13] and [15]). Subjectivity of the results of measurements is one of these challenges that needs to be dealt with: adjusting measurement parameters for pitch or intensity manually may lead to differing results depending on the researcher's goals, biological or articulatory dispositions of the speaker or the communicative circumstances of the investigated corpus. How can the parameters be adjusted (normalized) with regard to the speaker individually but plausibly, without the need for manual re-adjustments afterwards? What representation of intensity and F0-courses reflects the approximation of what human perception perceives most accurately?

This contribution outlines the first steps of an innovative method of phonetic measurement of prosodic features, focusing on F_0-slopes in local intonation patterns called *PEAT*. The technique presented is an algorithm aiming to measure and detect phonetic differences in speech signals by applying machine-learning techniques. The process, operating on the basis of speech analysis and statistical computation programs such as Praat and R (cf. [3], [4], [5], [7], and [14]), successively uses robust acoustic variables processing (sweeping) (cf. [15] and [16]), a smoothing process based on the physiology of natural articulation (DCT4) (cf. [1], [2], [6] and [13]) and extraction of compliant paths according to a generic cost function (best fit). This process allows a fully automated determination of the calibration parameters when conducting phonetic measurements with Praat (cf. [11] and [12]), thus eliminating subjectivity and making the results and illustrations reliable and comparable. In subsequent steps, PEAT furthermore can be trained to detect the prevalence of so-called binary features to automatically detect, measure and classify prosodic units in unknown speech signals, thus bridging phonetics and machine-learning. It is hereby assumed, that each intonational unit in a specific interval shows a highly characteristic pattern, defined by binary features. The binary feature must be regarded as certain phonetic criteria (i.e. the timely variation of variables over a speech interval) and is either present in an unknown signal or not. This allows for the statistical testing of whether or not two different groups of speech segments (usually phonologically assembled at different points in real-time) separate significantly and can thus be classified.

References

[1] Ahmed, N./Natarajan, T./Rao, R. (1974): *Discrete Cosine Transform*. In: IEEE Transaction on Computers 1: 90–93.

[2] Atkinson, J. E. (1978): *Correlation analysis of the physiological factors controlling fundamental voice frequency*. (Journal of the Acoustical Society of America 63: 211–222).

[3] Boersma, P. (1993): *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.* (IFA Proceedings 17: 97–110).

[4] Boersma, P. (2013): *Acoustic analysis*. In: Podesva, R. J./Devyani S. [eds.] (2013): *Research Methods in Linguistics*. New York: Cambridge University Press: 375–397.

[5] Boersma, P./Weenink, D. (2017): *Praat. Doing phonetics by computer*. Version 6.0.17.

[6] Collier, R. (1975): *Physiological correlates of intonation patterns*. (Journal of the Acoustical Society of America 58: 249–255).

[7] Escudero, P./Boersma, P. (2009): *A cross-dialect acoustic description of vowels: Brazilian and European Portuguese.* (Journal of the Acoustical Society of America 126 (3): 1379–1393).

[8] Féry, C. (2017): *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press (Key Topics in Phonology).

[9] Gilles, P. (2005): *Regionale Prosodie im Deutschen. Variabilität in der Intonation von Abschluss und Weiterweisung*. Berlin/New York: de Gruyter (Impulse & Tendenzen 6).

[10] Gussenhoven, C. (2004): *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press (Research Surveys in Linguistics).

[11] Keil, C. (2017): *Der VokalJäger. Eine phonetisch-algorithmische Methode zur Vokaluntersuchung. Exemplarisch angewendet auf historische Tondokumente der Frankfurter Stadtmundart*. Hildesheim: Olms (Deutsche Dialektgeographie 122).

[12] Keil, C. (2017a): *Der VokalJäger*. Deutsche Dialektgeographie, Vol. 122. Georg Olms Verlag, Hildesheim. *Enhanced Algorithmic Toolbox, vokaljaeger.org*.

[13] Mixdorff, H. (2012): *The application of the Fujisaki model in quantitative prosody research*. In: Niebuhr, O. [ed.] (2012): *Understanding prosody. The role of context, function and communication*. Berlin: de Gruyter: 55–57.

[14] R Development Core Team (2015): *R: A Language and Environment for Statistical Computing*. Software. R Foundation for Statistical Computing. Wien.

[15] Evanini, K./Lai, C./Zechner, K. (2011): *The importance of optimal parameter setting for pitch extraction.* Proceedings of Meetings on Acoustics 2010 15. – 19. November 2010, Cancun, Mexico: 1–10. Acoustical Society of America.

[16] De Looze, C./ Rauzy, S. (2009): *Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration.* Proceedings of Interspeech 2009, Brighton, England.

Konsonantencluster in der Wortverarbeitung vietnamesischsprachiger Deutschlernender

Mehrgliedrige Konsonantenverbindungen stellen Fremdsprachenlernende oft vor Herausforderungen, da sie in der temporalen Organisation der artikulatorischen Gesten einen hohen Grad an Kompaktheit aufweisen können [1] oder einem phonologischen System mit für Lernende neuen phonotaktischen Regeln unterliegen, d.h., erworben werden müssen. Um ungewohnte Konsonantenstrukturen aufzubrechen wenden Fremdsprachenlernende verschiedene Strategien an, darunter Vokalepenthese [2, 3] oder auch die Reduktion von Clustern durch Tilgung einzelner Phoneme. Das Vietnamesische erlaubt, abgesehen von wenigen Ausnahmen [4], keine Konsonantencluster, was zur Folge hat, dass für vietnamesischsprachige Deutsch als Fremdsprache (DaF) Lernende schon zweigliedrige Konsonantenverbindungen mit einer relativ geringen Komplexität, wie beispielsweise /ʃp/ und /ʃt/, Probleme verursachen [5, 6].

In der vorliegenden Studie untersuchen wir, ob diese Probleme bereits in der Perzeption vorliegen, d.h. ob vietnamesischsprachige DaF-Lernende Schwierigkeiten haben, zweigliedrige Konsonantenverbindungen von eingliedrigen zu unterscheiden. Zudem untersuchen wir die Frage, welche Unterschiede sich in der Diskriminierungsfähigkeit hinsichtlich der Position des Konsonantenclusters in der Silbe zeigen.

Dazu wurde ein Wahrnehmungsexperiment zur Differenzierung des Clusters /ʃt/ vs. /ʃ/ und /t/ konzipiert, an dem 60 vietnamesische DaF-Lernende in Hanoi teilnahmen. 105 Wortpaare mit dem Kontrast in Onset- oder Coda-Position wurden angeboten (AX-discrimination task). Analysen zeigen, dass die generelle Unterscheidung wenig problematisch erscheint. Der Faktor Position (initial vs. final) ist hierbei relevant und stützt Befunde aus dem L1-Kontext [7]. Ausgehend von dem phonologischen System des Vietnamesischen diskutieren wir die Ergebnisse des Wahrnehmungsexperiments mit Bezug auf die Faktoren Position und Komplexität von Konsonantenclustern. Im didaktischen Kontext beleuchten wir Fragen der Sprachlernerfahrung DaF (Markiertheit des Clusters /ʃt/ in der Erwerbsprogression [8, 9]), und erörtern den Einfluss von Englisch als erster Fremdsprache beim Erwerb phonologischer Strukturen in DaF als zweiter Fremdsprache nach Englisch (DaFnE) [10].

Literatur

[1] C. P. Browman, L. Goldstein "Articulatory gestures as phonological units", *Phonology*, 6, pp. 201-251, 1989

[2] E. Dupoux, K. Kazuhiko, H. Yuki Hirose, C. Pallier & J. Mehler, "Epenthetic vowels in Japanese: A perceptual illusion?", *Journal of Experimental Psychology: Human Perception and Performance*, 25, p. 1568–1578, 1999.

[3] N. Hall, "Vowel epenthesis", in *The Blackwell companion to phonology*, M. van Oostendorp, C. J. Ewen, E.Hume & K. Rice (Hrsg.), Malden, MA & Oxford: Wiley-Blackwell, pp. 1576—1596.

[4] X. G. Le, Phonetische Interferenzen der vietnamesischen Deutschlernenden. Jena: Thüringer Universitäts- und Landesbibliothek, 2014.

[5] Kelz, Heinrich P., "Vietnamesisch", in *Phonetik international: Von Afrikaans bis Zulu. Kontrastive Studien für Deutsch als Fremdsprache*, Hirschfeld, Ursula; Kelz, Heinrich P. & Müller, Ursula (Hrsg.), Waldsteinberg: Heidrun Popp, Nr. 11, 2004.

[6] D. Diehl, "Integration von Phonetik in den DaF-Unterricht am Goethe-Institut Hanoi," in: *Phonetische und rhetorische Aspekte der interkulturellen Kommunikation*, U. Hirschfeld, F. Lange & E. Stock (Hrsg.), Berlin: Frank & Timme.

[7] Byrd, D., "Influences on articulatory timing in consonant sequences". *Journal of Phonetics*, 4 (2), pp. 209-244, 1996.

[8] H. Y. T. Vu, An optimality theoretic analysis of ESL coda acquisition by Vietnamese learners, Master thesis, The Arctic University of Norway, 2014.

[9] F. Eckmann, "Typological markedness and second language phonology," in *Phonology and second language acquisition*, E. J. G. Hansen, M. L. Zampini (Hrsg.), Philadelphia: John Benjamins, pp. 95–115, 2008.

[10] H. Tropf, "Sonority as a variability factor in second language phonology", in: *Sound patterns in second language acquisition*, A. James, J. Leather, pp. 173-191.

Sprechgeschwindigkeit von Müttern und Vätern in kind- und erwachsenengerichteter Sprache

Katharina Thon, Melanie Weirich, Adrian Simpson
KatharinaThon@gmx.net, melanie.weirich@uni-jena.de, Adrian.Simpson@uni-jena.de

Kindgerichtete Sprache (IDS) zeichnet sich durch eine höhere mittlere Grundfrequenz, einen größeren Tonhöhenbereich, kürzere Äußerungen und einen vergrößerten Vokalraum aus [1, 2, 3, 4, 5]. Zudem wird oft von einer verlangsamten Sprechgeschwindigkeit gesprochen [3, 6, 7]. Die meisten Studien beschäftigen sich ausschließlich mit IDS bei Müttern. Fernald et al. [4] haben in ihrer mehrere Sprachen umfassenden Studie hingegen IDS bei Müttern und Vätern untersucht und sowohl Gemeinsamkeiten als auch Unterschiede zwischen den Geschlechtern in den untersuchten prosodischen Parametern die Grundfrequenz und Dauer betreffend gefunden. Uns ist keine größere Studie bekannt, die sich mit den Unterschieden der Sprechgeschwindigkeit zwischen kind- und erwachsenengerichteter Sprache (ADS) des Deutschen beschäftigt und neben Müttern auch Väter untersucht und die Einbindung in die Kinderbetreuung mit betrachtet.

Die vorliegende Studie ist Teil eines Langzeitprojektes (*AvarE*, [8]), welches IDS und ADS bei schwedischen und deutschen Müttern und Vätern innerhalb des ersten Lebensjahres ihres Kindes untersucht. Hier wird sich auf deutsche Daten zu zwei verschiedenen Zeitpunkten konzentriert: vor der Geburt und zwischen dem 9. und 11. Lebensmonat des Kindes. Analysiert wurden u.a. die Sprechrate (SR, inklusive Pausen) und die Artikulationsrate (AR, exklusive Pausen) in gelesener Sprache (IDS und ADS) von 12 Müttern (in Elternzeit), 5 Vätern (in Elternzeit) und 10 Kontrollvätern (ohne Elternzeit). Außerdem wurden von allen Sprechern mithilfe eines Fragebogens (TMF, [9]) Daten bzgl. der selbsteingeschätzten Maskulinität/Femininität erfasst. Die Ergebnisse zeigen, dass sowohl Mütter als auch Väter in IDS signifikant langsamere Artikulations- und Sprechraten aufweisen als in ADS (AR: $\chi^2(1) = 23.01$, $p<.001$, SR: $\chi^2(1)=25.02$, $p<.001$). Zudem hat sich gezeigt, dass Väter, die Elternzeit nehmen sowohl in IDS als auch in ADS die langsamste Sprechergruppe sind, gefolgt von Müttern und Kontrollvätern. Väter und Kontrollväter weisen somit tendenziell größere Differenzen als Mütter und jeweilige Vätergruppe auf. Außerdem wurde bei den männlichen Sprechern eine signifikante Korrelation ($p < .01$) zwischen selbsteingeschätzter Maskulinität/Femininität und Sprechgeschwindigkeit gefunden: je weiblicher sie sich selbst eingeschätzt haben, umso langsamer haben sie in ADS ($r = .80$) und IDS ($r = -.67$) gesprochen. Beim ersten Zeitpunkt vor der Geburt des Kindes bestand dieser Zusammenhang nicht (vgl. Abbildung 1).

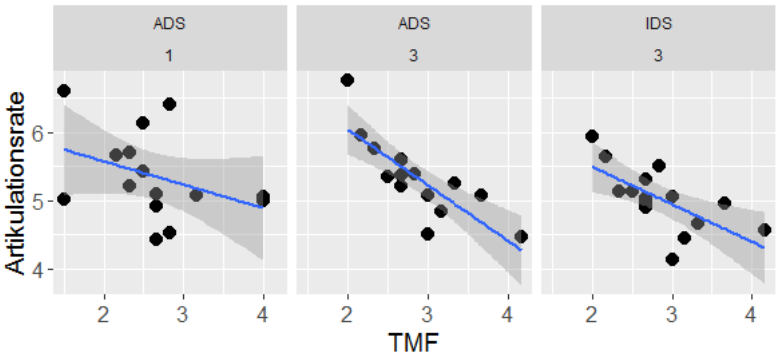


Abbildung 1: *Zusammenhang zwischen Artikulationsrate und TMF für alle Väter in ADS 1, ADS 3 und IDS*

Literatur

[1] Garnica, O. (1977) On some prosodic and paralinguistic features of speech to young children. *Talking to Children: Language Input and Acquisition*, ed. by C. Snow and C. Ferguson Cambridge: Cambridge University Press, 271-285.

[2] Kuhl, P., Andruski J., Chistovich I., Chistovich L., Kozhevnikova E., Ryskina V., Stolyarova E., Sundberg U. & Lacerda F. (1997) Cross-language analysis of phonetic units in language addressed to infants. *Science* , Vol. 277, Issue 5326, 684-686.

[3] Fernald, A. & Simon, T. (1984) Expanded Intonation Contours in Mothers' Speech to Newborns. *Developmental Psychology*, Vol. 20, No 1, 104-113.

[4] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B. & Fuku, I. (1989) A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang* 16, 477-501.

[5] Stern, D., Spieker, S., Barnett, R. & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes, *J. Child Lang.* 10, 1-15.

[6] Cooper, R. & Aslin, R. (1990) Preference for Infant-Directed Speech in the First Month after Birth. *Child Development*, Vol. 61, No. 5, 1584-1595.

[7] Tang, J. & Maidment, J. (1996) Prosodic aspects of child-directed speech in Cantonese, *Speech Hear. Lang.* 9, 257-276.

[8] Weirich, M. & Simpson, A. (2017) Acoustic correlates of parental role and gender identity in the speech of expecting parents. *Interspeech* 2017, Stockholm, 928-928.

[9] Kachel, S.**,** Steffens, M.C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. Frontiers in Personality and Social Psychology, 7:956.

How do anatomical differences affect individual speech rhythm?

Fabian Tomaschek[1], Adrian Leemann[2]
[1] University of Tübingen, [2] Lancaster University
fabian.tomaschek@uni-tuebingen.de, a.leemann@lancaster.ac.uk

It is widely accepted that languages, dialects and speakers differ in speech rhythm (Ramus, Nespor, and Mehler 1999; Grabe and Low 2002). One parameter to capture speech rhythm is %V, i.e. the percentage of vocalic intervals in the speech signal. Regarding this parameter, Wiget et al. (2010), V. Dellwo (2010), and V. Dellwo, Leemann, and M.-J Kolly (2015) and Leemann, M-J. Kolly, and V. Dellwo (2015) reported that %V, in particular, is speaker specific and speculated that anatomical differences between speakers might potentially account for these differences in %V. Their hypothesis seems to be supported by findings of idiosyncratic articulatory movement behavior due to differences in vocal tract size (Winkler, Fuchs, and Perrier 2006; Fuchs, Winkler, and Perrier 2008; Brunner, Fuchs, and Perrier 2009) or differences in the shape of the alveolo-palatal region (Weirich and Fuchs 2006; Rudy and Yunusova 2013).

In the present study, we investigated the relation between anatomical differences and the rhythm measure %V. It is a follow up to the study by Tomaschek, Arnold, and Volker Dellwo (2015), who showed that differences in mouth angle accounted for differences in speech rhythm. Here, we investigated articulations of the German pronoun 'sie' *they* articulated in roughly 380 [zi:] + verb phrases uttered by 21 native German speakers in two different speaking rate conditions (slow/fast), as recorded in Tomaschek, Tucker, et al. (2014). Recordings were done by means of the NDI articulograph for three sensors attached to the speakers' tongue, one at the tongue tip, one at the tongue middle and one at the tongue body (TB) (cf. Figure 1, left). Anatomical differences between speakers were parameterized by the individual movement area, i.e. a rhombus whose diagonals were defined by the distance between maximal and minimal horizontal and vertical tongue positions of a speaker (cf. area around the tongue body sensor TB in Figure 1, left).

We predicted that %V would increase, the larger a speaker's movement area. A random-forest analysis with movement area, speaking rate condition, syllable rate parameterizing local speaking rate and repetition of the word across the experiment as predictors for %V indicated that movement area is the strongest predictor for the %V. A mixed-effects linear regression analysis in which we controlled for possible linguistic confounds such as local speaking rate ($\beta = 0.07$, sd = 0.01, t = 11.96) and the identity of the upcoming word, indeed supported our hypothesis. Average individual %V were proportional to the tongue's movement area of the speakers (cf. Figure 1, right), with the effect being smaller in the slow condition (β = 0.04, sd = 0.01, t = 4.96) than in the fast condition (β = 0.09, sd = 0.01, t = 10.19).

Our results support the hypothesis by V. Dellwo, Leemann, and M.-J Kolly (2015) and Leemann, M-J. Kolly, and V. Dellwo (2015) that between-speaker variation in %V can be accounted by anatomical differences. In future work, we aim to determine whether further morphological features can be inferred from the speech signal.

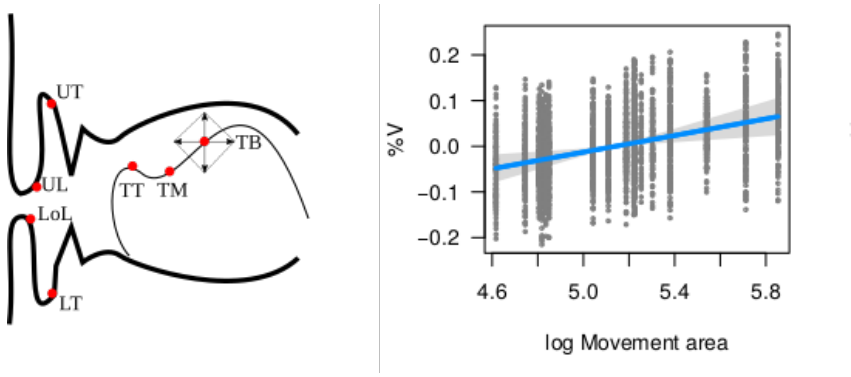


Figure 1: ***left:*** *Illustration of sensor position and movement area arount TB.* ***right:*** *Illustration of the partial effect of movement area on %V.*

References

[1] Jana Brunner, S. Fuchs, and P. Perrier. "On the relationship between palate shape and articulatory behavior." In: *The Journal of the Acoustical Society of America* 125.6 (2009), pp. 3936–3949. DOI: http://dx.doi.org/10.1121/1.3125313.

[2] V. Dellwo. *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence*. Rheinischen Friedrich-Wilhelms-Universität, Bonn, 2010.

[3] V. Dellwo, A. Leemann, and M.-J Kolly. "Rhythmic variability between speakers: Articulatory, prosodic and lexical factors." In: *Journal of the Acoustical Society of America* (2015).

[4] S. Fuchs, R. Winkler, and P. Perrier. "Do speakers' vocal tract geometries shape their articulatory vowel space?" In: 2008, pp. 333–336.

[5] E. Grabe and E. L. Low. "Duration variability in speech and the rhythm class hypthesis." In: *Laboratory Phonology* (2002).

[6] A. Leemann, M-J. Kolly, and V. Dellwo. "Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison." In: *Forensic Science International* (2015).

[7] F. Ramus, M. Nespor, and J. Mehler. "Correlates of linguistic rhythm in the speech signal." In: *Cognition* (1999).

[8] K. Rudy and Y. Yunusova. "The Effect of Anatomic Factors on Tongue Position Variability During Consonants." In: *Journal of Speech Language and Hearing Research* 56 (2013), pp. 137–149.

[9] F. Tomaschek, D. Arnold, and Volker Dellwo. "Face morphology affects speech rhythm: A pilot study." In: *PP 10 - 10. Tagung Phonetik und Phonologie im deutschsprachigen Raum, Konstanz*. Poster download at fabian-tomaschek.com/publications/, 2015.

[10] F. Tomaschek, B. V. Tucker, et al. "Vowel articulation affected by word frequency." In: *Proceedings of the 10th ISSP*. Cologne, 2014, pp. 425–428.

[11] M. Weirich and S. Fuchs. "Palatal morphology can influence speaker-specific realizations of phonemic contrasts." In: *Journal of Speech, Language, and Hearing Research* 56 (2006), pp. 1894–1908.

[12] L. Wiget et al. "How stable are acoustic metrics of contrastive speech rhythm?" In: *Journal of the Acoustical Society of America* (2010).

[13] R. Winkler, S. Fuchs, and P. Perrier. "The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modeling." In: 2006, pp. 509–516.

On the effects of state borders and standard varieties on the spread of the Bavarian /e-ɛ/ merger

Katrin Wolfswinkler[1], Felicitas Kleber[1], Conceição Cunha[1]
[1]Institute of Phonetics and Speech Processing, LMU Munich
katrin@phonetik.uni-muenchen.de, kleber@phonetik.uni-muenchen.de, cunha@phonetik.uni-muenchen.de

The main aim of this study was to investigate a potential spread of the East Central Bavarian /e-ɛ/ merger across the Austrian-German state border towards the West Central Bavarian dialect area. More specifically, we seek to better understand the extent to which the political border also acts as language divide, partly due to diverging influences of the two regional standard varieties (Standard German and Standard Austrian) on the same underlying dialect [7]. The well-documented merger is characterized by free variation between /e/ and /ɛ/ and has spread from Vienna (hence "Viennese *E*-confusion" [3]) to Salzburg albeit with different merging tendencies (Vienna: towards /e/ [8]; Salzburg: towards /ɛ/ [5]). Given the high frequency of (1) linguistically motivated *E*-vowel mergers (cf. [4]) and (2) cross-border contacts, a spread of the merger towards the German side is not unlikely. However, the influence of Southern Standard German where the distinction is still made [2] may impede a cross-border spread. We therefor predict tendencies towards a merged *E*-category as found in the city dialect of Salzburg [5] in German speakers of Bavarian (GB) though less pronounced than in Austrian speakers from the Salzburg area. Furthermore, if this is a sound change in progress then younger GB speakers should tend towards more neutralization than older GB speakers.

In an apparent-time study we are currently collecting acoustic recordings from younger and older GB speakers and from Austrian Bavarian (AB) control speakers using a picture-naming task. All speakers are living close to the border near the city of Salzburg, either on the German (GB speakers) or the Austrian (AB speakers) side. To date we have analyzed 458 *E*-tokens (223 with underlying /e/ and 235 with underlying /ɛ/) from 8 younger (average age 24) and older (average age 53) GB and 4 AB speakers. To measure the neutralization degree in *E*-vowels, the F1-difference between /e/ and /ɛ/ tokens was calculated per speaker.

The lower F1-difference in AB compared to GB speakers in Fig.1 is in line with previous findings reporting a westward spread of the merger to the Salzburg area (cf. [5]) but suggests that this change has not (yet) spread to the German side. The tendency towards an even greater F1-difference (e.g. in form of the median) but also the greater variability (greater dispersion of the data) in younger Germans compared to the older group points to two trends: some younger GB speakers tend towards a greater differentiation than older GB speakers while others do indeed confuse *E*-vowels more often. While the latter trend may indeed originate from contact with AB speakers the former may be due to a Southern Standard German influence. The results will be discussed in an exemplar-based [6] model of dialect levelling [1] that accounts for both the spread across the border and a deceleration of the change due to potentially diverging influences of regional standard varieties.

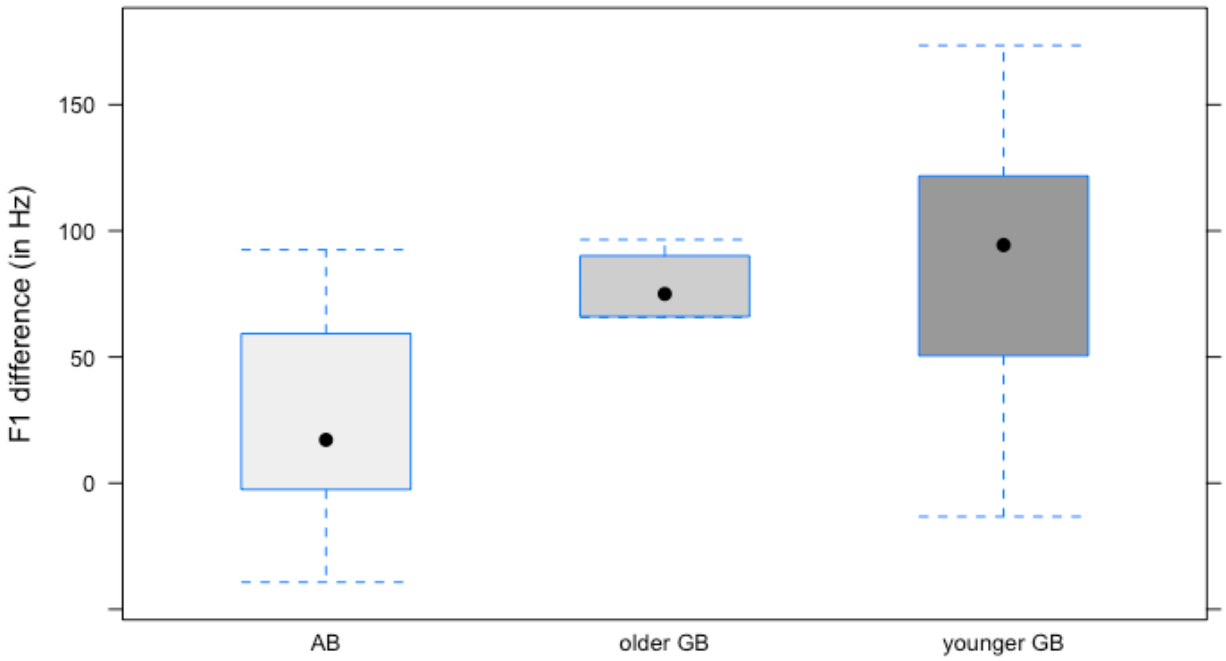


Figure 1. *F1-differences [Hz] between tokens with underlying /e/ and tokens with underlying /ɛ/ separately for Austrian Bavarian (AB), older and younger German Bavarian (GB) speakers.*

References

[1] Bukmaier, V., Harrington, J., & Kleber, F. 2014. An analysis of post-vocalic/s-ʃ/neutralization in Augsburg German: evidence for a gradient sound change. *Frontiers in psychology*, *5*.

[2] Kleiner, S. & Knöbl, R. 2011. Hochdeutsch und Hochdeutsch: Regionale Gebrauchsstandards im gesprochenem Deutsch. *No. Sprachreport 2*. Mannheim: Institut für Deutsche Sprache.

[3] Kranzmayer, E. 1953. Lautwandlungen und Lautverschiebungen im gegenwärtigen Wienerischen. *Zeitschrift für Mundartforschung*, *21*(H. 4), 197–239.

[4] Labov, W. 1994. Principles of language change: Internal factors. New York: Blackwell.

[5] Moosmüller, S., & Scheutz, H. 2013. Chain shifts revisited: The case of Monophthongisation and E-confusion in the city dialects of Salzburg and Vienna. *Language Variation–European Perspectives IV. Benjamins, Amsterdam*, 173–186.

[6] Pierrehumbert, J. 2002. Word-specific phonetics. *Laboratory phonology VII. Berlin: Mouton de Gruyter*, 101-40.

[7] Scheuringer, H. 1990. *Sprachentwicklung in Bayern und Österreich: eine Analyse des Substandardverhaltens der Städte Braunau am Inn (Österreich) und Simbach am Inn (Bayern) und ihres Umlandes* (Vol. 3). Buske Verlag.

[8] Seidelmann, E. 1971. Lautwandel und Systemwandel in der Wiener Stadtmundart. Ein strukturgeschichtlicher Abriß. *Zeitschrift für Dialektologie und Linguistik*, 145–166.

[9] Wagener, P. 2002. German dialects in real-time change. *Journal of Germanic Linguistics*, *14*(3), 271-285.

Abstract für die P&P 14 – Wien, 6.-7. September 2018

Lea-Sophie Adam, Luise Henneberg, Kim Jördens, Marco Krämer, Elizabeth Pankratz, Sarah Wesolek, Songyang Yan and Christine Mooshammer
Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin
christine.mooshammer@hu-berlin.de

The perception of the /ʃ - ç/ contrast in younger and older German listeners

Ongoing sound change has been experimentally shown in many languages. A prominent example is /ʃ/ and /ç/ in modern Standard German. Jannedy and Weirich (2014) [1] examined this phenomenon with regard to Kiezdeutsch and found that age was a reliable predictor of differing perceptions of the fricative, with young listeners identifying significantly more stimuli as /ʃ/ than older listeners.

The aim of our study was to replicate their findings concerning age. This would lend further support to their finding that there may be a sound change in progress in the way that German language users perceive these fricatives. We expanded the scope of the study to include speakers from across Germany, rather than focusing only on speakers from Berlin. In total, 18 speakers came from Berlin and 26 came from outside of Berlin.

We used the same stimuli as [1] which were synthetic stimuli of two acoustic continua between *fischte* ("fished") and *Fichte* ("spruce"). Two sets of stimuli were used here: the first set without vowel rounding (vowel [ɪ]), preceding /ʃ/ which consisted of 27 steps from naturally produced [fɪçtə] to [fɪʃtə] for which only the fricative was changed. The second set includes anticipatory lip rounding for /ʃ/, going from [fɪçtə] to [fʏʃtə] with fricative and vowel changed in 27 steps. Each subject takes 4 tests: two identification tests and two discrimination tests respectively with [ʏ] and [ɪ]. For the identification test the subject chose between two options which were presented as pictures on screen (a fishing young man and a spruce). For the AXB discrimination task the subject listened 3 stimuli and was asked to choose if the first or the third one is the same as the second stimulus by clicking the one of the two buttons on screen. The subjects were 18 older listeners from 50-65 years and 24 younger listeners from 18-30 years. Their regional background varied with listeners from Bavaria (1), Hesse (3), Berlin (18), Lower Saxony (9), Brandenburg (5), North Rhine-Westphalia (3) , Schleswig Holstein (2).

Figure 1 shows the identification results for the vowel continua, with percentage of Fichte answers for the younger group in red and for the older group in green. On the left side the [fɪçtə] to [fɪʃtə] continuum is shown, on the right the [fɪçtə] to [fʏʃtə]. There is a clear difference in the two curves: the younger listeners identify "fischte" more often than "fichte" in both vowel continua. This confirms the former study of Jannedy & Weirich (2014) showing that the category boundary between /ʃ - ç/ is indeed shifting for the younger generation. Whether this sound change is mainly lead by the Berlin listeners will be investigated in the near future.

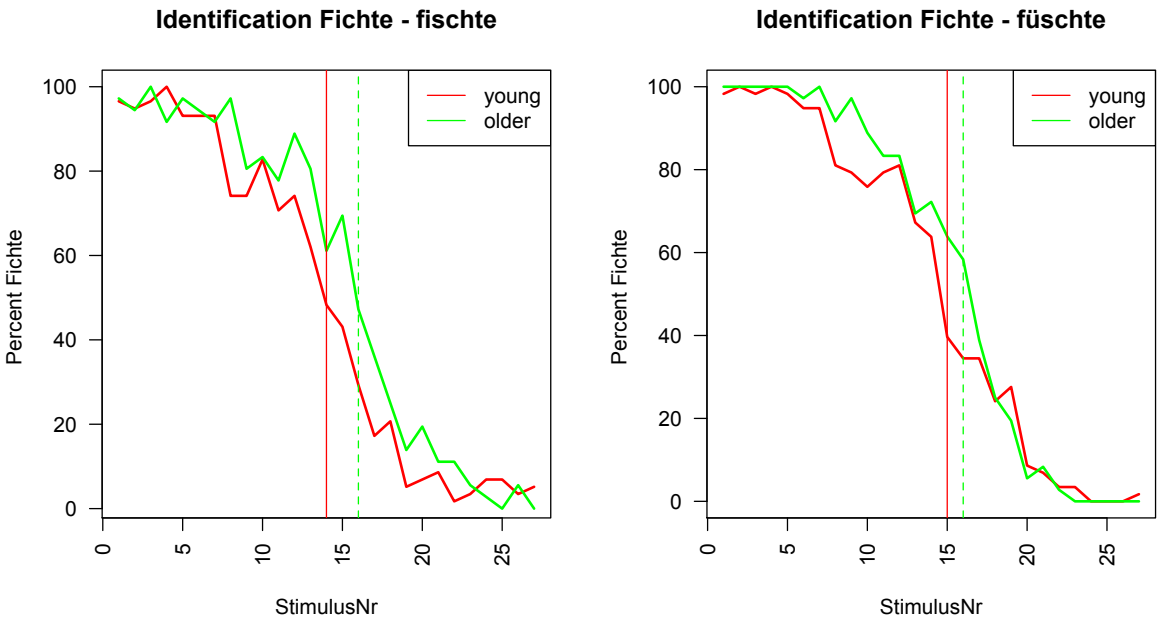


Figure 1: *Identification curves for younger (red) and older listeners (green) for the two stimulus sets. The red and green lines indicate the boundary, located at the 50% threshold.*

Literatur

[1] Jannedy, S., & Weirich, M. (2014). Sound change in an urban setting: Category instability of the palatal fricative in Berlin. *Laboratory Phonology*, *5*(1), 91-122.

The good, the bad, the bouba, and the kiki
Cross-modal correspondences between evaluative meanings, speech-sounds, and object shapes

Fabian Bross[1]
[1]University of Stuttgart
fabian.bross@ling.uni-stuttgart.de

A large body of evidence suggests the existence of a cognitive association between certain speech sounds and visual cues—a phenomenon known under the label 'bouba/kiki effect'. It was, for example, found that high, front vowels are judged to refer to small size objects and low, back vowels rather to big objects (e. g. [1], [2]). Other studies have found an association between pseudo-words with voiceless consonants, like *kiki*, and spiky objects and pseudo-words containing voiced consonants, like *bouba*, with round objects ([3]). The aim of the present poster is to present evidence for a cognitive association between pseudo-words of the *kiki*-type and negative meanings (e. .g, *hate*) and *bouba*-type pseudo-words and positive concepts (e. g. *love*). Additionally, it was tested if negative meanings are rather associated with spiky shapes and positive meanings with round objects.

Two experiments were conducted. In Experiment one, thirty-three native speakers of German with a mean age of 24.3 (*SD* = 5.81) were asked to judge eight predicates—four with positive and four with negative meanings (e. g., *to hate someone* or *to love someone*). Participants were told that they should guess the translation of these predicates in an unknown foreign language. For this, they were randomly presented with a set of pseudo-words that had proven to be useful in similar studies: (e. g., [4], [5], [6]): *takete*/*maluma*, *kiki*/*bouba*, *decter*/*bobolo*, and *ruki*/*lula*. So, for example, a participant saw the predicate *to love someone* and had to decide if this would translate to *decter* or *bobolo*. It was hypothesized that pseudo-words of the *kiki*-type would be preferred to relate to negative meanings and *bouba*-type words to positive ones. This was indeed found, as shown in Figure 1.

One problem with the results obtained in Experiment one is that participants were asked to associate words with (pseudo-)words. Hence, what was tested was not necessarily a cross-modal correspondence. This problem was tackled in Experiment two. As it is well known that *kiki*-type words are associated with spiky and *bouba*-type words with round shapes it was predicted that predicates referring to negative meanings would be judged to be better represented by spiky shapes and that positive concepts are rather associated with round objects. For this purpose, the same predicates as in Experiment one were used. Instead of using pseudo-words, four picture pairs were constructed consisting of spiky and round shapes as they are traditionally used in such tasks. Participants were twenty-nine native speakers of German with a mean age of 20.66 (*SD* = 2.51) who did not participate in Experiment one. The hypothesized effect was indeed found and was much stronger than the effect in Experiment one. See Figure 2 for details. The results of the two experiments present evidence for a cross-modal mapping between speech-sounds, evaluative meanings (evaluation as good vs. evaluation as bad) and object shapes.

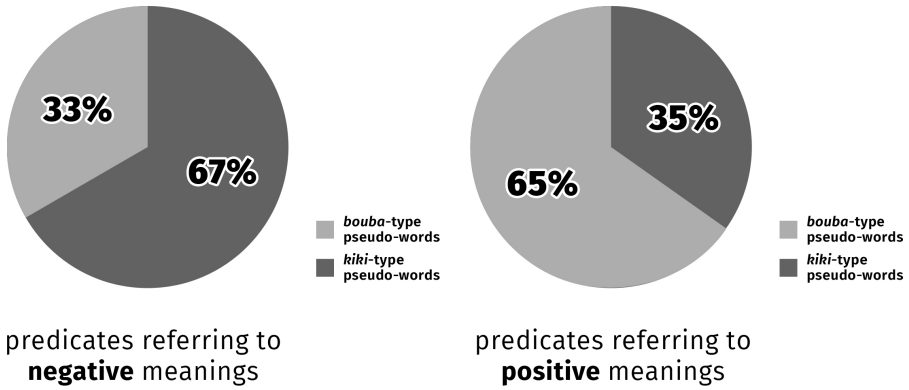


Figure 1: *Results of Experiment one: Participants chose more pseudo-words that were found to be rated spiky in previous studies (kiki-type) when it comes to negative meanings and vice versa. Assuming that there was no influence of the meaning being evaluative (i. .e. negative or positive), the null-hypothesis was that participants were guessing (i. e., H_0 predicted a 50/50 distribution of choices). Two exact binominal tests were calculated revealing a p-value of 0.0001602 and 0.0006313 respectively.*

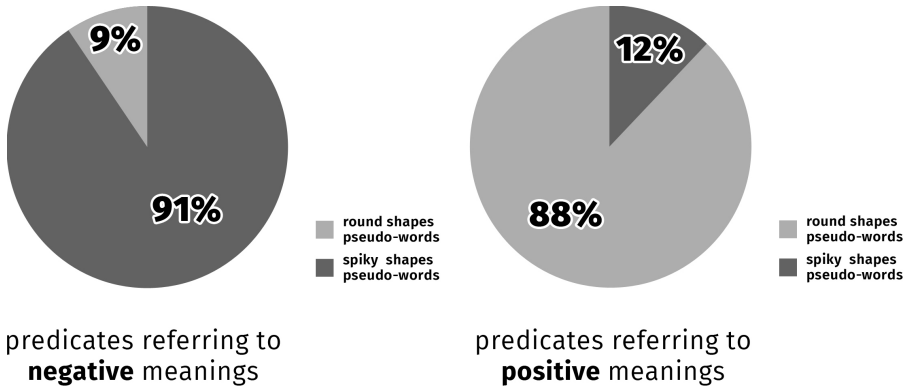


Figure 2: *Results of Experiment two: Participants preferred spiky shapes for negative meanings and vice versa. Assuming that there was no influence of the meaning being evaluative (i. .e. negative or positive), the null-hypothesis was that participants were guessing. Two exact binominal tests were calculated revealing a p-values smaller than 0.0000001.*

References

[1] Sapir, Edward (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12, 2251–1239.

[2] Blasi, Damián E., Wichmann, Søren, Hammarström, H., Stadler, P. F. & Christiansen, M. H. (2016). Sound meaning associations biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences: Anthropology, Psychological and Cognitive Sciences*, 113(39), 10818–10823.

[3] Ramachandran, Vilayanur Subramanian & Hubbard, Edward M. (2001). Synaesthesia–a window into perception. *Journal of Consciousness Studies*, 8, 3–34.

[4] Spence, Charles & Alberto Gallace (2010). Tasting shapes and words. *Food Quality and Preference*, 22, 290–295.

[5] Gallace, Alberto, Erica Boschin & Charles Spence (2011). On the taste of "Bouba" and "Kiki": An exploration of wordfood associations in neurologically normal participants. *Cognitive Neuroscience*, 2(1), 34–46.

[6] Cristinel, Anne-Sylvie, Sophie Jones & Charles Spence (2012). 'The sweet taste of maluma': Crossmodal associations between tastes and words. *Chemosensory Perception*, 5, 266273.

Examining the temporal extent of co-articulatory nasalization in German using real-time MRI

Christopher Carignan[1], Philip Hoole[1], Esther Kunay[1], Marianne Pouplier[1], Jonathan Harrington[1], Jens Frahm[2]
[1]*Institut für Phonetik und Sprachverarbeitung Ludwig-Maximilians-Universität München,*
[2]*Biomedizinische NMR Forschungs GmbH Göttingen*

Accurate knowledge of the temporal patterns involved in co-articulatory nasalization is paramount to our understanding of how phonemic vowel nasality can emerge diachronically. This paper examines the hypothesis that some aspects of variable vowel nasalization are due to variation in the temporal alignment of the nasal and oral gestures for nasal consonants [1]. This hypothesis predicts a greater degree of co-articulatory vowel nasalization of, e.g., /Vnt/ compared to /Vnd/, due to earlier onset of a constant-sized nasal gesture. Evidence in support of this hypothesis has been observed in English, Ikalanga, Thai, and Rwanda using aerodynamic data [2, 3, 4]. Here, we test this hypothesis for German using real-time magnetic resonance imaging (rtMRI), which allows direct observation of velum movement.

rtMRI data were collected at a temporal resolution of 50 fps and a spatial resolution of $\approx 2mm^2$. The corpus consists of $\approx$ 300 German lexical items; a subset containing only coronal coda consonants was used to examine the effect of consonant voicing on velum opening. The data from 11 (of 30+ recorded) speakers have been analyzed and are presented here. A velum opening (i.e., degree of nasalization) signal was created according to the following method. A region of interest (RoI) was manually selected around the area of velum movement for each speaker. The voxels in the RoI were then extracted for the images pertaining to words containing VN sequences. The voxel intensities were used as dimensions in principal components analysis models, and the coefficients from the first PC (relating to the axis of velum movement) were used to create a time-varying signal of velum opening for the entire data set. An example of the signal is shown in Figure 1 for the word *lehnt* [leːnt] "(he/she/it) leans". The velum opening signals were submitted to functional linear mixed models (FLMMs; [5]) using the *sparseFLMM* package in R [6], with either phonetic context (nasal-oral) or coda type (e.g., /n/-/nd/, /nd/-/nt/) as a fixed effect, and speaker and word as random effects. The results of the FLMMs are presented in Figures 2-5.

The results indicate that /Vn/ sequences exhibit extensive co-articulatory nasalization in German (Figure 2). However, comparison of /Vn/-/Vnd/-/Vnt/ sequences (Figures 3-5) reveals that co-articulatory nasalization is greater in /Vnd/ sequences compared to both /Vnt/ and /Vn/ sequences. This result is inconsistent with the hypothesis that this type of variation in vowel nasalization is due to variation in the temporal alignment of gestures [1], which predicts greater co-articulatory nasalization for /Vnt/ compared to /Vnd/. These findings suggest that: (1) the phonetic implementation of co-articulatory vowel nasalization in German is somehow different than in English, Ikalanga, Thai, and Rwanda [2, 3, 4], and/or (2) that direct observation of velum movement (e.g, using rtMRI) yields different kinematic characteristics when compared to indirect observation of velum movement (e.g., using aerodynamics).

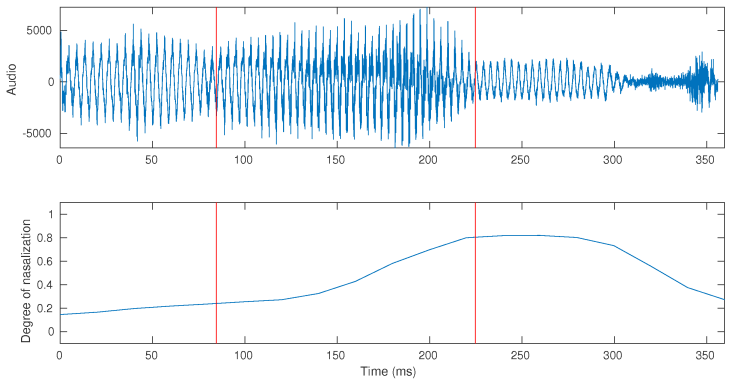


Figure 1: *Audio waveform and corresponding velum opening signal for an utterance of* lehnt *[leːnt] "(he/she/it) leans". The temporal boundaries of the vowel [eː] are marked by vertical red lines.*

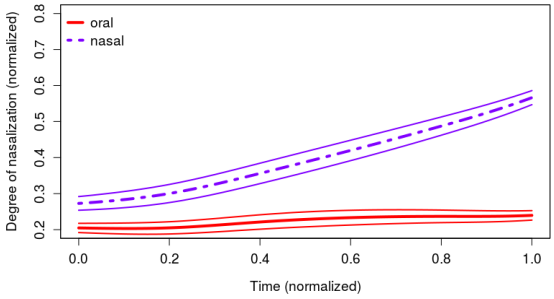


Figure 2: *FLMMs for pre-oral & -nasal vowels.*

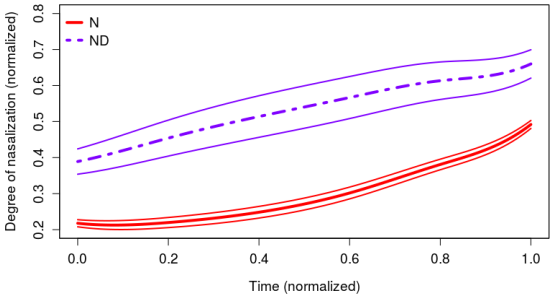


Figure 4: *FLMMs for pre-/n/ & -/nd/ vowels.*

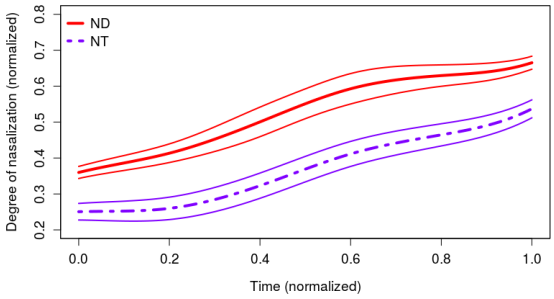


Figure 3: *FLMMs for pre-/nd/ & -/nt/ vowels.*

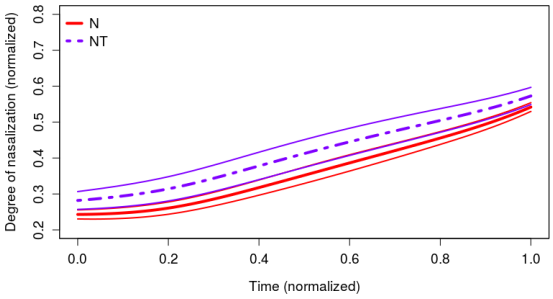


Figure 5: *FLMMs for pre-/n/ & -/nt/ vowels.*

References

[1] P. S. Beddor, "A coarticulatory path to sound change," *Language*, 85(4), pp. 785–821, 2009.

[2] P. S. Beddor, *Nasals and nasalization: The relation between segmental and coarticulatory timing*, Proceedings 16^{th} ICPhS Saarbrucken, pp. 249–254, 2007.

[3] D. Demolin, *Coarticulatory timing and aerodynamics of nasals and nasalization*, Proceedings 16^{th} ICPhS Saarbrucken, pp. 271–274, 2007.

[4] A. C. Cohn, *Phonetic and phonological rules of nasalization*, UCLA Working Papers in Phonetics 76, pp. 1–224, 1990.

[5] M. Pouplier, J. Cederbaum, P. Hoole, S. Marin, & S. Greven, "Mixed modeling for irregularly sampled and correlated functional data: Speech science applications," *Journal of the Acoustical Society of America*, 142(2), pp. 935–946, 2017.

[6] J. Cederbaum, "sparseFLMM: Functional linear mixed models for irregularly or sparsely sampled data. R package version 0.1.0," https://CRAN.R-project.org/packages/sparseFLMM, 2017.

Devoicing of /z/ in the Bavarian dialect: a large scale study

Johanna Cronenberg, Laura-Annika Haller, Iveta Patáková, Christoph Draxler
Institut für Phonetik und Sprachverarbeitung, LMU
johanna.cronenberg@campus.lmu.de, l.haller@campus.lmu.de,
iveta.patakova@campus.lmu.de, draxler@phonetik.uni-muenchen.de

This study investigates the devoicing of the alveolar sibilant /z/ by adolescent Bavarian speakers. It was hypothesised that adolescents from Bavarian small towns would devoice /z/ in word- and syllable-initial positions to a stronger degree than their contemporaries living in Bavarian cities who, in turn, would be more oriented towards a standardised pronunciation. These hypotheses are based on literature, which suggests that, especially in the younger generations, the use of dialect is declining in larger cities [1] whereas the relatively homogeneous population in small Bavarian towns demonstrates a positive attitude towards the dominant dialect and therefore uses it regularly [2]. It is also known that the phonologically voiced alveolar sibilant is devoiced in word- and syllable-initial positions in the Bavarian dialect in contrast to Standard German [3].

Three speaker groups were generated from the two speech databases RVG-J (http://hdl.handle.net/11022/1009-0000-0004-AE1D-9) and Ph@ttSessionz (http://hdl.handle.net/11022/1009-0000-0000-CC6A-4): "city" (185 speakers from Bavarian cities), "town" (214 speakers from Bavarian small towns), and "notBY" (398 speakers from other German federal states). In total, 96.270 items were measured for Bavarian speakers (either belonging to group "town" or group "city") and 95.028 for non-Bavarian speakers. Some examples for syllable-initial position and voiced context are "Besuch" and "Pause" and for syllable-final position and voiceless context "letztmalig" and "selbstverständlich" since in those two cases, the / t / in the coda is not produced. The fundamental frequency of all phonetic segments labelled [s] or [z] was determined using the ksvF0 function [4], and for each condition, the voicelessness was calculated in percent (see Figure 1). In each of the six conditions, [s] is relatively stable in its amount of voiceless measuring points over all three speaker groups. Speakers from group "town" always show a stronger degree of voicelessness in [z] than speakers from group "city". Speaker groups "city" and "notBY" exhibited a more similar behaviour in the voicing or devoicing of alveolar sibilants than they did in comparison to speaker group "town". The exception to this rule can be found in the word-final condition where the non-Bavarian speaker group shows an almost equal degree of voicelessness in [z] as the speaker group "town". It can also be noted that there are no labels [z] in syllable-final position and voiceless phonetic context.

For the moment, we conclude that there is indeed a difference in the pronunciation of the alveolar sibilants between speakers from Bavarian small towns and Bavarian cities, as well as between Bavarian and non-Bavarian adolescents. In comparison to other dialectological studies, the present one is unique in terms of the amount of data used and the automatic procedures of data processing. More detailed investigations will also take into account factors like age and gender of participants, speaking rate, and speaking modality which could have an influence on the fundamental frequency in sibilants.

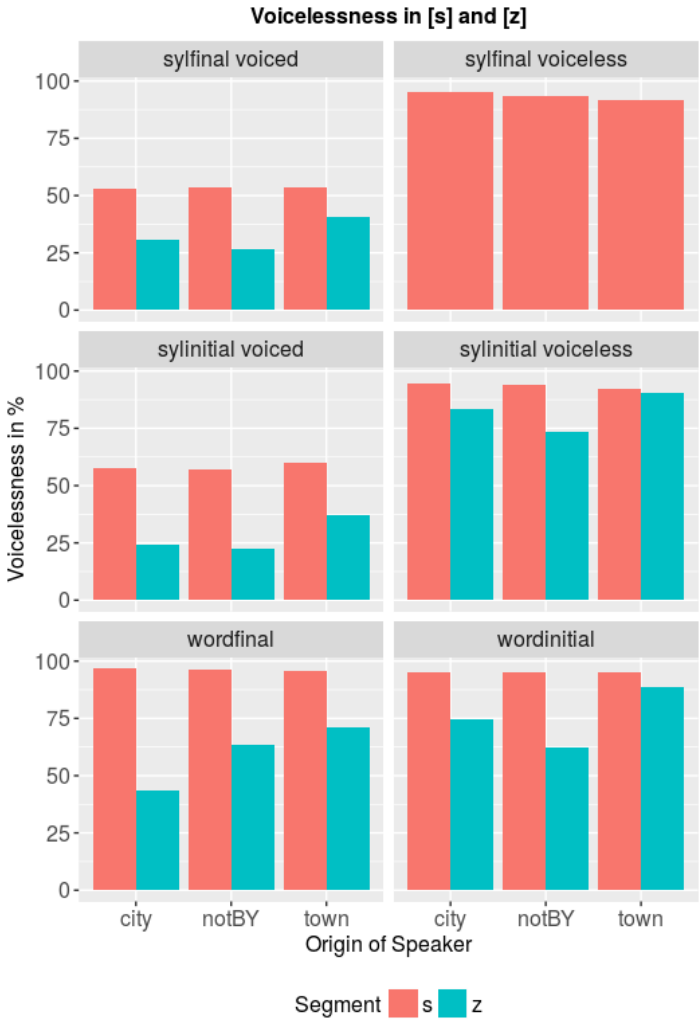


Figure 1: *Percentages of voicelessness of segments labelled [s] (red) and [z] (blue) for all six conditions, separated by speaker group (city, town, and notBY). The condition "sylfinal voiced" includes segments that appeared in syllable-final position and voiced phonetic context, "sylinitial voiceless" comprises segments in syllable-initial position and voiceless context, etc. In word-final and word-initial position, the phonetic context was not considered.*

References

[1] M. Renn, *Die Mundart im Raum Augsburg: Untersuchungen zum Dialekt und zum Dialektwandel im Spannungsfeld großstädtisch-ländlicher und alemannisch-bairischer Gegensätze*, Heidelberg: Universitäts-Verlag Winter, 1994.

[2] G. Steinegger, *Situativer Sprachgebrauch und Spracheinschätzung in Österreich und Südtirol. Auswertung einer Umfrage*, Wien: Universität Wien, 1997.

[3] H. G. Piroth & P. M. Janker, "Speaker-dependent differences in voicing and devoicing of German obstruents," *Journal of Phonetics*, 32, pp. 81-109, 2004.

[4] K. Schäfer-Vincent, "Pitch period detection and chaining: method and evaluation," *Phonetica*, 40, pp. 177-202, 1983.

How to Measure a Pleasant Voice

Jula Marie Carlsen[1], *Mona Doberaß*[1], *Lena-Marie Huttner*[1], *André Erwin Radtke*[1]
[1] Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität Munich
j.carlsen@campus.lmu.de, m.doberass@campus.lmu.de, lenamarie.huttner@campus.lmu.de, a.radtke@campus.lmu.de

While the evaluation of voice attractiveness has been the subject of numerous studies, the mere pleasantness of a voice is a topic rarely examined. The concept of voice pleasantness is of significant interest to the field of speech synthesis. Previous research [1], [2], [3], [4] aimed to find the best natural voices which could be used as models upon which to recreate a pleasant synthesized voice. Objective measurements of the model voices should help accomplish the endeavor.

In order to investigate the pleasantness of the sound of continuous speech, we constructed a listening experiment, a partial reproduction of studies already conducted. German native listeners were asked to rate 50 German utterances of continuous speech on *pleasantness*, *expressiveness*, *variability*, *intelligibility*, *clarity* and aptitude of the voice to be used for narration in an *audiobook* on a seven-point scale. They also had the possibility to give a qualitative feedback on each voice. The subjective data collected during the experiment was then matched against the objective measures f0, minimum f0, maximum f0, range, jitter, shimmer, speaking rate and harmony to noise ratio (HNR).

Our first hypothesis should test the assumption that the pleasantness evaluation of a voice will be homogenous among participants: Listeners discriminate pleasant and unpleasant voices (H1). As studies in speech pathology and voice quality have shown, HNR, jitter, and shimmer contribute as measures to a voice being diagnosed as dysphonic [5]. Thus we hypothesize that a voice which would rank higher on a pathological scale, would be rated as less pleasant (H2). Therefore high HNR values will be rated more favorably (H2.1) as will voices with low jitter and shimmer values (H 2.2) [1], [2], [3], [4], [5].

It could be shown that voices are consistently perceived as pleasant or unpleasant by the participants confirming our first hypothesis. However the results for our second hypothesis went against our expectations. Higher jitter and lower harmony to noise values received more favorable scores. Our objective measures were all together only able to explain 15% of the variance of the pleasantness ratings indicating that voice pleasantness is highly complex. A broader approach and the testing of more categories will be necessary to gain a more comprehensive understanding of the subject matter at hand.

Table 1: Correlates between the subjective responses and objective measures

Pearson's r	Syl/s	F0	F0 min	F0 max	Range	Jitter	Shimmer	HNR
Pleasantness	-0,01	**0,23**	**0,19**	0,14	0,05	-0,12	-0,08	**0,16**
Expressiveness	-0,08	**0,18**	**0,19**	0,08	-0,05	**-0,20**	-0,01	**0,16**
Clarity	-0,02	0,01	0,02	-0,03	-0,06	-0,04	0	0,06
Intellegibility	-0,02	0,04	0,04	0	-0,04	-0,06	0	0,07
Variability	-0,13	**0,18**	**0,20**	0,06	-0,09	**-0,24**	-0,01	**0,16**
Audiobook	-0,07	**0,24**	**0,21**	0,13	0,01	**-0,16**	-0,06	**0,16**

References

[1] D. Braga, L. Coelho, F. Gil V. Resende and M. Sales Dias, "Subjective and Objective Evaluation of Brazilian Portuguese TTS Voice Font Quality" *Advances in Speech Technology, 14th International Workshop. June 27-29 2007, Maribor, Slovenia.* 2007.

[2] L. Coelho, D. Braga and C. Garcia-Mateo, "Voice pleasantness: on the improvement of TTS voice quality", in *Jornadas en Tecnología del Habla,* pp. 211-214, 2008.

[3] L. Coelho, H. Hain, O. Jokisch and D. Braga, "Towards an Objective Voice Preference Definition for the Portuguese Language." *Proceedings of the I Iberian SLTech 2009.* pp. 67-70.

[4] L. Pinto-Coelho, D. Braga, M. Sales-Dias, C. Garcia-Mateo, "On the development of an automatic voice pleasantness classification and intensity estimation system", in *Computer Speech and Language*, vol. 27, no. 1, pp. 75-88, 2012.

[5] B. Schneider-Stickler and W.Bigenzahn, *Stimmdiagnostik: Ein Leitfaden für die Praxis*, Wien: Springer-Verlag, 2013.

Sprechgeschwindigkeit in Lehrvideos auf YouTube

Alexandra Ebel[1]
[1]Martin-Luther-Universität Halle-Wittenberg
alexandra.ebel@sprechwiss.uni-halle.de

Die Videoplattform YouTube ist zum Inbegriff für Video im Internet geworden. Längst dient das Portal nicht mehr nur zur Unterhaltung seiner Nutzer, sondern besitzt auch wissensvermittelnde Funktionen. Die Gründe, die hauptsächlich für das Lernen mit YouTube sprechen, sind Unabhängigkeit von Ort und Zeit, eine breite Themenvielfalt, die vom groben Überblick bis zum Detailwissen nahezu alles bereithält, sowie die multisensorische Darstellung der Inhalte. Bislang ist das Genre der Wissens- und Wissenschaftsvideos auf YouTube allerdings noch wenig erforscht, was zum einen an der Personalisierung der Suchergebnisse liegt, die eine systematische Analyse erschweren [1] und zum anderen den methodischen Herausforderungen: Während die Daten aus textbasierten Social Media relativ leicht mithilfe von Web-Crawlern extrahiert werden können, gibt es noch kein Tool, das gezielt das Bewegtbildangebot auf YouTube durchsuchen könnte [2]. Da mittlerweile auch in den Fachdidaktiken verschiedener Disziplinen ein Interesse an YouTube-Videos aufkommt, wächst der Wunsch nach Anhaltspunkten, wie das nutzerseitige Verständnis des präsentierten Wissens sichergestellt werden kann. Es gibt bereits einige wissenschaftliche Auseinandersetzungen, die versuchen, die Merkmale erfolgreicher Videos zu ermitteln, oft bleibt es im Hinblick auf die sprecherisch-sprachliche Gestaltung aber bei Allgemeinplätzen.

Auch die Frage, welche Sprechgeschwindigkeit angestrebt werden sollte, um das Verständnis der Lerninhalte in YouTube-Videos zu unterstützten, wird in der Literatur unterschiedlich beantwortet: So empfiehlt zum Beispiel Becher [3] lediglich generalisierend eine ‚angemessene' Sprechgeschwindigkeit, ebenso tut es Merkt [4]. Auf die Besonderheiten der Plattform YouTube, wie beliebige Reproduzierbarkeit des Schallereignisses oder die Möglichkeit, Videos mit verringerter oder erhöhter Geschwindigkeit abzuspielen, wird dabei jedoch nicht eingegangen. Welbourne / Grant konnten in einer Studie ermitteln, dass Videos mit schnell sprechenden Sprechern erfolgreicher waren, als die, in denen langsamer präsentiert wurde [5] und vermuten als Grund die Wiederholbarkeit der Videos, regen jedoch weitere Studien zu dieser Thematik an.

Um exaktere Aussagen treffen zu können, welche Sprechgeschwindigkeit innerhalb eines YouTube-Lehrvideos als angemessen gelten kann, wurden in einem Korpus aus 40 Videos die Sprech- und Artikulationsgeschwindigkeiten gemessen. Die Daten werden diskutiert und zu den Spezifika des Formats ins Verhältnis gesetzt.

Literatur

[1] J. Allgaier, „Wo Wissenschaft auf Populäkkultur trifft“, in: *Web Video Wissenschaft. play science. Ohne Bewegtbild läuft nichts mehr im Netz: Wie Wissenschaftsvideos das Publikum erobern*, T. Körkel & K. Hoppenhaus (Hrsg.), Heidelberg: Spektrum der Wissenschaft, 2016, o. S.

[2] A. Geipel, „Wissenschaft@YouTube. Plattformspezifische Formen von Wissenschaftskommunikation“, in: *Knowledge in Action. Neue Formen der Kommunikation in der Wissensgesellschaft*, E. Lettkemann, R. Wilke & H. Knoblauch (Hrsg.), Wiesbaden: Springer Fachmedien, 2018, pp. 137–163.

[3] A. Becher, *Lernvideos auf YouTube*, Masterarbeit, Technische Universität Dresden, 2012.

[4] M. Merkt, *Didaktische Optimierung von Videos in der Hochschullehre*, e-teaching.org. https://www.e-teaching.org/news/eteaching_blog/didaktische-optimierung-von-videos-in-der-hochschullehre, 2015.

[5] D. J. Welbourne & W. J. Grant, „Science communication on YouTube: Factors that affect channel and video popularity“, *Public understanding of science*, 25(6), pp. 706–718, (2016).

Junge oder Mädchen?
Zur Geschlechtsidentifikation präpubertärer Stimmen

Riccarda Funk[1], Susanne Voigt-Zimmermann[1], Adrian P. Simpson[1]
[1]Martin-Luther-Universität Halle-Wittenberg
[2]Friedrich-Schiller-Universität Jena
riccarda.funk@web.de, susanne.voigt-zimmermann@sprechwiss.uni-halle.de, adrian.simpson@uni-jena.de

Das auditive Erkennen des Geschlechts einer Person wird durch verschiedene Faktoren beeinflusst. Eine wichtige Rolle spielen sowohl anatomische und akustische Merkmale (Größe des Kehlkopfes, Grundfrequenz, Formanten) als auch sozial gelernte Verhaltensmuster des Sprechenden. Die Unterschiede zwischen Männern und Frauen sind dabei so groß, dass Hörer/-innen „männliche“ und „weibliche“ Stimmen nahezu 100% richtig bestimmen können [1]. Obwohl Jungen und Mädchen vor der Pubertät lediglich geringe Differenzen hinsichtlich ihres Stimmapparates aufweisen [2], belegen einige Studien, dass das Geschlecht eines Kindes nur anhand der Stimme (sprachenunabhängig) ebenfalls signifikant richtig bestimmt werden kann [3] [4] [5] [6]. Interessanterweise zeigt sich zudem ein starker Zusammenhang zwischen wahrnehmbaren Attributen wie Heiserkeit, Sprechgeschwindigkeit, Melodieführung und akustischen Messungen [6].

Die kommende Arbeit wird ebenfalls untersuchen, wie erwachsene Hörer/-innen das Geschlecht präpubertärer Kinder zuordnen. Bisher durchgeführte Studien weisen einige Nachteile auf: Meist waren die untersuchten Kinder älter als sieben Jahre (hier können beginnende anatomische und endokrinologische Veränderungen der Stimme im Zuge der Pubertät nicht ausgeschlossen werden [7]), es wurden nur wenige Aufnahmen angefertigt oder die Anzahl der Befragten im Hörexperiment war sehr gering [3] [4] [5] [6]. Um bisher gewonnene Erkenntnisse zu bestätigen und zu erweitern, ist es nötig, ähnliche Experimente wie bei [3] und [6] mit einer großen Anzahl jüngerer Kinder und vielen Hörer/-innen durchzuführen.

Dafür wurden im Dezember 2017 Aufnahmen von 50 deutschsprachigen Kindern im Alter von vier bis fünf Jahren (Gruppe 1) und sechs bis sieben Jahren (Gruppe 2) angefertigt. Für spontansprachliche Stimuli beschrieben die Kinder einfache Situationsbilder, zudem wurden das Bennenen von Bildkarten, das Zählen bis zehn und das Nachsprechen kurzer Sätze aufgenommen. Mithilfe dieser Aufnahmen sollen zwei Hörexperimente durchgeführt werden: Die erste Hörergruppe beurteilt spontane und vorgegebene Äußerungen aus Gruppe 1, die zweite Hörergruppe beurteilt selbige Stimulitypen aus Gruppe 2. Ziel ist es zunächst, herauszufinden, wie gut das Geschlecht der jeweiligen Kindergruppe identifiziert werden kann und welchen Einfluss Alter und Stimulustyp haben. Aus akustischen Messungen lassen sich weiterhin physikalische Stimmunterschiede (Grundfrequenz, Formanten, HNR, Sprechgeschwindigkeit) feststellen, die in späteren Wahrnehmungsstudien überprüft werden sollen.

Literatur

[1] R.M. Krauss, R. Freyberg, E. Morsella, „Inferring speakers physical attributes from their voices“, *Journal of Experimental Social Psychology*, 38(6), pp. 618-625, 2002.

[2] J. Kahane, „A morphological study of the human prepubertal and pubertal larynx*“, Developmental Dynamics*, 151, pp. 11-19, 1978.

[3] D. Günzburger, A. Bresser, M. Keurs, „Voice identication of prepubertal boys and girls by normally sighted and visually handicapped subjects*“, Language and Speech, 30*, pp. 47-58, 1987.

[4] I. Karlsson, M. Rothenberg„Inter-cultural variations in gender-based language differences in young children“, *Speech research summary report STL-QPSR*, 33, pp. 1-17, 1992.

[5] C. Klein, „Acoustic and perceptual gender characteristics in the voices of pre-adolescent children“, *Phonus, 9*, pp. 221-329, 2004.

[6] A.P. Simpson, R. Funk, F. Palmer, „Perceptual and acoustic correlates of gender in the prepubertal voice“, *Interspeech Stockholm*, pp. 914-918, 2017.

[7] F.M. Biro, M.P. Galvez, L.C. Greenspan, P.A. Succop, N. Vangeepuram, S.M. Pinney, S. Teiteil-baum, G.C. Windham, L.H. Kushi, M.S. Wolff, M.S., „Pubertal Assessment Method and Baseline Characteristics in a Mixed Longitudinal Study of Girls“, *Pediatrics, 126(3)*, pp. 583-590.

Realisations of /ɹ/ in Learner English – The Role of Bilingualism

Kathrin Feindt M.A.
Universität Hamburg
Kathrin.Feindt@googlemail.com

To investigate which linguistic system leaves traces in early bilinguals' foreign language English, the oral performance of 16 pupils in the 9th grade growing up with Turkish and German as first languages and learning English in Germany was examined for realisations of the /ɹ/ phoneme, complemented by a control group in the same age of 18 monolingually raised German participants. In this language

constellation, the alveolar approximant is uniquely used in English and thus equally foreign to participants across groups, as the standard German phoneme inventory contains the uvular fricative /R/, while standard Turkish uses the alveolar tap /ɾ/. In analysing 135 tokens (64 from bilingual and 71 from

monolingual speakers) phonetically with the focus on formant frequencies, surprisingly few

substitutional sounds from either language are found. The vast majority of segments were neither produced with considerable friction noise, as could be expected by negative transfer from German, nor by stopping the pulmonic airflow as associated with the articulation of taps. Bilinguals as well as monolinguals show an equally high performance in target-like production of /ɹ/ with its characteristic low F3 values and the steep raise towards the following vowel [1]. In more than 90% of the cases where the alveolar approximant was aimed at, it was successfully produced. The few non target-like sounds can rather be traced back to general developmental mechanisms of first and foreign language acquisition. As concluded by Wode [2], substituting /ɹ/ with the bilabial approximant /w/ is a universal learning pattern irrespective of age or number of languages spoken; a strategy participants in both groups deployed. Moreover, bilinguals as well as monolinguals not only show the ability to adjust phonetic settings, but were also able to deploy superordinate English phonotactical rules in switching to the allophonic rhotic vowel given the appropriate phonetic surrounding. In a C_r_V environment, 93% of instances in the bilingual and 97% in the monolingual group are realised as /ɹ/. In those cases, where the phoneme is preceded by a vowel and followed by a consonant, 97% of the monolinguals and 93% of the bilinguals produce the rhotic vowel, in compliance to native speakers of RP. At least in this particular area of English phonology, bilinguals show no clear preference for either of the known languages. In addition, studying the role of early bilingualism for foreign language acquisition reveals that phonological development is not solely driven by transfer effects but that universal processes for language learning have an additional influence learner's oral performance.

Literatur

[1] Ladefoged, Peter (2004): Vowels and consonants. An introduction to the sounds of languages. Malden, MA: Blackwell

[2] H. Wode, “Perception and Production in learning to talk.” In: *Focus on phonological acquisition.* S. J. Hannahs and M. Young-Scholten (Hrsg.), Amsterdam: Benjamin, 1997, pp. 17-47.

The prosody of Yes-No questions in Pulaar

Susanne Genzel, Department of Linguistics, University of Potsdam, Germany
Frank Kügler, Institute of Linguistics, Goethe-University Frankfurt &
Institute of Linguistics – Phonetics, University of Cologne, Germany

In this paper, we compare the prosodic realization of string-identical statement/Yes-No questions pairs in the experimentally understudied language Pulaar. Results show that Yes-No questions are realized with a high boundary tone at the right edge of the utterance (H%) and a raised register.

Fulfulde or Fula is a language of the West-Atlantic branch of the Niger-Congo family. It is spoken by 13 million speakers in 17 African countries from Senegal to Sudan [1], [2]. The name Pulaar is used in Senegal and will be used throughout the paper. The paper reports data from the dialect Futa Toro spoken in Northern Senegal. The basic word order is SVO and the language is non-tonal [3]. [4] show that stress is initial. Pulaar is related to Wolof; see [5] for an overview of the intonational system. [6] investigating Maasinankoore Fulfulde observes that statements show pitch lowering on each stressed syllable with the final syllable having the lowest pitch. To the best of our knowledge, there is no experimental study investigating the default prosody of Pulaar, but see [7] for phonetic effects of focus. Concerning Yes-No questions, [6] observes that the pitch on each stressed syllable in Yes-No questions is rising. [8] classifies Fula, based on [6], as having a H% and register expansion/reduction of downdrift.

We recorded 108 sentences (3 (1 female, 2 male) speakers x 6 items x 2 sentence modi x 3 repetitions) in Linguère. Materials consist of 2 SV, 2 SVDO and 2 SVIODO string-identical statement/Yes-No question pairs. An SV test sentence is presented in (1). The abbreviation PTCP refers to participle. Participants were recorded digitally with a headset on a laptop.

(1) Amel mb-on-i yim-a
Amel 3SG-be-PRS.SG sing-PTCP
Amel sings./?

We extracted time-normalized F0 in Hz with Praat [9] for each word. The investigation of further phonetic parameters is planned.

As shown in figure 1, the F0 contour of the statement consists of a rising pitch movement on the subject, contrary to [6]. The peak is generally located on the second syllable also in polysyllabic subject constituents. After the peak, the F0 falls towards a low target at the beginning of the VP. For the female speaker, the F0 rises smoothly until the end of the sentence. The sentence terminates breathy. For the male speakers, in line with [6], F0 falls smoothly to the end of the sentence and show a final falling F0 and no breathy termination, everything else being equal. [5] describe a similar contour for Wolof. Yes-No questions share the same intonational events, however, as shown in figure 1, they terminate super-high and the pitch register is raised. The findings are broadly in line with the observation of [6] and the classification of [8].

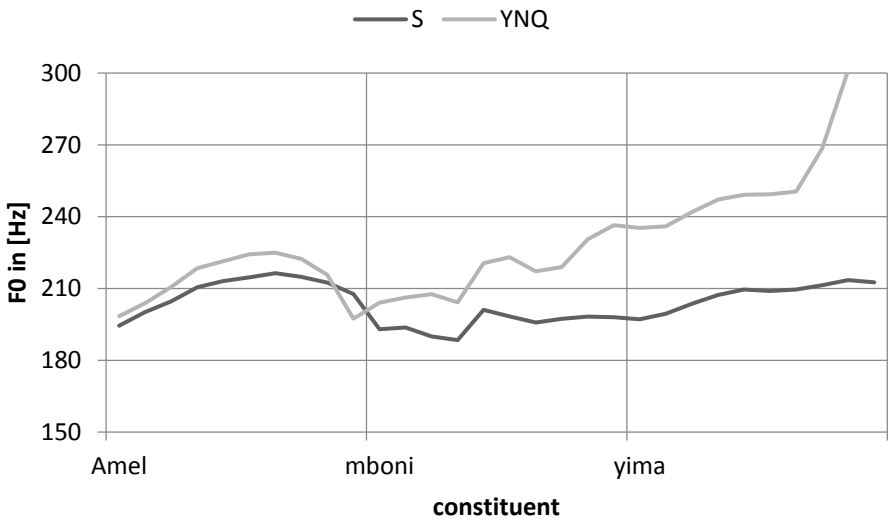


Figure 1: *Time normalized course of mean F0 in Hz (female speaker) of the sentence in (1), black line – statement (S) & grey line – Yes-No question (YNQ), vertical lines indicate constituent boundaries.*

Literatur

[1] G. F. Simons & C. D. Fennig (eds.), *Ethnologue: Languages of the World*, Twentieth edition. Dallas, Texas: SIL International, online version: http://www.ethnologue.com, 2017.

[2] A. Harrison, *Fulfulde Language Family Report*, SIL Electronic Survey Reports 2003-009, http://www-01.sil.org/silesr/2003/silesr2003-009.html, 2003.

[3] D. W. Arnott, *The nominal and verbal systems of Fula*, Oxford: The Clarendon Press, 1970.

[4] F. Laughlin & C. Wiltshire, "Syllable weight in the phonology of Pulaar", in *Syllable Weight in African Languages*, P. Newman (ed.), John Benjamins, 2017, pp. 161–176.

[5] A. Rialland & S. Robert, "The intonational system of Wolof", *Linguistics*, 39(5), pp. 893–939, 2001.

[6] J. O. Breedveld, *Form and meaning in Fulfulde: A morphophonological study of Maasinankoore*, Leiden: Research School CNWS, 1995.

[7] B. Mingzhen, "Prosody of Focus among Information Structures in Pulaar", *Afrikanistik online*, Vol. 2012, http://www.afrikanistik-aegyptologie-online.de/archiv/2012/3528, 2012.

[8] A. Rialland, "Question prosody: an African perspective", in *Tones and tunes, vol.1: Typological studies in word and sentence prosody*, T. Riad & C. Gussenhoven, Berlin, New York: Mouton de Gruyter, 2007, pp. 35–64.

[9] P. Boersma & D. Weenink, *Praat: Doing phonetics by computer* version 5.4.05, [Computer program], http://www.praat.org/, 2015.

Voice quality of ironic utterances

Hannah Leykum
Acoustics Research Institute, Austrian Academy of Sciences, Vienna
hannah.leykum@oeaw.ac.at

Irony is regularly used in everyday communication. Irony cues can be verbal, nonverbal, or paraverbal. If both verbal and nonverbal cues are missing in interaction (e.g., ambiguous utterances on the telephone), listeners have to rely on paraverbal cues in order to understand irony. The risk of misunderstanding of ironic expressions is reduced if the speakers apply disambiguating (e.g. paraverbal) cues. The most prominent paraverbal cues to indicate verbal irony are average fundamental frequency (F0), F0-contour, intensity, and duration. Acoustic analyses performed on German utterances of sarcasm/ironic criticism showed a lower average F0, smaller F0 variation, and longer segment durations in verbal irony [1–3]. Concerning voice quality, Niebuhr [4] found a more variable, mainly breathier or tenser, voice quality for ironic utterances. In addition, Schmiedel [1] revealed a higher Harmonics-to-Noise Ratio (HNR) for sarcastic utterances. Until now, no investigations of voice quality in ironic utterances of speakers of Standard Austrian German (SAG) exist.

The present study is part of a larger project investigating on the one hand the acoustic characteristics of verbal irony and on the other hand the irony recognition by listeners, when context-free stimuli are presented. For the production part, 20 speakers (two age groups, balanced for gender) of SAG are presented with 20 scenarios that evoke short utterances either in the literal sense or in an ironic sense. The speakers are instructed to read the scenarios and the answers to the scenarios while audio, video, and electroglottographic (EGG) recordings are conducted. In addition to the acoustic analysis of the audio recording, the EGG signal is analysed to get objective measurements for voice quality.

First analyses indicate that creaky voice and breathiness occur more often in ironic realisations of utterances compared to their literal counterparts. In the EGG signal, creaky voice is characterised mainly by a relatively short rise time, a low opening quotient and irregular pulses [5]. Breathiness has a low peak amplitude and a long opening phase in the EGG signal [5]. In Figure 1, the literal (upper part) and the ironic (lower part) realisation of the utterance “unglaublich” (“*incredible*”) by one male speaker is shown. In the lower part of Figure 1 the breathy beginning of the ironic utterance is clearly visible in the EGG signal.

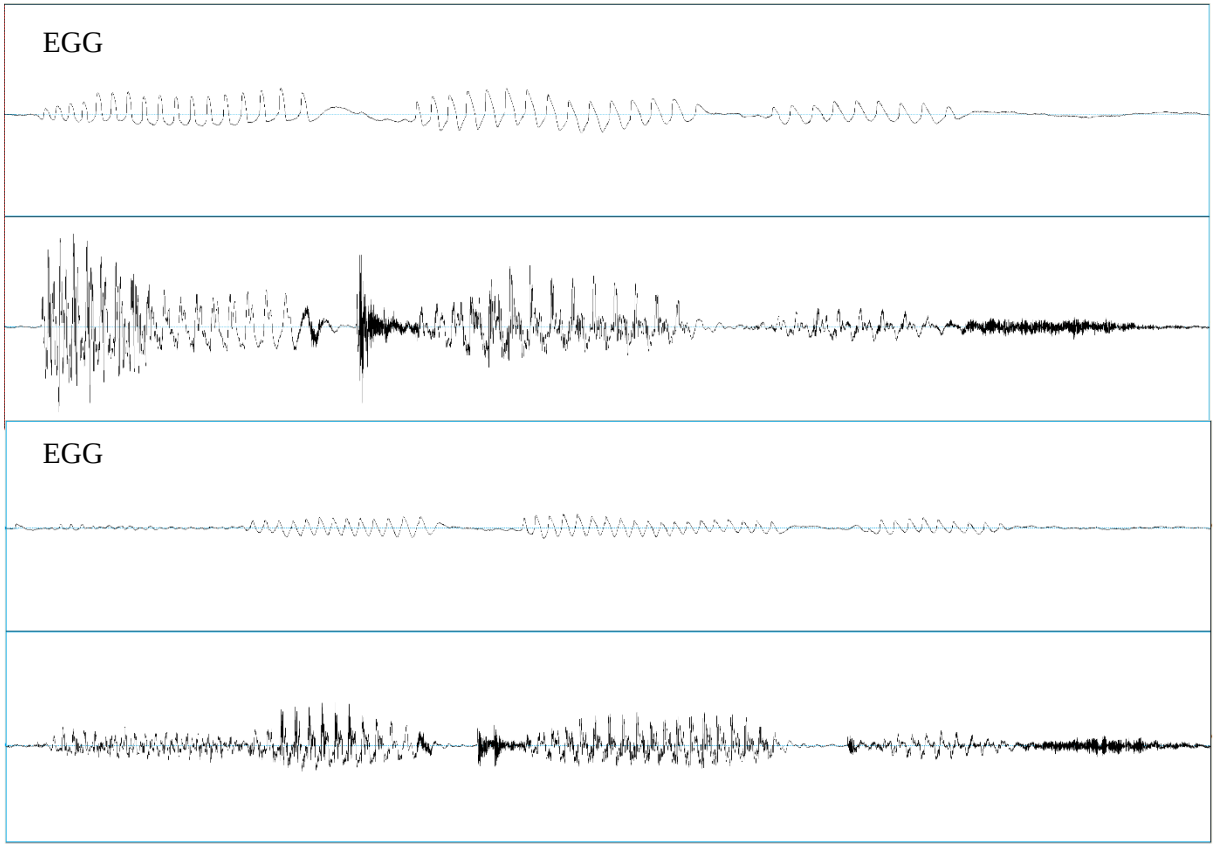


Figure 1*: "unglaublich" (“incredible”) literal (upper part) vs. ironic (lower part): EGG signal and waveform of the audio recording*

References

[1] A. Schmiedel, *Phonetik ironischer Sprechweise: Produktion und Perzeption sarkastisch ironischer und freundlich ironischer Äußerungen*, Berlin; Frank & Timme, 2017.

[2] A. Nauke & A. Braun, “The production and perception of irony in short context-free utterances,” *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, pp. 1450–1453, 2011.

[3] L. Scharrer, U. Christmann & M. Knoll, “Voice Modulations in German Ironic Speech,” *Language and Speech*, 54(4), pp. 435–465, 2011.

[4] O. Niebuhr, “"A little more ironic" – Voice quality and segmental reduction differences between sarcastic and neutral utterances,” *7th International Conference of Speech Prosody*, pp. 608–612, 2014.

[5] K. Marasek, “Electroglottographic Description of Voice Quality,” Habilitation, University of Stuttgart, Stuttgart, 1997.

German dorsal fricative assimilation (DFA) *revisited*

Kathrin Linke[1], Francesc Torres-Tamarit[2]
[1]University of Vienna
[2]CNRS/Paris 8
kathrin.linke@univie.ac.at, francescjosep.torres@gmail.com

We present the well-known case of *ich-laut* vs *ach-laut* alternation in the Standard German varieties spoken in Austria and Germany. Whereas the Austrian variety shows a transparent pattern, we show that the German variety presents some interactions with related phonological processes which lead to opacity in such a way that is problematic for derivational versions of Optimality Theory (OT). We will illustrate that OT is, nevertheless, a useful tool to study typologies of dialectal variation.

1 | The issue. The German consonantal segment inventory provides a voiceless dorsal fricative with a velar and a palatal allophonic variant as illustrated in *Buch* [buːx] 'book' and *Bücher* [byːçɐ̯] 'books'.

The dorsal fricative surfaces as the velar variant [x] following non-front vowels and as the palatal variant [ç] after front vowels and, additionally, sonorant coronal consonants, which suggests that they are regressively assimilated to the tautomorphemically preceding front or non-front segment.

2 | Interacting processes leading to opacity. A second allophonic process in German is /ʀ/-vocalisation, in which /ʀ/ is replaced by [ɐ̯] in syllable codas (Itô & Mester 2001). In Standard German spoken in Austria, vocalised /ʀ/ is (transparently) followed by the velar variant [x] like in *durch* [dʊɐ̯x] 'through'. However, in Standard German spoken in Germany, vocalised /ʀ/'s create incidences in which the palatal fricative opaquely surfaces after 'back' segments like in *durch* [dʊɐ̯ç] 'through'.

In addition, in Standard German spoken in Germany, DFA also interacts with the phonological processes of /g/-spirantisation and coda devoicing, which account for the differences in word pairs such as *König* [køːnɪç] 'king' / *Könige* [køːnɪgə] 'kings'. So, underlying /g/ in syllable codas with nucleus /ɪ/ are spirantised, opaquely surfacing as the palatal fricative [ç] (Hall 1989, Itô & Mester 2001).

3 | Account using derivational versions of OT. Firstly, we will discuss OT based approaches to DFA in German and their accounts for the opaque results of the interacting processes mentioned above. It will be shown that OT with Candidate Chains (OT-CC) can account for the case of double opacity in Standard German because Precedence constraints emulate rule ordering within candidate chains. Harmonic Serialism (HS), however, cannot account for phonological opacity. A representational solution based on Turbidity Theory will also be entertained.

4 | Typological considerations. We will evaluate the DFA patterns and interacting phonological processes found in the Standard varieties of German spoken in Austria and Germany and show that OT provides a helpful tool to account for dialectal variation. Additional data from regional dialects will be used to substantiate the proposed analysis and to derive a typology of the interactions of DFA with /g/-spirantization and /ʀ/-vocalization in German.

References

[1] T. A. Hall, “Lexical phonology and the distribution of German [ç] and [x]”, *Phonology* 6(1), pp. 1–17, 1989.

[2] J. Itô & A. Mester, “Structure preservation and stratal opacity in German”, in *Segmental phonology in Optimality Theory: constraints and representations*, L. Lombardi (Ed.), Cambridge: Cambridge University Press, 2001, pp. 261–295.

[3] J. J. McCarthy, *Hidden generalization: phonological opacity in Optimality Theory*. London; Oakville, CT: Equinox Pub, 2006.

[4] J. J. McCarthy, “The theory and practice of Harmonic Serialism”, in *Harmonic Grammar and Harmonic Serialism*, J. J. McCarthy and J. Pater (Eds.), London: Equinox, 2016, pp. 47–87.

Changes in vowel stem articulation due to suffixation – Investigating the effects of practice

Fabian Tomaschek[1], Harald Baayen[1] and Benjamin V. Tucker[2]
[1] University of Tübingen, [2] University of Alberta
fabian.tomaschek@uni-tuebingen.de, harald.baayen@uni-tuebingen.de, bvtucker@ualberta.ca

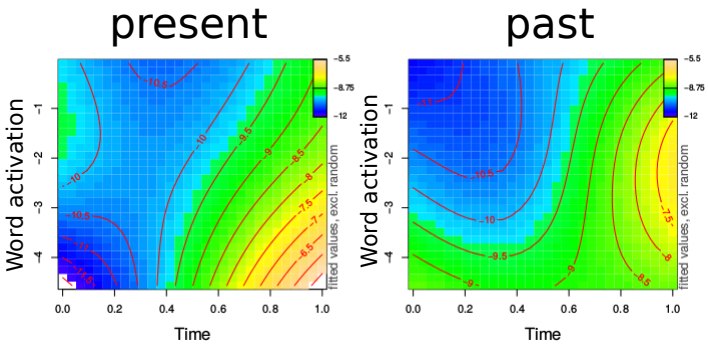


Figure 1: *Height of tongue body sensor as a function of time (x-axis) and word activation during vowel articulation in present (left) and past-tense (right) verbs. Warmer colors indicate higher vertical positions.*

Evidence has arisen showing that phonologically homophonous units such as stems and phones differ in their fine phonetic detail with respect to their lexical, semantic and morphological structures (e.g. Plag, Homann, and Kunter 2017; Lee-Kim, Davidson, and Hwang 2012; Drager 2011; Kemps et al. 2005a). In the present study, we investigated the articulatory processes behind these fine phonetic details in stems undergoing morphological changes in suffixed and unsuffixed AE verbs (e.g. "clean" vs. "cleans", "cleaned", "cleaning"). We built our hypotheses on two pillars: The first is in regards to consistent findings which state that phonologically identical phones show varying fine phonetic details due to anticipatory coarticulation (e.g. Öhman 1966). The second pillar is that kinematic skills improve with practice, reflected among others by a stronger overlap of consequtive movement gestures (Sosnik et al. 2004). Consequently, we hypothesized that articulatory movements of verbal stem vowels should differ systematically between suffixed and unsuffixed verbs and that coarticulation of the suffix should be stronger in more practiced verbs. Using electromagnetic articulography, we recorded tongue movements of 25 native speakers of Canadian English, each uttering 210 suffixed and unsuffixed verbs containing the vowels [ɑː] and [iː] in their stems. We parameterized practice of words by means of word activations calculated by the Naïve Discriminative Learner (Baayen et al. 2011).

Analyses of tongue body and tongue tip movements revealed strong effects of coarticulation and word activation. Figure 1 exemplarily illustrates the regression surfaces of

tongue body height in the stem form (left) and the past tense (right) during [ɑː] articulation. In present verbs, a lower tongue position is reached later and maintained longer for more practiced than for less practiced words. The past tense suffix (Figure 1, right) modifies the trajectory with more practiced words showing a deeper articulation of the [ɑː] and a raising towards the dental [t] suffix much earlier.

Our findings show that suffixes strongly affect articulatory patterns in homophonous stems and that they are modulated by the amount of practice speakers have with a word. Contradicting modular models of speech production (Levelt, Roelofs, and Meyer 1999), these findings have important implications for models of the phonology-morphology interaction.

References

[1] R. H. Baayen et al. "An amorphous model for morphological processing in visual comprehension based on naive discriminative learning." In: *Psychological review* 118.3 (2011), pp. 438–481.

[2] Katie K. Drager. "Sociophonetic variation and the lemma." In: *Journal of Phonetics* 39.4 (2011). Cross-Language Speech Perception and Variations in Linguistics Experience, pp. 694–707.

[3] R. J. Kemps et al. "Prosodic cues for morphological complexity in Dutch and English." In: *Language and Cognitive Processes* 20.1/2 (2005a), pp. 43–73.

[4] Sang-Im Lee-Kim, Lisa Davidson, and Sangjin Hwang. "Morphological effects on the darkness of English intervocalic /l/." In: *Laboratory Phonologya* 4.2 (2012), pp. 475–511.

[5] W. J. Levelt, A. Roelofs, and A. S. Meyer. "A theory of lexical access in speech production." In: *The Behavioral and brain sciences* 22.1 (Feb. 1999).

[6] S.E.G. Öhman. "Coarticulation in VCV Utterances: Spectrographic Measurements." In: *Journal of the Acoustical Society of America* 39.151 (1966), pp. 151–168.

[7] I. Plag, J. Homann, and G. Kunter. "Homophony and morphology: The acoustics of word-final S in English." In: *Journal of Linguistics* 53.1 (2017), pp. 181–216.

[8] R. Sosnik et al. "When practice leads to co-articulation: the evolution of geometrically defined movement primitives." In: *Exp Brain Res* 156 (2004), pp. 422–438.

Zu phonetischen Partikeln in "stillen" Pausen

Jürgen Trouvain & Bernd Möbius
Sprachwissenschaft und Sprachtechnologie, Universität des Saarlandes, Saarbrücken
trouvain|moebius@coli.uni-saarland.de

Eine häufig vorgenommene Einteilung bei Sprechpausen ist diejenige in so genannte "gefüllte" und "ungefüllte" Pausen. "Gefüllte" Pausen beziehen sich auf eine Füll-Silbe wie [@] oder [@m], wohingegen Pausen, bei denen solche Füllsel fehlen gemeinhin als "stille" Pausen bezeichnet werden [1]. Aus einer phonetischen Perspektive, die auch sprechphysiologische Vorgänge betrachtet, sind viele dieser "stillen" Pausen mit Atmungsgeräuschen, aber eventuell auch mit anderen non-verbalen Vokalisierungen angereichert (um den Begriff "gefüllt" an dieser Stelle zu vermeiden). Es ist daher aus einer phonetischen Warte relevant und notwendig, die so genannten "stillen" Pausen phonetischer genauer zu betrachten.

Die Atemgeräusche können (mehr oder minder verlässlich) der Ein- oder der Ausatmung zugeordnet werden. Zudem kann auch die Beteiligung des oralen Raumes beim ingressiven oder egressiven Luftstrom unterschieden werden, auch wenn es den Anschein hat, dass die allermeisten Atemgeräusche durch nasale Einatmung hervorgerufen werden [2]. Zu den wichtigen Beiträgen von Atmungsgeräuschen gehört es, den Zuhörer über die zu erwartende Länge der nachfolgenden prosodischen Phrase zu informieren [3] oder die Memorierbarkeit des danach Gesprochenen zu erhöhen [4]. Darüber hinaus können auch noch weitere phonetische Partikeln in "stillen" Pausen beobachtet werden, wie z.B. Schnalzlaute mit der Zungenspitze [5] oder auch Lachen [6]. Hierzu gibt es einen Forschungsbedarf, da nicht anzunehmen ist, dass alle denkbaren nicht-verbalen Vokalisierungen [7] gleich häufig in Sprachdaten und mit ähnlichen Funktionen zu finden sind. Es ist aber anzunehmen, dass in verschiedenen Sprechregistern verschiedene pausen-interne Partikeln vorkommen.

Für die vorliegende Pilotstudie wurden Sprachproben aus sieben verschiedenen Sprechstilen bzw. Sprechsituationen ausgewählt (Radionachrichten, Erzählung, spontane Nacherzählung, Parlamentsrede, Verdolmetschung, Vortrag, Dialog). Jede Sprachprobe hat die Dauer von genau einer Minute und wurde nach dem Zufallsprinzip aus vorliegenden Datensammlungen ausgewählt. Die ersten Ergebnisse zeigen, dass Dauer der Pause und Dauer des Atemgeräuschs korrelieren. Dabei gibt es aber etliche Pausen, die länger als 500 ms sind, die aber **kein** Atemgeräusch zeigen, bei denen aber sehr wahrscheinlich Einatmung stattgefunden hat. Dieser Befund zeigt die Lücke zwischen akustischer und auditiver Beobachtung einerseits und physiologischer Bewegung andererseits. Ein weiteres Ergebnis ist, dass die zeitliche Variation der mittleren Einatmungsdauer sich zwischen 270 und 420 ms bewegt, die mittlere Pausendauer pro Sprecher hingegen zwischen 400 und 900 ms. Dabei sind auch zwischen Textsorten mit Lesesprache große Unterschiede festzustellen. Füll-Silben kommen nur in unvorbereiteter Sprache vor, Schnalzlaute auch in gelesener Sprache. Die Ergebnisse veranschaulichen, dass "stille Pausen" auf Grund der darin enthaltenen phonetischen Partikeln alles andere als eine homogene Klasse darstellen.

Literatur

[1] F. Goldman Eisler, *Psycholinguistics: Experiments in Spontaneous Speech*, London: Academic Press, 1968.

[2] M. Kienast & F. Glitza, "Respiratory sounds as an idiosyncratic feature in speaker recognition", *Proceedings of the 15th International Congress of the Phonetic Sciences* (ICPhS), Barcelona, pp. 1607-1610, 2003.

[3] S. Fuchs, C. Petrone, J. Krivokapić, & Ph. Hoole, "Acoustic and respiratory evidence for utterance planning in German", *Journal of Phonetics* 41, pp. 29-47, 2013.

[4] D.H. Whalen, Ch.E. Hoequist & S. Sheffert, "The effects of breath sounds on the perception of synthetic speech", *Journal of the Acoustical Society of America* 97, pp. 3147-3153, 1995.

[5] R. Ogden, "Forms and functions of clicks in English conversation", *Journal of the International Phonetic Association*, 43(3), pp. 299-320, 2013.

[6] J. Trouvain, "Laughing, breathing, clicking — the prosody of nonverbal vocalisations," *Proceedings of the Conference on Speech Prosody* (SP7), Dublin, pp. 598-602, 2014.

[7] J. Trouvain & K. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," *Proceedings of the 4th Int'l Workshop on Corpora for Research on Emotion Sentiment & Social Signals*, Istanbul, pp. 36-39, 2012.

Automatic Acoustic Prominence Annotation: Re-Implementing a Tool for Model Exploration and Refinement

Jana Voße[1,2], Petra Wagner[1,3]
[1]Phonetics and Phonology Work Group, Bielefeld University, Germany
[2]University of Gothenburg, Sweden
[3]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany
j.vosse@uni-bielefeld.de, petra.wagner@uni-bielefeld.de

Prominence is an essential concept to manifold research areas within linguistics, such as information structure [1][2], phonetic variation [3], and word disambiguation [2]. With this, prominence also plays a significant role in technological contexts, as its parameters influence speech understanding and the generation of natural-sounding speech [2] to a great extent. As it is costly in terms of time to annotate prominence phenomena manually [4], the automatic detection and annotation of such events is crucial to train speech recognition and speech synthesis systems on databases of a sufficient size [1].

With this paper, we contribute to the state of the art of automatic data processing by providing a re-implemented version of an existing prominence tagger. This tool is originally presented by [1] and is specialized for the automatic detection and annotation of prominent syllables in German. For this, the tool defines prominence as a continuous rather than a categorical parameter and considers acoustic features exclusively. In the algorithm, these features represent an interplay between the concepts of pitch accent and force accent defined by Kohler [5], where the contribution of each of the two accent types is weighted by language specific constants [6]. Although the perception of prominence is assumed to be guided by both, acoustic features and linguistic expectancies, the tool delivers good predictions [1].

In order to make the prominence tagger more accessible to the community and more inviting in the manipulation of features, we will provide a re-implemented version of this tool, accessible as a praat script. With this, we do not aim to implement additional features into the model in [1], but rather to provide a simplified way for experimentation with attributes of the tagger's model, e.g. to adapt it to new languages or to investigate new features, and to contribute to the further development of acoustic prominence modeling.

In [1], the prominence tagger has been tested on a read speech corpus exclusively. We will refine the evaluation of this tagger by additionally testing it on a corpus of semi-spontaneous speech representing different levels of motivational speech. The corpus consists of motivational speeches of six different female speakers, summing up to 51 minutes of speech. Choosing these corpora for the evaluation of the prominence tagger will enable us to compare the tagger's performance across different speaking styles. The evaluation is currently under preparation.

Literatur

[1] F. Tamburini and P. Wagner, "On Automatic Prominence Detection for German", *Proceedings of Interspeech 2007*, 2007, pp. 1809-1812.

[2] O. Kalinli, S. Narayana, "Prominence detection using auditory attention cues and task-dependent high level information", *IEEE Transactions on audio, Speech, and language processing*, 17(5), 2009, pp. 1009-1024.

[3] G. Abete, F. Cutugno, B. Ludusan, A. Origlia, "Pitch behavior detection for automatic prominence recognition", *Speech Prosody 2010-Fifth International Conference*, 2010.

[4] M. Avanzi, A. C. Simon, J. P. Goldman, A. Auchlin, "C-prom: An annotated corpus for french prominence study", *Speech Prosody 2010-Fifth International Conference*, 2010.

[5] K. J. Kohler, "Form and Function of Non-Pitch Accents", *Prosodic Patterns of German Spontaneous Speech*, AIPUK, 35a, 2005, pp. 97-123.

[6] F. Tamburini, "Automatic detection of prosodic prominence in continuous speech." In *LREC*. 2002.

[7] P. Wagner, "Great expectations-introspective vs. perceptual prominence ratings and their acoustic correlates", *Ninth European Conference on Speech Communication and Technology*, 2005.

Adaptation of Harmonic Spectrum-based Voice Quality Parameters in Spontaneous Dialogues

Wolfgang Wokurek, Antje Schweitzer
Institut für Maschinelle Sprachverarbeitung, Uni-Stuttgart
{wokurek,schweitzer}@ims.uni-stuttgart.de

The adaptation of voice quality between dialogue partners in natural spontaneous conversations is investigated using data from the GECO [1][2] database. While adaptation with regard to a variety of phonetic parameters have often been investigated, voice quality has mostly been neglected by work in this area. Each dialog in the database is temporally split in two halves of equal duration, and a set of harmonic spectrum-based voice quality parameters are compared by MANOVA. Only voiced /a/-like vowel segments are analyzed which occurred during pauses of the dialogue partner. The voice quality parameter set consists of open quotient, glottal opening, skewness, rate of closure and incompleteness of closure. They have been named after temporal phenomena by [3] but are defined as spectral amplitude decay gradients between the fundamental and selected higher harmonics. Only the incompleteness of closure parameter is a relative formant bandwidth. The vocal tract transfer function is estimated by formant frequencies and bandwidths and subtracted from the harmonic amplitudes to approximate source parameters. All processing steps aim to reduce the effect of changes of articulation and intonation on the voice quality parameters. Most speakers show significantly different voice quality parameters in the two halves of each dialog, but with different amounts of change. The amount of change seems to vary with the dialogue partner. It is investigated whether the voice quality parameter sets of the dialogue partners differ in each dialogue half, and whether this difference decreases or increases in the second half of the dialogue, i.e. whether it signals phonetic convergence or divergence in terms of voice quality.

Literatur

[1] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013, Lyon)*, 2013, pp. 525–529.

[2] ——, "Social factors in convergence of f1 and f2 in spontaneous speech," in *Proceedings of the 10th International Seminar on Speech Production, Cologne*, 2014.

[3] K. M. Stevens and H. M. Hanson, "Classification of glottal vibration from acoustic measurements," in *Vocal Fold Physiology*, O. Fujimura and M. Hirano, Eds. Cambridge MA: Hiltop University Press, 1998, pp. 147–170.

Methodische Betrachtungen zur Erschließung großer Datenmengen in der Phonetik

Denis Arnold[1], Fabian Tomaschek[2]
[1]Institut für Deutsche Sprache, Mannheim
[2]Eberhard Karls Universität Tübingen, Tübingen
arnold@ids-mannheim.de, fabian.tomaschek@uni-tuebingen.de

In allen Bereichen der Wissenschaft ergeben sich durch immer größer werdende Datenmengen neue, spannende Möglichkeiten des Erkenntnisgewinns, aber auch immer größere Herausforderungen bei der Erschließung des Datenschatzes. Bei der Erschließung bedienen wir uns immer öfter verschiedener Algorithmen zur Segmentierung des Signals oder zum Messen verschiedener Parameter des Sprachsignals. Es ist allgemein bekannt, dass verschiedene Algorithmen zu unterschiedlichen Ergebnissen kommen und auch gewisse Fehler machen. Hess hat dies am Beispiel der Grundfrequenzmessung eindrucksvoll dargestellt [1]. In der Regel wird jedoch davon ausgegangen, dass die Unterschiede, sofern die Algorithmen mit sinnvollen Annahmen benutzt wurden, nicht zu entscheidenden Unterschieden in den Ergebnissen führen. Es ergibt sich jedoch die Frage, in wie fern beispielsweise die Segmentierung zweier Korpora mittels unterschiedlicher Aligner vergleichbar sind.

In diesem Beitrag soll der Frage nachgegangen werden, in wie fern die Unterschiede in den Resultaten verschiedener Werkzeuge tatsächlich nicht zu signifikanten Unterschieden führen. Dazu wollen wir bestehende Annotationen mit zwei Forced-Alignern [2,3] neu segmentieren lassen. Auf Basis der verschiedenen Segmentgrenzen wollen wir mit verschiedenen gebräuchlichen Werkzeugen z.B. [4,5] verschiedene Parameter extrahieren und die Ergebnisse vergleichen.

Die Ergebnisse sollen dabei die Frage beantworten, ob sich beispielsweise Korpora mit Segmentierungen auf der Basis verschiedener Aligner beliebig für neue Studien kombinieren lassen. Sind die Unterschiede in den ausgegeben Werten der verschiedenen Algorithmen größer als die Unterschiede durch Variationen in den Parametern. Können wir abschätzen, wie vergleichbar Studien auf der Basis unterschiedlicher Werkzeuge sind.

Literatur

[1] HESS, Wolfgang. *Pitch determination of speech signals: algorithms and devices*. Springer Science & Business Media, 2012.
[2] KISLER, Thomas; REICHEL, Uwe; SCHIEL, Florian. Multilingual processing of speech via web services. *Computer Speech & Language*, 2017, 45. Jg., S. 326-347.
[3] RAPP, Stefan. Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models. An Aligner for German. 1995.
[4] BOERSMA, Paul. Praat: doing phonetics by computer. *http://www. praat. org/*, 2018.
[5] BARREDA, Santiago. phonTools: Functions for phonetics in R. *R package version 0.2-2.1*, 2015.

Voice quality as a function of information density and prosodic factors

Erika Brandt, Bistra Andreeva, Bernd Möbius
[1]Language Science and Technology, Saarland University, Saarbrücken
[ebrandt,andreeva,moebius@coli.uni-saarland.de]

This study investigated the influence of information density (ID) on cepstral peak prominence (CPP) and CPP-smoothed (CPPS) in German content words. CPP measures the difference in amplitude (in dB) between the cepstral peak and the corresponding fundamental quefrency. CPP and CPPS correlate well with perceived breathiness and hoarseness [3, 5]. Speech signals with well-defined harmonic structure show prominent peaks, and thus higher CPP(S) values than signals with less well-defined harmonic structure [4]. We expected to find higher CPP values in vowels that were difficult to predict from the context and that appeared in low-frequency words. As controls, primary lexical stress, prosodic boundary, articulation rate, average vowel duration, and sentence position were used.

Vocalic segments (n = 40,203) of the Siemens Synthesis corpus (SI1000P) [7] were fed into the CPPS analysis tool [5] using the default settings for sustained vowels. CPP is calculated every 10 ms, and then averaged for every signal. CPPS is measured every 2 ms and then averaged and smoothed in cepstral magnitude across quefrency bins.

ID factors were surprisal ($S(unit_i) = -log_2 P(unit_i|context)$) and word frequency. Surprisal values for the preceding and following context were calculated from an *n*-phone language model using SRILM [8]. As a text corpus for language modeling and word frequency counting, SDeWaC was preprocessed using the g2p tool in German-Festival [1]. Articulation rate (phonemes / second) was calculated excluding pauses on the sentence (global) and word level (local). Primary lexical stress (stressed vs. unstressed) was based on the canonical transcription of the SI1000P corpus. Boundary was defined as word, phrase or no boundary. Statistical analysis was performed with lme4 [2] and lmerTest [6].

We found a significant positive effect of biphone surprisal of the preceding context on CPP(S), and a significant effect of triphone surprisal of the following context for CPP. There was only a tendency for a negative effect of word frequency. Vowels immediately preceding both boundary positions showed significantly lower values in both CPP and CPPS. Primary lexical stress was not significant, however, in interaction with triphone surprisal it had a positive effect on both metrics. Vowels in sentences at fast global speech rate showed lower CPP(S) values than at slow tempo. The opposite effect was observed for local speech rate. Average vowel duration had a significant strong positive effect on CPP. For CPPS, however, we only found a tendency for this effect. This result was due to the durational averaging that was part of the smoothing procedure. As expected, vowels in the last word of a sentence showed less well-defined harmonic structure than vowels in words with non-final position. To conclude, ID and voice quality were related, while controlling for other variables: Vowels that were difficult to predict showed less breathiness and hoarseness than easily predictable vowels.

References

[1] Marco Baroni and Adam Kilgarriff. “Large Linguistically-processed Web Corpora for Multiple Languages”. In: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. EACL ’06. Trento, Italy: Association for Computational Linguistics, 2006, pp. 87–90.

[2] Douglas Bates et al. “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48.

[3] Yolanda D. Heman-Ackah, Deirdre D. Michael, and George S. Goding. “The relationship between cepstral peak prominence and selected parameters of dysphonia”. In: *Journal of Voice* 16.1 (2002), pp. 20–27.

[4] James Hillenbrand, Ronald A. Cleveland, and Robert L. Erickson. “Acoustic Correlates of Breathy Vocal Quality”. In: *Journal of Speech Language and Hearing Research* 37.4 (1994), p. 769.

[5] James Hillenbrand and Robert A. Houde. “Acoustic Correlates of Breathy Vocal QualityDysphonic Voices and Continuous Speech”. In: *Journal of Speech, Language, and Hearing Research* 39.2 (1996), pp. 311–321.

[6] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. “lmerTest Package: Tests in Linear Mixed Effects Models”. In: *Journal of Statistical Software* 82.13 (2017), pp. 1–26.

[7] Florian Schiel. *Siemens Synthesis Corpus - SI1000P*. 1997. URL: `https://www.phonetik.uni-muenchen.de/Bas/BasSI1000Peng.html`.

[8] Andreas Stolcke. “Srilm — an Extensible Language Modeling Toolkit”. In: *Proceedings of Interspeech* Denver, Colorado (2002), pp. 901–904.

Abstract für die P&P 14 – Wien, 6.-7. September 2018

Olga Buchmüller, Minying Feng, Sabrina Knott, Marco Krämer, Frederik Palmer, Elizabeth Pankratz, Sophie Reule, Zarina Rizaeva, Alina Zöllner and Christine Mooshammer
Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin
christine.mooshammer@hu-berlin.de

The influence of phonemic similarity on speech errors and variability in CVC syllables

Speech errors can offer insights into the workings of our phonetic and phonological systems. Traditional studies assume that substitution errors are the most frequent error type, resulting a complete segment (e.g. [3, 10]). Recent research has indicated, however, that putative substitution errors are rarely cases of "either/or" [4, 5, 8]. Goldrick & Blumstein's (2006) [5] suggest that these intermediate items are traces of the intended VOT of the target segment. In their view cascading activation of competing segments leads to an increase in variability. The first aim of this study is to replicate [5] in German. The second aim is to test whether the number of differing phonological feature has an effect on acoustical voicing measures. As was shown in [8, 11] speakers exhibited shorter reaction times and lower error rates for words differing in two features than for those differing in only.

Experiment and measurements: Seven German native speakers (m = 3, f = 4) participated in the experiment. Each subject read 3 repetitions of word quadrupels (e.g. "Paff Gaff Gaff Paff") to a metronome. The initial consonants differed in either one or two phonological features, e.g. *giff kiff* (voice[1]) vs. *giff tiff* (voice and place of articulation) or were identical e.g. *giff giff*.. For annotation, formant and f0 extraction the programs WebMAUS Basic, Praat and EmuR [2, 6, 12] were used. Substitution errors were annotated by two students. Delta values [7] were calculated as the Euclidean distance between measurements for each repetition of a particular segment and its mean position for each, VOT, F1, f0 and combined for all 3 measurement. Statistical analysis was carried out by Lmer package of R [1].

Results: For the annotated impressionistic errors the error rates indicate that onsets varying only in one feature are produced with higher error rates (4.26%) than minimal pairs varying in two features (2.61%). This is supported by the delta value analysis: The delta values for VOT (see Fig. 2) are smallest for the control items with no features varying ("0" in Fig. 2) and largest for items that differed in 1 feature. The delta values for two features alternating are in-between. Linear mixed models show a significant interaction with the voicing status. Therefore the data were subset. For both, voiced and voiceless stops the VOT delta values for two alternating features was significantly larger than for the controls and smaller than for one alternating feature. The secondary voicing parameters f0 and F1 as well as the combined delta value did not vary with the number of features. In conclusion, error rates and variability of VOT varied with similarity of the alternating segments as was predicted.

[1] *voice* is used here but for German *spread glottis* would be more appropriate.

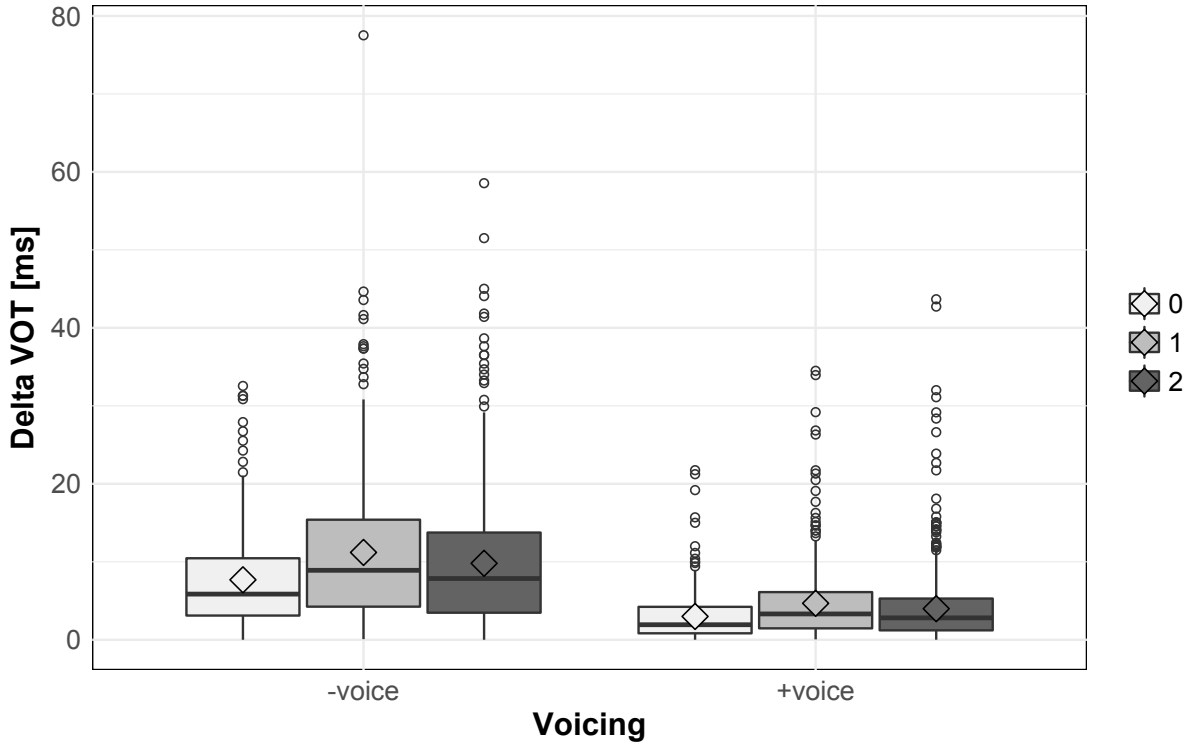


Figure 1: *Delta values for VOT for voiceless (left) and voiced stops. Light grey boxes show controls with zero alternating features (e.g. giff giff), medium grey with one (voice, e.g. giff kiff) and dark grey with two (giff piff).*

Literatur

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

[2] Boersma, P. (2006). Praat: doing phonetics by computer. *http://www. praat. org/*.

[3] Fromkin V (1973) The Non-Anomalous Nature of Anomalous Utterances. Speech Errors as LinguisticEvidence (ed. Fromkin V), 215-242.

[4] Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649-683.

[5] Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386-412.

[6] Kisler, T. and Reichel U. D. and Schiel, F. (2017): **Multilingual processing of speech via web services**, Computer Speech & Language, Volume 45, September 2017, pages 326–347.

[7] McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*(3), 243-260.

[8] Meyer, D. E., & Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory and Language*, *24*(1), 3-26.

[9] Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America*, *88*(3), 1299-1312.

[10] Shattuck-Hufnagel S, Klatt D (1979) The Limited Use of Distinctive Features and Markednessin Speech Production:Evidence from Speech Error Data. JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 18, 41-55.

[11] Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive psychology*,*41*(2), 101-175.

[12] Winkelmann, R., Jaensch, K., Cassidy, Steve & Harrington (2017), J.. emuR: Main Package of the EMU Speech Database Management SystemR package version 0.2.3..

Aspiration und Akkommodation im Slowakischen

Sofia Chasiotaki, Iveta Patáková, Kobrinian Slavik, Felicitas Kleber
Institut für Phonetik und Sprachverarbeitung, LMU
s.chasiotaki@campus.lmu.de, iveta.patakova@campus.lmu.de,
K.Slavik@campus.lmu.de, kleber@phonetik.uni-muenchen.de,

Trudgill zufolge ist die Entstehung neuer sprachlicher Varietäten (z. B. australisches Englisch) zwangsläufig ein Produkt von Dialektvermischung, da Sprecher im Gespräch automatisch akkommodieren [1]. Evidenz für die Operation eines solchen Mechanismus in Dialogsituationen findet sich in einer australisch-neuseeländischen Imitationsstudie [2]. Ziel der vorliegenden Untersuchung ist es, zu prüfen, ob Akkommodation auch bei Kontaktsprachen und untypischen Aussprachevarianten zu beobachten ist, genauer, ob sich Sprecherinnen des Slowakischen an Sprecherinnen des Tschechischen anpassen und dabei den künstlich eingefügten Parameter Aspiration imitieren, der weder in der Ausgangs- noch Zielsprache vorkommt [3].

Das Sprachmaterial bestand aus sieben Minimalpaaren mit derselben Bedeutung in Slowakisch und Tschechisch (z. B. /dɛn/, ‚Tag'; /tɛn/, ‚der'). Mit einer Ausnahme handelte es sich dabei um einsilbige Wörter mit initialem Plosiv (d. h. /p, t, k, b, d, ɡ/). Diese Wortlisten wurden von zwei tschechischen Modellsprecherinnen aus Prag und Karvina eingesprochen. Anschließend wurde in Praat Aspiration mit einer durchschnittlichen Dauer von 25 ms in die Zielwörter mit einem stimmlosen Plosiv eingefügt (z. B. /t^hɛn/).

Es wurden zwei Gruppen von Sprecherinnen mit Slowakisch als Muttersprache aufgenommen: neun 20–25-Jährige, die Englisch („Aspirationssprache") als Fremdsprache erworben haben, und neun 35–55-Jährige ohne Fremdspracherfahrung. Die Hypothesen waren, dass (1) beide Gruppen akkommodieren, basierend auf der Annahme, dass Imitation automatisch geschieht, und dass (2) jüngere Sprecherinnen aufgrund der Fremdspracherfahrung Aspiration stärker imitieren als ältere. Der Versuchsaufbau orientierte sich an dem in [2] und bestand aus drei separaten Teilen. Im Pretest benannten die Versuchspersonen Bilder der Zielwörter. Dies diente der Ermittlung einer sprecherspezifischen Baseline für die Aspiration. Im Test wurden die von den Modellsprecherinnen vorgesprochenen Wörter wiederholt. Der Posttest entsprach dem Pretest und diente der Prüfung eines Langzeiteffekts von Imitation. Eine erste Analyse von drei Wörtern mit stimmlosem Plosiv zeigte bei beiden Gruppen eine längere Aspiration in der Testphase als im Pretest, d. h. Sprecherinnen imitierten die für beide Sprachen untypische Aspiration, vermutlich automatisch. Auch wenn es keinen signifikanten Gruppenunterschied im Grad der Imitation gab, so wiesen jüngere Sprecherinnen grundsätzlich eine längere Aspiration als ältere Sprecherinnen auf, was möglicherweise auf den größeren Kontakt mit Englisch zurückzuführen ist. Eine Tendenz zu einem Langzeiteffekt ist auch in der geringfügig längeren Aspiration im Posttest zu erkennen. Die finale Analyse umfasst zusätzlich eine Auswertung eines Fragebogens zu Kontakt mit und Einstellungen gegenüber dem Tschechischen.

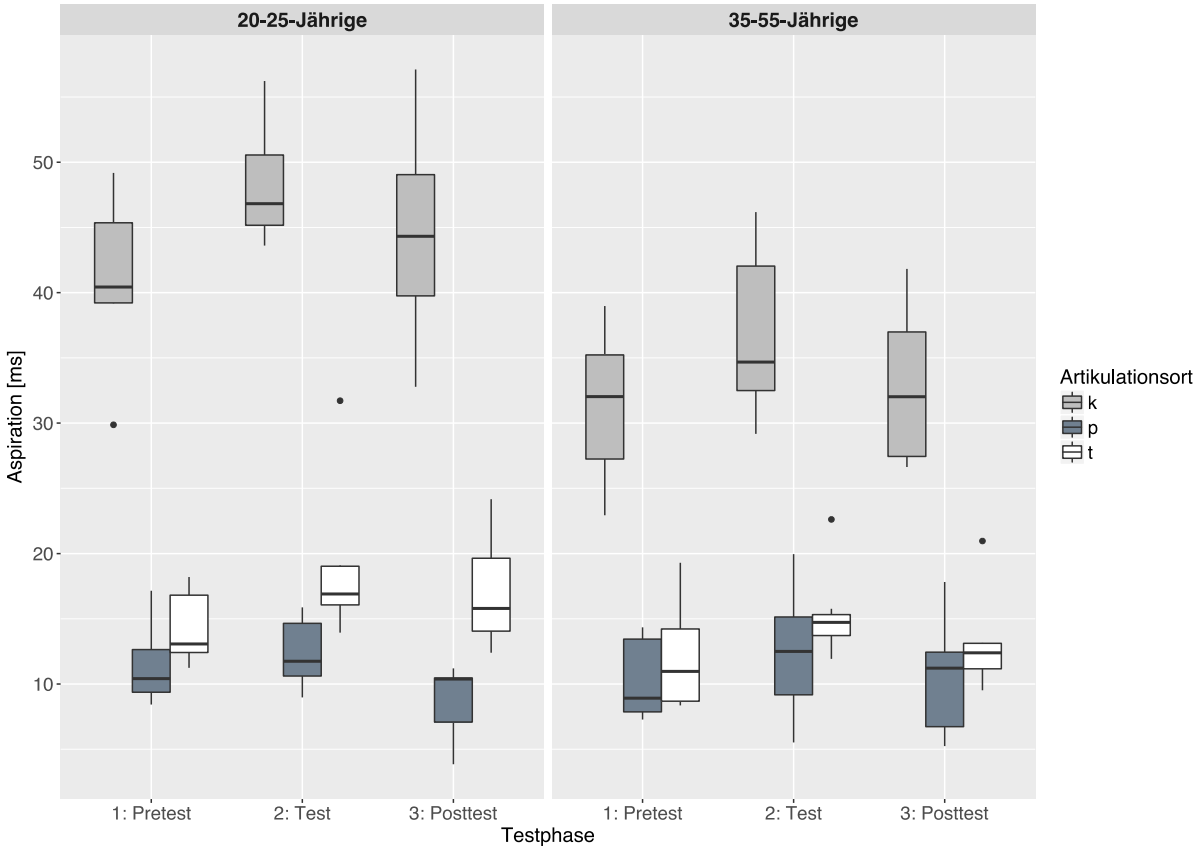


Abbildung 1: *Dauer der Aspiration [ms] getrennt nach Sprechergruppe, Testphase und Artikulationsort des Plosivs.*

Literatur

[1] P. Trudgill, "Colonial dialect in the history of European languages: On the irrelevance of identity to new-dialect formation," *Language in Society 37*, pp. 241-254, 2008.

[2] M. Babel, "Dialect divergence and convergence in New Zealand English", *Language in Society 39*, pp. 437-456, 2010.

[3] Á, Kráľ, J. Sabol, "Fonetika a fonológia," Bratislava: SPN, 1989, pp. 282-289.

Die Aussprache des Deutschen in Luxemburg

François Conrad[1], Judith Manzoni[2]
[1]Leibniz Universität Hannover
[2]Universität Trier
francois.conrad@germanistik.uni-hannover.de, manzoni@uni-trier.de

Das Luxemburgische (Lb.) grenzt als westlicher Ausläufer der Kontinentalgermania direkt an die Romania (Belgien/Frankreich) an. Hierdurch ergibt sich für das Land eine über Jahrhunderte gewachsene Mehrsprachigkeit, die insbesondere Lb., Deutsch (Dt.) und Französisch umfasst [1] und sich in allen sprachlichen Bereichen niederschlägt [2]. Als Alphabetisierungssprache in der Schule und wichtige Sprache in den (Print)Medien ist das Dt. im Großherzogtum schriftlich sehr präsent [1, 2]. Dialektal dem moselfränkischen Kontinuum zugeordnet, emanzipiert das Lb. sich von diesem seit einigen Jahrzehnten insbesondere auf der lautlichen Ebene [3]. Zugleich teilt die Sprache nach wie vor wesentliche moselfränkische Merkmale mit den benachbarten bundesdeutschen Dialekten, etwa die Koronalisierung von /ç/ und dessen Zusammenfall mit /ʃ/ zu /ɕ/ [4], die Aussprache von <r> als [χ] vor stimmlosen Plosiven [ˈpɑχ̲.kən] ‚parken' [5], die durchgängige Realisierung von Schwa in Flexiven ([ˈpɑχ.kə̲n] ‚parken', [ˈiə.zə.lə̲n] *Ieselen* ‚Esel, Pl.') [6] oder Lenisierung bzw. Ausfall inlautender Plosive (*Lie[v]en* ‚Leben', *Reen* ‚Regen') [7]. Während zahlreiche Beiträge aus verschiedensten sprachlichen Bereichen in den letzten Jahrzehnten das Wissen um das Lb. stark erweitert haben, ist die linguistische Beschreibung der Aussprache des Deutschen durch lb. Muttersprachlerinnen und Muttersprachler bisher ein Forschungsdesiderat. In einer experimentellen Erhebung (n=29) im Rahmen einer variationslinguistischen Studie [6] lasen lb. Muttersprachlerinnen und Muttersprachler aus der Hauptstadt Luxemburg – Teil der zentrallb. Varietät, von der der aktuelle Standardisierungsprozess ausgeht [8] – die Fabel ‚Nordwind und Sonne' auf Deutsch vor. Die auditive und akustische Analyse (Formantmessungen und Spektralanalyse sowie Pitch-Messungen von Tonhöhenunterschieden innerhalb der nuklearen Struktur) dieser Aufnahmen zeigt, dass zahlreiche Interferenzen im intendierten Standarddt. (so die Vorgabe) auftreten. Diese umfassen für den Konsonantismus viele der oben angeführten moselfränkischen bzw. westmitteldeutschen Merkmale, während im Vokalismus häufig etwa die Qualität der lb. Kurzvokale [ɑ] und [æ] (standarddt. [a] und [ɛ]) sowie die nur gespannt vorkommenden lb. Kurzvokale [i] und [u] vom Standarddt. abweichen. Diese und weitere phonologische wie phonetische Auffälligkeiten auf segmentaler und supra-segmentaler (phonologische Anpassungsprozesse, Betonung, Intonation) Ebene geben ein ausführliches Bild über die deutsche Aussprachevarietät in Luxemburg. Ein Vergleich der verschiedenen soziodemographischen Gruppen innerhalb des Sprechersamples (u.a. drei Generationen, zwei Geschlechter, unterschiedliche Bildungshintergründe) erlaubt zudem eine soziolinguistische Situierung bestimmter Aussprachevarianten, deren Vorkommen zwischen den Gruppen sowie auf individueller Ebene eine große Variation zeigt.

Literatur

[1] F. Fehlen, *BaleienBis. Une enquête sur un marché linguistique en profonde mutation. Luxemburgs Sprachenmarkt im Wandel*, RED No. 12, Luxemburg: Sesopi, 2009.

[2] F. Conrad, *Variation durch Sprachkontakt. Lautliche Dubletten im Luxemburgischen*, Luxemburg-Studien No. 14, 2017.

[3] P. Gilles, "Die Emanzipation des Lëtzebuergeschen aus dem Gefüge der deutschen Mundarten", in *Regionale Sprachgeschichte*, W. Besch & H. J. Solms (Hrsg.), Berlin: Erich Schmidt, 1998, pp. 20-35.

[4] F. Conrad, "Der Zusammenfall von /ʃ/ > /ɕ/ < /ç/ im Luxemburgischen", *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13)*, pp. 29-32, 2018.

[5] P. Gilles & J. Trouvain, "Luxembourgish", *Journal of the International Phonetic Association*, 43(1), pp. 67-74, 2013.

[6] F. Conrad, "Schwa-Elision und die Wortprosodie des Luxemburgischen", *Linguistik online*, 80(1), pp. 21-49, 2017.

[7] C. V. J. Russ, "Lëtzebuergesch: A linguistic description", in *Luxembourg and Lëtzebuergesch. Language and communication at the crossroads of Europe*, G. Newton (Hrsg.), Oxford: Clarendon Press, 1996, pp. 67-96.

[8] P. Gilles & C. Moulin, "Luxembourgish", in *Germanic standardization – past and present*, A. Deumert & W. Vandenbussche (Hrsg.), Amsterdam: John Benjamins, 2003, pp. 303-329.

Abstract für die P&P 14 – Wien, 6.-7. September 2018

Johanna Fanta[1]
[1]Johanna Fanta
Johanna.fanta@univie.ac.at

Intersituative language dynamics among the dialect-standard-axis in rural Austria

Austria's language situation can be characterized by a so-called 'inner multilingualism', allowing speakers to switch and shift effortlessly among the dialect-standard-axis according to certain situational parameters (e. g. formality, closeness etc.). Hitherto, there are several thorough investigations focusing on the 'architecture' of the dialect-standard-axis for specific regions in Germany (cf. LENZ 2003, KEHREIN 2012 and LAMELI 2004), comparable endeavours for the Austrian context, however, are usually limited to specific locations (e. g. SCHEUTZ 1985, SCHEURINGER 1990) or focus primarily on just one part of the dialect-standard-axis (cf. MOOSMÜLLER 1991). To contribute to the description of the overall 'vertical' spectrum, this poster aims to show early auditory phonetic results on the analysis of areal-horizontal and social-vertical variation.

To this end, a large-scale multivariate investigation is conducted within the framework of the Special Research Programme (FWF) "German in Austria. Variation – Contact – Perception". The empirical input consists of language data from 'autochthonous' speakers in rural areas of Austria, representing the major dialect regions of the country: the central Bavarian, the south central Bavarian transition zone, the southern Bavarian and the Alemannic area. The data is collected in various 'natural' and standardised survey settings: an interview led by a foreign academic, an unguided conversation among friends, two translation tasks and reading-aloud tasks.

First results based on the analysis of specific phonological variables show intersituative language behaviour patterns for selected speakers with varying sociodemographic backgrounds. While some speakers show strong binary patterns, others seem to shift and switch more gradually along the dialect-standard-axis. For each phonological phenomenon (e. g. MHG /oe/ and l-vocalization) different degrees of variability can be distinguished regarding the vertical-social as well as the horizontal-diatopic distribution.

Literature

[1] R. Kehrein, *Regionalsprachliche Spektren im Raum. Zur linguistischen Struktur der Vertikale,* Zeitschrift für Dialektologie und Linguistik. Beihefte 152, 2012.
[2] A. Lameli, *Standard und Substandard. Regionalismen im diachronen Längsschnitt,* Zeitschrift für Dialektologie und Linguistik. Beiheft 128, 2004.
[3] A. Lenz, *Struktur und Dynamik des Substandards. Eine Studie zum Westmitteldeutschen (Wittlich/Eifel),* Zeitschrift für Dialektologie und Linguistik. Beihefte 125, 2003.

[4] S. Moosmüller, *Hochsprache und Dialekt in Österreich. Soziophonologische Untersuchungen zu ihrer Abgrenzung in Wien, Graz, Salzburg und Innsbruck.* Wien / Köln / Weimar: Böhlau. (= Sprachwissenschaftliche Reihe 1), 1991.
[5] H. Scheuringer, *Sprachentwicklung in Bayern und Österreich. Eine Analyse des Substandardverhaltens der Städte Braunau am Inn (Österreich) und Simbach am Inn (Bayern) und ihres Umlandes.* Hamburg: Buske (= Beiträge zur Sprachwissenschaft. Band 3), 1990.
[6] H. Scheutz, *Strukturen der Lautveränderung. Variationslinguistische Studien zur Theorie und Empirie sprachlicher Wandlungsprozesse am Beispiel des Mittelbairischen von Ulrichsberg/Oberösterreich* (= Schriften zur deutschen Sprache in Österreich 10), Wien: Braumüller, 1985.

Uptalk bei weiblichen Sprecherinnen im Deutschen – eine Korpusstudie

Fabian Fey[1], Natalie Lewandowski[2]
[1]fabian.fey@ims.uni-stuttgart.de
[2]natalie.lewandowski@ims.uni-stuttgart.de

Uptalk – das Phänomen steigender Intonation, ähnlich der in Fragen, jedoch am Ende von deklarativen Phrasen – ist eine Sprecheigenschaft, die außerhalb des englischen Sprachraumes, darunter auch im Deutschen, noch wenig untersucht wurde. Wir konnten anhand einer Korpusstudie auf der *GECO* (**Ge**rman **Co**nversations) Datenbank [1] nachweisen, dass Uptalk, ähnlich wie in der englischen Literatur beschrieben, auch von weiblichen deutschen Sprecherinnen verwendet wird. Zusätzlich können erste Rückschlüsse auf mögliche Korrelationen zu Persönlichkeitseigenschaften der Sprecherinnen gezogen werden.

Nach dem derzeitigen Kenntnisstand kann Uptalk verwendet werden um dem Gesprächspartner Unsicherheit zu signalisieren [2]. Andererseits wird *Uptalk* auch häufig vom dominanteren Sprecher verwendet [3], beispielsweise um kontrollieren zu können, ob der Konversationspartner dem Gespräch noch folgen kann [4]. Darüber hinaus kann *Uptalk* eine minimale verbale [5] und eine positive nonverbale Reaktion hervorrufen [6]. *Turn-taking* [7] und *floor-holding* [8, 9] werden ebenfalls durch Uptalk gesteuert.

Die GECO Datenbank beinhaltet 92 Aufnahmen weiblicher deutscher Muttersprachlerinnen aus insgesamt 46 Dialogen [1], die in zwei Konditionen entstanden sind – mit und ohne Sichtkontakt. Zusätzlich zu den Aufnahmen verfügt GECO u.a. auch über Daten aus einem Selbstüberwachungstest [12]. Die Aufnahmen wurden automatisch anhand von *PaIntE-Parametern* und *GTobiS*-Labels auf steigende Intonation an Phrasenenden analysiert. Phrasenendwörter (ausgenommen von intermediären Phrasenenden) wurden durch ein Satzzeichen (Punkt, Ausrufezeichen oder Fragezeichen) im Transkript identifiziert. So konnten maschinell 3978 von 9429 (42,19%) Phrasenenden mit Uptalk ermittelt werden, wobei nur *H%*-Labels aus Nicht-Fragesätzen eingeschlossen waren. Das tatsächliche Auftreten von Uptalk in dieser Teilmenge wurde anschließend zusätzlich stichprobenweise auditorisch überprüft und verifiziert.

Um Korrelationen zwischen Persönlichkeitsmerkmalen der Probandinnen und dem Auftreten von Uptalk zu ermitteln, wurde ein *generalisiertes lineares Modell* (*glm*) mit den Faktoren *Modalität* (d.h. mit oder ohne Sichtkontakt), den vier Persönlichkeitsmerkmalen (*Acting, Extraversion, Other-directedness und Sensitivität für soziale Hinweise* erstellt. Höhere *Acting* und *Other-directedness* Werte und die Worthäufigkeit beeinflussen das Vorhandensein von Uptalk signifikant. Während die Modalität allein sich nicht auf das Auftreten von Uptalk auswirkt, wurde eine Interaktion von Modalität in Kombination mit *Acting*, *Extraversion* und *Sensitivität* festgestellt. Ein weiteres *glm* zeigt starke individuelle Unterschiede in der Menge von Uptalk pro Sprecherin und innerhalb konkreter Sprecherin-Partnerin-Konstellationen, was den Schluss zulässt, dass Uptalk ein sehr dynamisches und partnerabhängiges Phänomen ist.

Literatur

[1] A. Schweitzer und N. Lewandowski. Convergence of Articulation Rate in Spontaneous Speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon*, S. 525–529, 2013.

[2] M. Shepherd. Functional significance of rising-intonation declaratives in settings with special discursive norms. *LSA Annual Meeting Extended Abstracts*, 2(0):10–1–5, 2011.

[3] W. Cheng und M. Warren. The use of rise and rise-fall tones in the Hong Kong Corpus of Spoken English. *International journal of corpus linguistics*, 10:85–107, 2005.

[4] J. House. The role of prosody in constraining context selection: A procedural approach. *Cahiers de Linguistique Francaise 28: Interfaces discours-prosodie*, 28:369–383, 2007.

[5] S. Allan. *The rise of New Zealand intonation*, S. 115–128. Multilingual Matters, 1990.

[6] P. Bryant. *Australian questioning intonation: An addition to speakers' response-seeking repertoire*. Australian National University, Faculty of Arts, Department of Linguistics, nicht veröffentlichte Bachelorthesis, 1980.

[7] A. Wichmann und J. Caspers. Melodic cues to turn-taking in English: evidence from perception. *Proceedings of SIGDIAL '01*. Proceedings of the Second SIGdial Workshop on Discourse and Dialogue, Volume 16, S. 1–6, 2001.

[8] A. Arvaniti und M. Atkins. Uptalk in Southern British English. *Speech Prosody 2016*, 2016.

[9] A. Ritchart und A. Arvaniti. The form and use of uptalk in Southern Californian English. In *Proceedings of Speech Prosody*, volume 7, S. 20–23, 2014.

[10] D. Britain. In W. Labov und R. T. Cacoullos (Hrsg.) *Language variation and change*, S. 77–104. Cambridge University Press, 1992.

[11] V. Shokeir. Evidence for the stable use of uptalk in South Ontario English. *University of Pennsylvania Working Papers in Linguistics*, 14, 2009.

[12] G. von Collani und S. Stürmer. Deutsche Skala zur Operationalisierung des Konstrukts Selbstüberwachung (Self-Monitoring) und seiner Facetten. Zusammenstellung sozialwissenschaftlicher Items und Skalen. doi: 10.6102/zis50, 2014.

The Merger of [eː] and [ɛː] in Standard German

Marina Frank[1]
[1]Forschungszentrum Deutscher Sprachatlas, Philipps-Universität Marburg
Frankma@students.uni-marburg.de

The question as to the phoneme status of [ɛː] (in words such as *Käse* 'cheese' and *Mädchen* 'girl') in Standard German has received considerable attention in the phonological literature (e.g. Ramers 1988, Stiel 2018). This discussion may have led to the claim in the literature that /ɛː/ has recently undergone a shift to [eː] in Standard German ([keːzə], [meːtçən], cf. Herrgen 2015: 139).

The present study examines the realization of /ɛː/ from 12 female native speakers of German (age: 22–26) by means of formant measurements. Using a corpus of 850 realizations of both /eː/ and /ɛː/, the following hypotheses will be tested:

- Hypothesis 1: [ɛː] is being replaced by [eː] (cf. Herrgen 2015: 139)
- Hypothesis 2: Speakers from northern Germany tend to replace [ɛː] by [eː] more often than speakers from other regions (cf. Ternes 1999: 95 and Ramers 1988: 156–157)
- Hypothesis 3: [ɛː] tends to be replaced by [eː] in the context before /r/ (cf. Ramers 1988: 160–162, Wiese 2000: 17, Stiel 2018)

The results of the present study show that [ɛː] is not being replaced by [eː]. Instead, both phones approach each other phonetically, but they remain phonologically distinct (cf. Ramers 1988: 163, Stiel in print). This is shown by the fact that all 12 speakers differentiate the two phones. The speakers also did not differentiate in terms of their regional origin. Furthermore, [ɛː] is not replaced before /r/, instead the two phones slightly approach one another. These findings contradict previous claims by Herrgen (2015: 139) and Wiese (2000: 17); however, they are in line with the formant analyses by Stiel (2018) and Sendlmeier / Seebode (2008).

Literature

[1] J. Herrgen, "Entnationalisierung des Standards. Eine perzeptionslinguistische Untersuchung zur deutschen Standardsprache in Deutschland, Österreich und der Schweiz", in: *Standarddeutsch im 21. Jahrhundert – Theoretische und empirische Ansätze mit einem Fokus auf Österreich*, A. N. Lenz & M. M. Glauninger (eds.), Göttingen: V&R unipress (Wiener Arbeiten zur Linguistik. 1), 2015, pp. 139–164.

[2] K. H. Ramers, *Vokalquantität und -qualität im Deutschen*, Tübingen: Niemeyer (Linguistische Arbeiten. 213), 1988.

[3] W. F. Sendlmeier & J. Seebode, *Formantkarten des deutschen Vokalsystems*, 2008. URL: https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf, letzter Zugriff: 20.02.2018.

[4] R. Stiel, "Symmetriebildung im deutschen Langvokalsystem. Eine variationslinguistische Untersuchung zum Abbau des /ɛː/-Phonems im intendierten Standarddeutschen", in: *Perspektiven der Variationslinguistik II. Neue Beiträge aus dem Forum Sprachvariation*, B. Ganswindt & C. Purschke (eds.), Hildesheim [u. a.]: Olms (Germanistische Linguistik), 2018, pp. 291–337.

[5] R. Wiese, *The phonology of German*, Oxford [u. a.]: Oxford University Press (The Phonology of the World's Languages), 2000.

Abstract für die P&P 14 – Wien, 6.-7. September 2018

Christoph Gube
Humboldt-Universität zu Berlin
christoph.gube@gmail.com

Normalisierung von Vokalen im DaF-Kontext: Vergleich von Normalisierungsmethoden

DaF-LernerInnen mit L1 Chinesisch haben oft große Schwierigkeiten mit dem deutschen Vokalsystem. Insbesondere das Konzept der Kurz- und Langvokale und die damit einhergehende Änderung der Vokalqualität ist in der mündlichen Produktion eine häufige Fehlerquelle. Um solche Abweichungen der chinesischen LernerInnen systematisch zu analysieren, ist es hilfreich, Zungenposition (F_2) und Mundöffnung (F_1) direkt mit den entsprechenden Daten deutscher MuttersprachlerInnen vergleichen zu können.

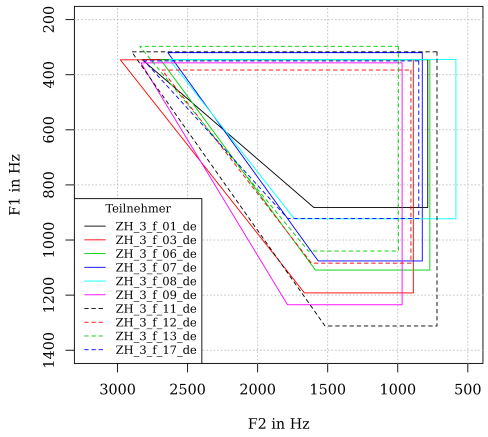
"Rohe" Formantdaten in Hz sind jedoch für einen direkten Vergleich zwischen verschiedenen SprecherInnen nicht geeignet, weil diese Werte mit der individuellen Anatomie des Vokaltrakts stark variieren. Es ist deshalb für Vergleiche zwischen SprecherInnen nötig, die Formantdaten zu normalisieren. Hierfür steht eine große und ständig wachsende Anzahl von Methoden und Werkzeugen zur Verfügung.

Am Beispiel von chinesischen und deutschen Audio-Daten aus Sprachproduktions-Experimenten vergleiche ich 12 verschiedene Methoden (und einige Modifikationen) der Vokal-Normalisierung mit dem Ziel, ihre Eignung für den direkten Vergleich im DaF-Kontext zu beurteilen. Hierfür verwende ich den quadrierten Variationskoeffizienten (SCV) sowie die Vereinigung und Schnittmenge der theoretischen maximalen Vokalräume (Bild 1). Kriterien für den Vergleich der Methoden sind SCV-Werte sowie Kongruenz der theoretischen Vokalräume in Prozent (Flynn 2011).

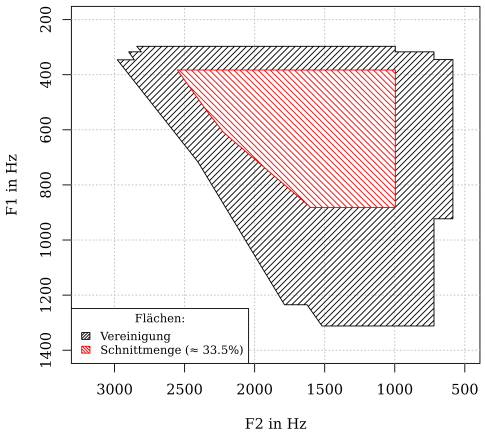
Die für diese Arbeit verwendeten Audio-Daten stammen aus Produktionsexperimenten mit deutschen MuttersprachlerInnen sowie mit chinesischen Germanistik-Studierenden der *Universität Nanchang* (Provinz Jiangxi, VR China). Aufgrund der begrenzten Anzahl männlicher Germanistik-Studierender dort beschränkt sich meine Arbeit in beiden Sprachen auf Daten von Sprecherinnen.

Ich komme zu dem Ergebnis, dass die Normalisierungsmethoden *Gerstman*, *Watt & Fabricius (Variante 2, 2009)* und *Bigham* am besten geeignet sind, um die Formantdaten von chinesischen und deutschen Muttersprachlerinnen zu vergleichen (exemplarisch in Bild 2: *Gerstman*). Die Methoden *Watt & Fabricius* und *Bigham* können außerdem leicht modifiziert werden, um die Vokalräume verschiedener Sprachen besser zu repräsentieren, als das in ihrer ursprünglichen Form der Fall ist.

Perspektivisch könnte mit einem solchen Vergleich – eine solide Datenbasis vorausgesetzt – bestimmt werden, welche Abweichungen LernerInnen mit L1 Chinesisch typischerweise produzieren. Diese Ergebnisse können helfen, im Aussprache-Training im DaF-Unterricht besser auf die speziellen Bedürfnisse chinesischer LernerInnen einzugehen.

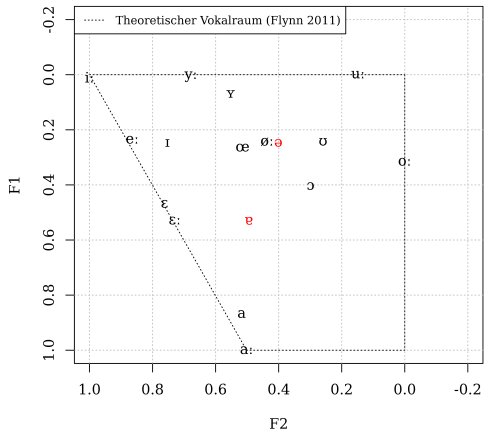


(a) Theoretische maximale Vokalräume (Deutsch), chinesische Probandinnen

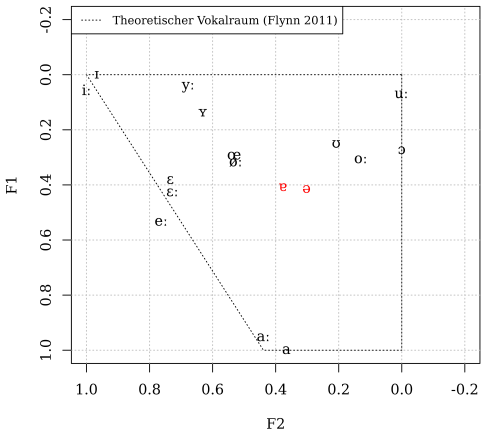


(b) Vereinigung und Schnittmenge (Deutsch), chinesische Probandinnen

Abbildung 1: Maximale Vokalräume und Schnittmenge/Vereinigung



(a) Deutsche Sprecherin DE-2-f-12



(b) Chinesische Sprecherin ZH-3-f-09

Abbildung 2: *Gerstman*-Normalisierte Formantkarten im Vergleich. ZH-3-f-09 produziert z.B. keine Unterschiede zwischen /e/ und /ɛ/ bzw. /i/ und /ɪ/. Möglicherweise wurden /e/ und /ɪ/ in Phonemkategorien ihrer L1 assimiliert.

Literatur

[1] Douglas Stephan Bigham: „Dialect Contact and Accommodation Among Emerging Adults in a University Setting“. Diss. Austin: The University of Texas at Austin, Aug. 2008. URL: `https://repositories.lib.utexas.edu/bitstream/handle/2152/17759/bighamd.pdf` (besucht am 18. 03. 2018).

[2] Anne H. Fabricius, Dominic Watt und Daniel Ezra Johnson: „A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics“. In: *Language Variation and Change*, 21 (3) (Okt. 2009), S. 413–435.

[3] Nicholas Flynn: „Comparing vowel formant normalisation procedures“. In: *York papers in linguistics*, 2 (11) (März 2011), S. 1–28.

[4] Cordula Hunold: *Untersuchungen zu segmentalen und suprasegmentalen Ausspracheabweichungen chinesischer Deutschlernender*. Hallesche Schriften zur Sprechwissenschaft und Phonetik 28. Frankfurt am Main etc.: Peter Lang, 2009.

[5] Dominic Watt und Anne H. Fabricius: „Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F_1~F_2 plane“. In: *Leeds Working Papers in Linguistics and Phonetics*, 9 (9) (2002), S. 159–173.

Speaking Style Alignment Initiated by Virtual Agents

Jennifer Jaekel[1], Petra Wagner[1,2], Simon Betz[1,2]

[1]AG Phonetik & Phonologie, Universität Bielefeld
[2]CITEC (Excellence Cluster Cognitive Interaction Technology), Universität Bielefeld

j.jaekel@uni-bielefeld,de

Smart-home environments and their application steadily increase in importance. Simultaneously, Western-European societies undergo a demographic change, the average age of people increases. This study investigates communicative aspects of human-machine interaction with regard to age in an interaction study within a smart-home environment [1]. It has been observed that humans align in communication on various levels [2]. In Kriz et al 2009 [3], it was found that humans change their speaking style when addressing a robot. In this study, we investigate alignment in human-machine communication, with regard to two main questions:

1) Is it possible for an intelligent, interactive system to negotiate a new speaking style via variation of pitch range and verbosity in an utterance?
2) Is linguistic alignment between humans and machines influenced by age?

The study was conducted with 44 participants in a smart-home environment. During the experiment, participants verbally addressed a virtual agent from a pre-defined list of topics and the agent replied. The replies could be in any of four conditions regarding verbosity (high/low) and pitch range (high/low). The verbosity was manipulated by changing the utterances of the virtual agent, the pitch range was artificially increased or decreased using PSOLA (cf. Figure 1). We used a between subjects design, so every participant only interacted with the agent in one of the possible four conditions. It was then measured whether the condition induced alignment; i.e. if the participants' verbosity or pitch range changed over the course of the interaction.

First analyses of the results yield surprising results. The most striking effect is a disalignment: participants significantly decrease their pitch range the longer the utterances of the virtual agent become. Furthermore, participants increased their pitch range when the agent addressed them with a monotone pitch contour. Furthermore, we found an age effect: "younger" (age 18-35) participants seem to align more to a monotonously speaking virtual agent as compared to "older" (age 36+) ones.

This study yields several insights for future studies in human-machine alignment. In short-term interactions as studied here, there appears to be no alignment, despite the fact that extremely contrasting conditions were used. Alignment observed was in the opposite direction than the expected, which has to be borne in mind when designing systems in the future. Age seems to have an influence, so interacting machines have to be designed aptly for the individual user groups. An open question is at the moment, how users would align in long-term interactions with virtual agents, e.g. after years of living in a smart-home environment.

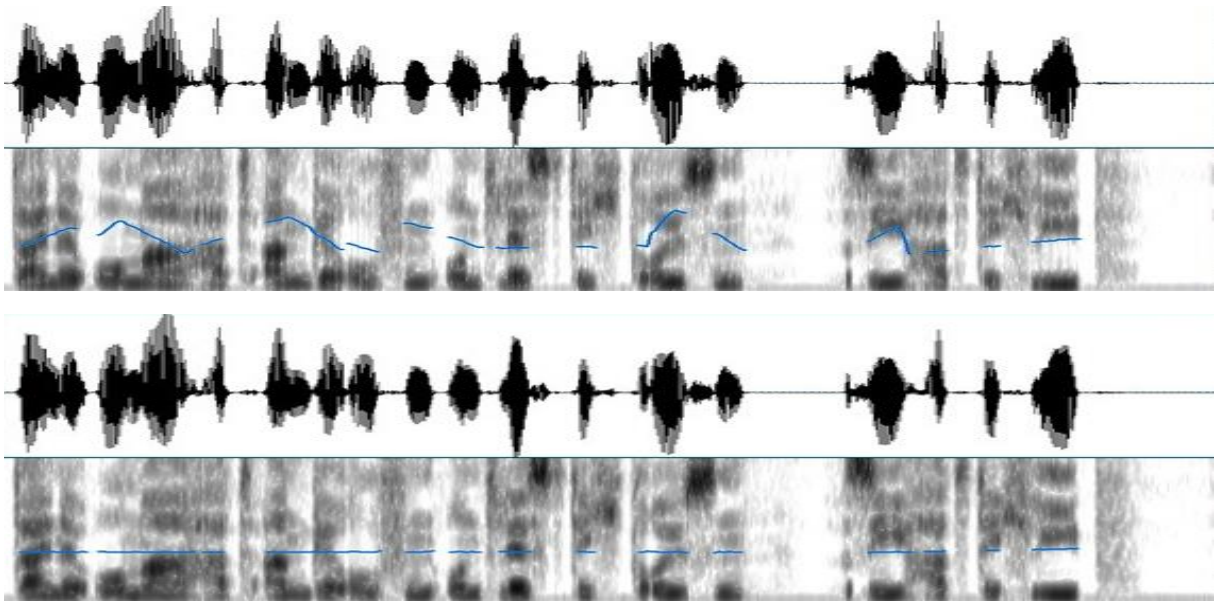


Figure 1. Virtual Agent's synthetic utterances, top panel: with increased pitch (blue line) range, bottom panel: with decreased pitch range.

Literatur (Beispiele)

[1] Wrede, S.; Leichsenring, C.; Holthaus, P.; Hermann, T.; Wachsmuth, S. The Cognitive Service Robotics Apartment: A Versatile Environment for Human-Machine Interaction Research.*Kuenstliche Intell.*2017,*31*, 299–304.

[2] Garrod S. und Pickering M.J: "Interactive alignment and prediction in dialogue". In: Ipke Wachsmuth, Jan de Ruiter, Petra Jaecks und Stefan Kopp (Hg.): Alignment in Communication. Towards a new theory of communication, Bd. 6, 2013 Amsterdam/Philadelphia: John Benjamins Publishing Company (Advances in Interaction Studies), S. 193–205.

[3] Kriz, Sarah; Anderson, Gregory; Trafton, J. Gregory; Bugajska, Magdalena: "Robot-Directed Speech as a Means of Exploring Conceptualization of Robots". *HRI*'09, March 11–13, 2009, La Jolla, California, USA.

Annotation of Haitian Creole prosody and intonation

Alexander Teixeira Kalkhoff
Albert-Ludwigs-Universität Freiburg
alexander.teixeira.kalkhoff@romanistik.uni-freiburg.de

Since, for the most part, Haitian Creole (HC henceforth) prosody and intonation still represent desiderata of linguistic research [1], an experimental phonetic approach to the data sets is mandatory. Its outcome must be first hypotheses about the HC prosodic system and HC intonational phonology. Unfortunately, there is a considerable methodological gap within prosodic and intonational research, because, due to its strong phonological claim, the generally-accepted and well-established ToBI (Tones and Break Indices) annotation system [2] appears inappropriate for that test stage. ToBI-annotated data sets represent the final stage of the analytic process and are readable only in the light of the intonational phonology of a given language.

To fill this gap and to annotate intonational variation of still unknown phonological systems of English varieties, the IViE (Intonation Variation in English) annotation system was developed at the Cambridge University under the auspices of Francis Nolan [3]. Brechtje Post and Elisabeth Delais-Roussarie expanded the language-specific IViE model to the language-non-specific Intonational Variation Transcription System (IVTS) [4]. The IVTS model encodes orthographic, prosodic, and intonational information on six annotation levels, i.e. a comment, a phonological (or tonal), a global phonetic, a local phonetic, a rhythmic (or prominence), and an orthographic tier. Starting point for the annotation process is the identification of prominent syllables and a narrow phonetic annotation of local intonational events. Intonation at the discourse level is annotated phonetically at the global phonetic tier. First phonological assumptions can be also annotated at the phonological tier. The IViE/IVTS set of labels is transparent, easy to manage and re-uses well-established ToBI symbols, such as L, H, and % [5].

Under the condition of the largely unknown HC prosodic and intonational system, I adopted, according to my first working hypotheses about HC prosody [6], the IVTS model (for a full sketch of the annotation hierarchy see the additional page). Accordingly, I added five annotation tiers to the original six IVTS annotation levels: Three to collect more prosodic information about HC penultimate and final lengthening, syllable structure, and prosodic constituency; one to correlate pragmatic information with local and global intonational events; and one to correlate specific word classes with tonal movements. The annotation hierarchy can be easily implemented into Praat [7].

The goal of my contribution is to discuss the suggested annotation hierarchy and the enlarged full set of annotation labels on the basis of an annotated HC test sample taken from the APiCS online [8]. The discussion is designed to take place within the greater context of data-driven corpus annotation of prosody and intonation of spontaneous speech data [9] [10].

Annotation hierarchy for annotating prosody and intonation based on the IViE/IVTS model (Grabe & Post 2004; Post & Delais-Roussarie 2006)

(i) **Comment tier** (for remarks)
(ii) **Segmental tier** (to enable durational measurements of vocalic segments to test penultimate and final lengthening)
(iii) **Syllable tier** (including an evaluation of syllabic complexity)
(iv) **Phonological tier** (for building hypotheses about the intonational phonology of the variety under scrutiny; usage of a limited set of labels)
(v) **Break Indices tier** (to annotate prosodic constituency)
(vi) **Macro-prosodic (phonetic) tier** (to annotate prosodic/intonational events related to the discourse level)
(vii) **Micro-prosodic (phonetic) tier** (to annotate phonetically the realized pitch movement within the Implementation Domain (ID) upon the prominent syllables)
(viii) **Rhythmic/Prominence tier** (to annotate prominent syllables; stress and pitch accented syllables)
(ix) **Pragmatic tier** (pragmatic annotation of speech acts and focus types; with a limited set of pragmatic tags; in order to make searchable speech acts (illocution) and pragmatic tones within the annotated corpus)
(x) **Glossed tier** (interlinear glossing of the lexical and grammatical items; to make searchable pitch movements upon functional word classes (grammatical tones), e.g. upon DET)
(xi) **Orthographic tier** (orthographic transcription of the speech signal)

References

[1] A. Teixeira Kalkhoff, "Éléments et structures suprasegmentaux," in *Manuel des langues créoles à base française*, P. Krämer, K. Mutz & P. Stein (eds.), Berlin: De Gruyter (Manuals of Romance Linguistics), in press.
[2] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert & J. Hirschberg, "ToBI: A standard for labeling English prosody," in *Proceedings from the 2nd* International *Conference on Spoken Language Processing (ICSLP 92), Banff, Alberta (Canada)*, 1992, pp. 867–870.
[3] E. Grabe & B. Post, "Intonational variation in the British Isles," in *Corpus Linguistics: Readings in a widening discipline*, G. Sampson & D. McCarthy (eds.), London: Continuum International, 2004, pp. 474–481.
[4] B. Post & E. Delais-Roussarie, "Transcribing intonational variation at different levels of analysis," in *ISCAM Archives, Speech Prosody 2006, Dresden, Germany, May 2-5, 2006.*
[5] *The IViE Labelling Guide*, version 3, copyright by Esther Grabe, 2001 (http://www.phon.ox.ac.uk/files/apps/IViE/guide.html) (Accessed 2018-03-15).
[6] A. Teixeira Kalkhoff, "Using corpus data for testing a working hypothesis about Haitian Creole prosody," manuscript.
[7] P. Boersma & D. Weenink, *Praat: Doing phonetics by computer* [Computer program], version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/.
[8] D. Fattier, "Structure dataset 49: Haitian Creole" (http://apics-online.info/contributions/49) (Accessed 2018-03-15).
[9] E. Delais-Roussarie & B. Post, "Corpus annotation: methodology and transcription systems," in *The Oxford Handbook of Corpus Phonology*, J. Durand, U. Gut & G. Kristoffersen (eds.), Oxford: Oxford University Press, 2014, pp. 46–88.
[10] C. Brierley & E. Atwell, "ProPOSEC: A prosody and POS annotated spoken English corpus," in *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, pp. 1266–1270.

Quantitative analysis of Glossolalia – A case of phonetic signals without semantics

Samantha Link, R. Harald Baayen, Fabian Tomaschek
Department of General Linguistics, University of Tübingen, Germany

It is repeatedly reported that frequent words and syllables are articulated shorter than rare words and syllables, while everything else is held constant (Gahl 2008; Aylett and Turk 2004; Bell et al. 2009). Several explanations have been put forward for this phenomenon. For example, Cholin and Levelt (2009) argue that frequent syllables can be accessed from a syllabary, which stores motor commands for frequent syllables, making the speech production process faster. By contrast, rare syllables have to be generated ad hoc, reducing articulatory pace, thus lengthening the phonetic signal. However, the idea of the syllabary does not allow for gradual distinction between frequent and infrequent syllables.

Another explanation is put forward by Bell et al. (2009), according to which higher frequency of occurrence allows for faster access times to the linguistic lexicon, again increasing articulatory pace. In contrast to Levelt, Roelofs, and Meyer (1999), these allow gradual changes in duration.

In the current paper, we investigated the locus of the frequency effect by investigating glossolalia ('speaking in tongues'), which describes the vocalization of semantic-deprived syllable- sequences during public or private prayer. 15 recordings of glossolalic speech from 11 German native speakers were analyzed.

Although glossolalia has no semantics, it is still determined by underlying phonotactic models (Enninger and Raith 1983). Furthermore, we show that syllable frequencyies can be modelled by a Zipfian distribution, indicating that glossolalia behaves like natural speech.

Controlling for local speaking rate variation, we find that syllable dura-tion in glossolalia correlates with the frequency of occurrence of glossolalic syllables ($\beta = -0.023$, se = 0.01, $t = -2.3$) as well as with frequency of occurrences of syllables in German ($\beta = -0.034$, se = 0.004, $t = -8.78$). Im-portantly, glossolalic and German syllable frequencies are mildly correlated ($r = 0.25$), indicating that the glossolalic system is not primarily based in the native language of the speaker. Our results show that the access of concepts in the mental lexicon is not the only source for shorter acoustic duration. Also, the syllabary is not an either-or mechanism but shows gradual effects of experience, as we have already argued in Tomaschek et al. (2014).

References

[1] M. Aylett and A. Turk. “The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech.” In: Language and Speech 47.1 (2004), pp. 31–56.

[2] Alan Bell et al. “Predictability effects on durations of content and func-tion words in conversational English.” In: Journal of Memory and Lan-guage 60.1 (2009), pp. 92–111. issn: 0749-596X.

[3] J. Cholin and W. J. Levelt. “Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level.” In: Cognition 24 (2009), pp. 662–684. doi: 10.1080/01690960802348852.

[4] Werner Enninger and Joachim Raith. “The Design Features of Glosso-lalia.” In: Kodikas 6.1-2 (1983), pp. 59–70.

[5] S. Gahl. “”Thyme” and”Time” are not homophones. Word durations in spontaneous speech.” In: Language 84.3 (2008), pp. 474–496.

[6] W. J. Levelt, A. Roelofs, and A. S. Meyer. “A theory of lexical access in speech production.” In: The Behavioral and brain sciences 22.1 (Feb. 1999).

[7] F. Tomaschek et al. “Vowel articulation affected by word frequency.” In: Proceedings of the 10th ISSP. Cologne, 2014, pp. 425–428.

Der Zusammenhang zwischen selbsteingeschätzter und wahrgenommener Maskulinität in männlichen Stimmen

Fabienne Neder, Melanie Weirich und Adrian P. Simpson
Institut für germanistische Sprachwissenschaft, Friedrich-Schiller-Universität Jena
fabienne-neder@gmx.net, melanie.weirich@uni-jena.de, adrian.simpson@uni-jena.de

Verschiedene Studien haben mögliche akustische Parameter analysiert, die dazu führen, dass ein männlicher Sprecher als maskulin klingend wahrgenommen wird [1, 2, 3 und 4]. Sprecher, die als weniger maskulin wahrgenommen werden, weisen tendenziell eine höhere mittlere Grundfrequenz und Grundfrequenzvariation, einen größeren Vokalraum, höhere CoG und niedrigere skewness-Werte im /s/ auf [2 und 3]. Männer, die einen höheren F2 haben, werden als weniger maskulin beurteilt, wohingegen Sprecher mit einem hohen F1 eher als schwul beurteilt werden [5]. Studien, die sich mit dem Zusammenhang von akustischen Parametern und *selbst*eingeschätzter Maskulinität beschäftigen, sind seltener. Kachel et al. (2017) konnten zeigen, dass Männer, die sich als maskuliner beurteilen, einen tieferen F2 und niedrigere CoG im /s/-Laut aufweisen. Weirich & Simpson (2017) haben einen Zusammenhang von Vokalraumgröße und selbsteingeschätzter Geschlechteridentität (gemessen anhand der F+ Skala der *German version of the Extended Personal Attributes Questionnaire*, GEPAQ, [6]) gefunden.

Diese Untersuchung überprüft, ob selbsteingeschätzte und wahrgenommene Maskulinität korrelieren. Dazu haben 21 Hörer*innen (11m, 10w) die Maskulinität von 19 männlichen Sprechern eingeschätzt. Die Teilnehmer*innen haben Stimulus-Paare gehört und mussten entscheiden, welcher der beiden maskuliner klingt (auf einer Skala von 0 bis 3). Der Stimulus bestand nur aus dem Wort *Tasse*, entnommen aus einem semi-spontansprachlichen Korpus [7 und 8]. Des Weiteren wurden akustische Parameter untersucht, die für die Wahrnehmung als mehr oder weniger maskulin verantwortlich sein könnten. Die untersuchten Parameter waren F1 und F2 im /a/, CoG und skewness im /s/ sowie die Grundfrequenz des /a/. Ein lineares gemischtes Model mit der wahrgenommenen Maskulinität als abhängige Variable und den *random effects* Hörer, Sprecher 1 und Sprecher 2 des Paares wurde über die Daten des Perzeptionstests gerechnet. Als mögliche Einflussfaktoren wurden das Geschlecht der Hörer*innen, die selbsteingeschätzte Maskulinität der Sprecher (GEPAQ) und die akustischen Parameter (F1, F2, CoG, skewness und f0) betrachtet. Das Modell, welches die Daten am besten erklärt, enthält den Einflussfaktor f0 ($\chi^2(1) = 37.07$, $p < .001$) und die Interaktionen von GEPAQ * Geschlecht der Hörer ($\chi^2(3) = 13.81$, $p < .01$) und F1 * Geschlecht der Hörer ($\chi^2(2) = 8.45$, $p < .05$). Abbildung 1 verdeutlicht die gefundenen Zusammenhänge: Je maskuliner ein Sprecher wahrgenommen wurde (positiver Wert auf der y-Achse), desto geringer war sein GEPAQ-Wert, sein F1 im /a/ und seine Grundfrequenz. Alle Korrelationen (Pearson) sind negativ und signifikant ($p < .001$). Die Interaktionen mit dem Geschlecht der Hörer*innen spiegeln sich dadurch wider, dass der gefundene Zusammenhang bei den Hörerinnen (rechts) stärker ist als bei den Hörern (links).

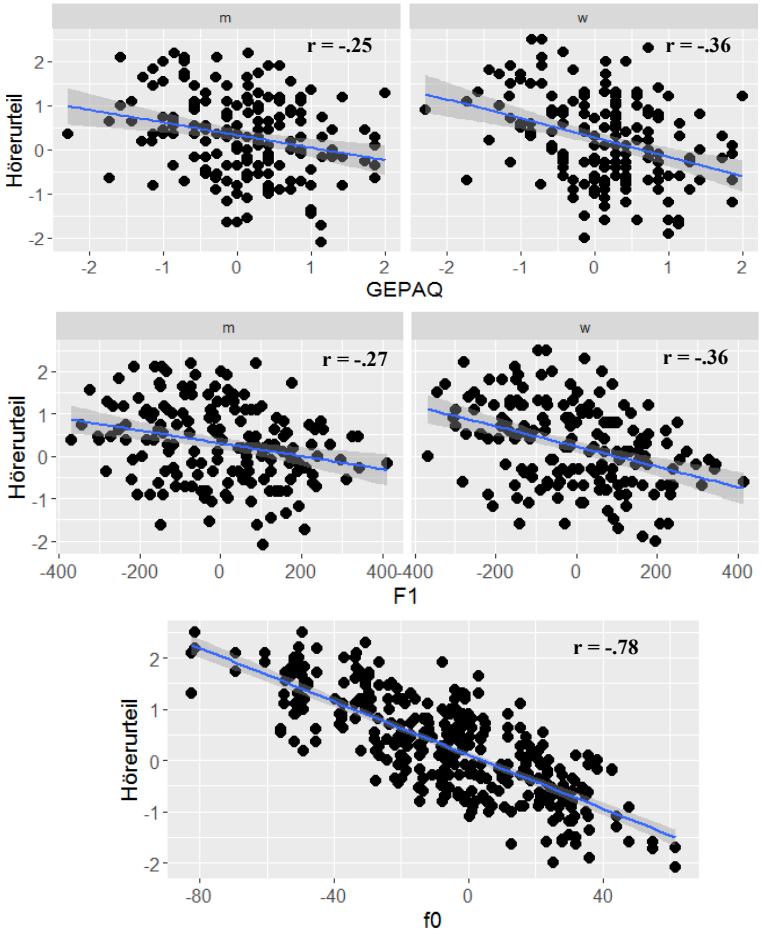


Abbildung 1: Zusammenhang zwischen wahrgenommener Maskulinität (Hörerurteil) und selbsteingeschätzter Maskulinität (GEPAQ), F1 und f0

Literatur

[1] Terango, L. (1966). Pitch and duration characteristics of the oral reading of males on a masculinity-femininity dimension. *Journal of Speech & Hearing Research*, 9(4). pp.590-595.

[2] Avery, J. D., and Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *The Journal of the Acoustical Society of America*, 99(6). pp.3738-3748.

[3] Guzik, K. M. (2004). Acoustic analysis of phonetic parameters of less masculine sounding German speech. In J. Harrington & C. Mooshammer (Eds.), *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK),* 36. pp.15–29.

[4] Kachel, S., Simpson, A.P., & Steffens, M.C. (in press). "Do I sound straight?" – Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men's speech. *Journal of Speech, Language, and Hearing Research.*

[5] Munson, B. (2007). The Acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech*, 50(1). pp.125-142.

[6] Runge, T. E., Frey, D., Gollwitzer, P. M., Helmreich, R. L., and Spence, J. T. (1981). Masculine (instrumental) and feminine (expressive) traits. A comparison between students in the United States and West Germany. *Journal of Cross-Cultural Psychology*, 12(2). pp.142-162.

[7] Weirich, M. und Simpson, A.P. (2016). Changes in IDS and ADS during parental leave – project sketch and first results of pilot studies. *Proceedings of the 12th meeting on "Phonetik und Phonologie im deutschsprachigen Raum* (P&P12), Munich. pp.224-227.

[8] Weirich, M. und Simpson, A.P. (2017). Acoustic correlates of parental role and gender identity in the speech of expecting parents. *Interspeech*, Stockholm. pp.924-928.

Wanted! - Assessing effects of distraction on working memory in speech perception

Matthias Wilke[1], Daniel Duran[2], Natalie Lewandowski[1]
[1]University of Stuttgart
[2]University of Freiburg
matthias.wilke@ims.uni-stuttgart.de, daniel.duran@germanistik.uni-freiburg.de, natalie.lewandowski@ims.uni-stuttgart.de

We present "*Wanted!*", a 3D video game for assessing the effects of distraction on working memory during speech perception. With recent technological advances and widespread social acceptance for video games, numerous studies have shown their usefulness in psychological and psycholinguistic experiments (e.g. [1]). A crucial aspect of games is their interactivity, distinguishing them from other media such as film, still images or audio recordings as well as from classical computer-based experimentation frameworks, allowing subjects to immerse in given tasks in a more natural way. Previous experiments implement tasks "of alertness, orienting, and executive control" [2] or reaction tests [3] within virtual environments, applying distraction in order to interfere with participants' cognitive, sensory and motoric capacities. There is, however, a lack of research harnessing computerized games for memory tests – especially in research on speech perception.

For that reason, we have developed a state-of-the-art, first-person 3D video game to examine the interactions between attention, cognitive load, working memory and speech processing, which have been adressed in previous laboratory experiments [4, 5, 6]. In a free recall task [7], we apply varying combinations of auditory (speech and non-speech) and visual distraction during memorization and recall. A comparison is conducted to identify visual effects similarly to the *irrelevant sound* or *unattended speech* effect [8], which describes the phenomenon of working memory deterioration during exposure to background noise or speech. During nine in-game days or *Trials*, players assume the role of the identikit expert's holiday replacement in a police station. Seven verbally presented crime features need to be memorized during a conversation with an eyewitness (memorization) and reproduced in a visual multiple-choice test shortly thereafter to create a wanted poster (recall). The game is set in a strongly scripted, linear 3D environment, guiding players through each phase of the experiment while still allowing for exploration of the detailed game world.

In our pilot study with 16 German native speakers playing the game, auditory distraction during memorization and recall was found to impair working memory efficiency by over 10% each. Visual distraction had significant impact during memorization (see Figure 1) with a performance deterioration of 7%, yet it did not yield significant effects during recall (see Figure 2).

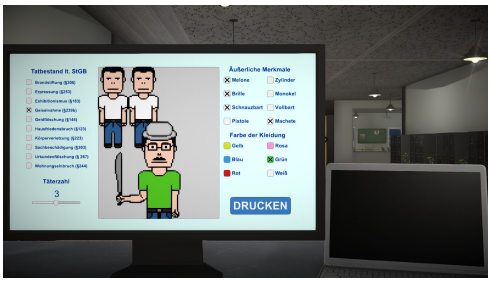
Finally, we make numerous suggestions for harvesting the full potential of our game prototype for research into speech perception, processing and cognition in the future.

Figure 1: *Memorization phase with and without visual distraction*

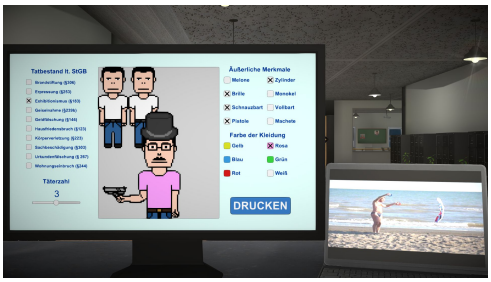
(a) No visual distraction during memorization.

(b) One of four possible visual distractions during memorization.

Figure 2: *Recall phase with and without visual distraction*



(a) No visual distraction during recall.



(b) One of four possible visual distractions during recall.

References

[1] D. Duran, N. Lewandowski, and A. Schweitzer, "A 3D computer game for testing perception of acoustic detail in speech," *POMA*, vol. 28, p. 060004, 2016. [Online]. Available: http://asa.scitation.org/doi/abs/10.1121/2.0000493

[2] A. Berger, L. Jones, M. K. Rothbart, and M. I. Posner, "Computerized games to study the development of attention in childhood," *Behavior Research Methods, Instruments, & Computers*, vol. 32, no. 2, pp. 297–303, Jun 2000.

[3] J. L. Harbluk, P. C. Burns, M. Lochner, and P. L. Trbovich, "Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces," in *Proc. of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2007, pp. 16–22.

[4] V. Taler, G. P. Aaron, L. G. Steinmetz, and D. B. Pisoni, "Lexical neighborhood density effects on spoken word recognition and production in healthy aging," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 65B, no. 5, pp. 551–560, 2010.

[5] S. L. Mattys and L. Wiget, "Effects of cognitive load on speech recognition," *Journal of Memory and Language*, vol. 65, no. 2, pp. 145–160, Aug. 2011.

[6] A. A. Zekveld and S. E. Kramer, "Cognitive processing load across a wide range of listening conditions: Insights from pupillometry: Processing load across a wide range of listening conditions," *Psychophysiology*, vol. 51, no. 3, pp. 277–284, Mar. 2014.

[7] K. A. Klein, K. M. Addis, and M. J. Kahana, "A comparative analysis of serial and free recall," *Memory & Cognition*, vol. 33, pp. 833–839, 2005.

[8] L. A. Farley, I. Neath, D. W. Allbritton, and A. M. Surprenant, "Irrelevant speech effects and sequence learning," *Memory & Cognition*, vol. 35, no. 1, pp. 156–165, 2007.

The timing of turn-taking in high-functioning autism

Simon Wehrle[1], Francesco Cangemi[1], Kai Vogeley[2] & Martine Grice[1]

[1]*IfL Phonetik, University of Cologne*
[2]*Department of Psychiatry, University Hospital Cologne*

Conversational partners in dialogue usually aim to reach a high degree of entrainment and rhythmicity. One objective is the smooth transition from the conversational turn of one participant to that of the interlocutor. To ensure this, there should be neither an overlap across turns nor too large a gap between them. Following [1] we can identify three types of turn transition: gaps (silent intervals), between-overlaps (overlapping speech before the transition) and within-overlaps (one turn being contained entirely within another). Cross-linguistic evidence has shown overwhelmingly that gaps are the preferred type of transition. Whilst turn-taking can be observed in domains other than language (e.g music [2]) and is found in a number of other species (e.g. marmosets [3]), the split-second timing typical of human spoken communication (with a modal value for gaps of around 250ms [4]) seems to be a unique achievement [5]. Such rapid transitions require considerable cognitive effort, as interlocutors must predict the end of a dialogue partner's turn while at the same time planning their next utterance. Such skills in predicting the behaviour of others are said to be impaired in people with autism [6], who are known to show non-typical behaviour in perspective-taking and in the decoding of linguistic cues [7]. Nothing is known to date about turn-taking in autistic speech. We investigated turn-taking in a parallel corpus of Map Task speech by adults diagnosed with high-functioning autism (HFA) and neurotypical controls (NT). Our general hypothesis is that HFA subjects will reach a lesser degree of entrainment. Across a total of 1514 turn transitions, for gaps and between-overlaps we found that in the HFA dyads there is a larger number of overlap transitions (Fig. 1) compared to NT dyads, which instead show a clear preference for gap transitions. Using the continuous measure of Floor Transfer Offset (FTO) reveals a similar trend (Fig. 2) and shows that autistic speakers produce more long overlaps. Analysing within-turn overlaps reveals the same pattern, with twice as many for the HFA group taken as a proportion of the number of turns. The fact that this increase in overlaps in HFA is shown both in discrete terms (Fig. 1) and in durational measurements (Fig. 2) is of interest regarding the broader theoretical issue of the interplay between discrete categories and continuous parameters in the characterisation of speech [8]. Excluding any overlaps containing backchannels, we still find a higher number of the remaining conversationally "unprincipled" overlaps for HFA (29.4%; NT: 21.5%). These differences in turn-taking are of considerable relevance for everyday communication in themselves, but taken together with other deviances in autistic speech concerning e.g. prosody, reference or irony, these features might (only) together add up to form the often described but always underspecified impression of "oddness" of speech in high-functioning autism.

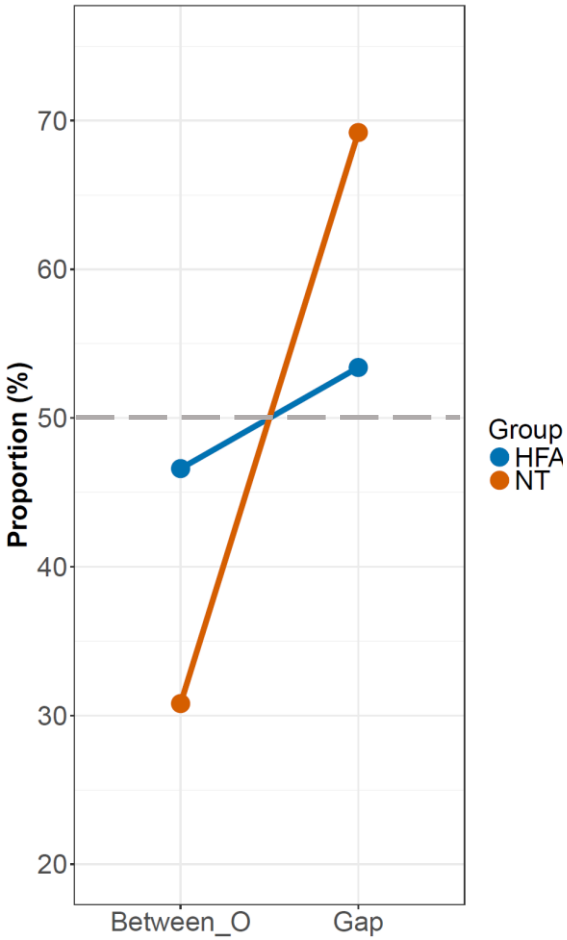


Figure 1: *Proportion of turn transitions as overlaps between turns (Between_O) and gaps between turns (Gap) by group.*

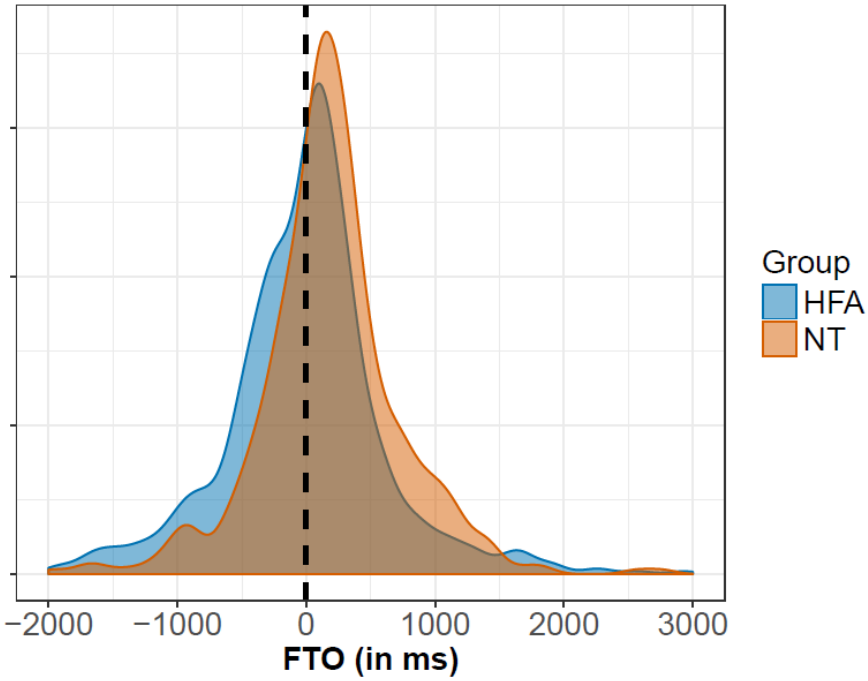


Figure 2: *Floor Transfer Offset (FTO) values of gap- and overlap transitions between turns by group.*
Negative values represent overlaps, positive values represent gaps.

References

[1] S. C. Levinson & F. Torreira, "Timing in turn-taking and its implications for processing models of language", *Frontiers in Psychology*, 6, pp. 10-26, 2015.

[2] S. Hawkins, I. Cross & R. Ogden, "Communicative interaction in spontaneous music and speech", in *Music, Language and Interaction*, M. Orwin, C. Howes & R. Kempson (Eds.), London: College Publication, 2013, pp. 285-329.

[3] D.Y. Takahashi, D.Z. Narayanan & A.A. Ghazanfar, "Coupled oscillator dynamics of vocal turn-taking in monkeys", *Current Biology*, 23(21), pp. 2162-2168, 2013.

[4] T. Stivers, N.J. Enfield, P. Brown, C. Englert, M. Hayashi, M., T. Heinemann, ... & S. C. Levinson, "Universals and cultural variation in turn-taking in conversation", *Proceedings of the National Academy of Sciences*, 106(26), pp. 10587-10592, 2009.

[5] S.C. Levinson, "Turn-taking in human communication–origins and implications for language processing", *Trends in cognitive sciences*, 20(1), pp. 6-14, 2016.

[6] K. Vogeley, "Two social brains: neural mechanisms of intersubjectivity", *Philosophical Transactions of the Royal Society B*, 372(1727), 2017.

[7] M. Grice, M. Krüger & K. Vogeley, "Adults with Asperger syndrome are less sensitive to intonation than control persons when listening to speech", *Culture and Brain*, 4(1), pp. 38-50, 2016.

[8] M. Grice, S. Ritter, H. Niemann & T.B. Roettger, "Integrating the discreteness and continuity of intonational categories", *Journal of Phonetics*, 64, pp. 90-107, 2017.

AUTORENVERZEICHNIS