SPPL2020: 2nd Workshop on Speech Perception and Production across the Lifespan

Book of Abstracts

UCL, London, UK, 30-31 March 2020

Division of Psychology and Language Sciences, Chandler House

Sponsored by

UCL

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Workshop website: http://sppl2020.org

Front cover photo by Tomas Sereda

Book of abstracts of 2nd Workshop on Speech Perception and Production across the Lifespan (SPPL2020) DOI: 10.5281/zenodo.3732383

Editor: Linda Taschenberger

Foreword

This workshop is the second in the 'Speech Perception and Production across the Lifespan' (SPPL) series, following the initial workshop that took place at UCL in April 2017 (SPPL2017). Lifespan studies can further our understanding of the extent and direction of changes for key measures of speech communication and of how these changes interact with cognitive, social or sensory factors. However, studies that take a full lifespan approach are still rare, partly due to the difficulty of finding materials and experimental designs that are appropriate for use across an extended age range. The aim of SPPL is to bring together researchers carrying out lifespan studies in speech and language sciences with researchers that may be focused on different developmental stages, e.g. early development and ageing. We hope that the fruitful interchanges that can come from bringing researchers together can further spur the development of lifespan studies.

World events linked to the Covid-19 pandemic have affected our plans for SPPL2020. By the time that this crisis had started to fully develop, abstracts had been received and reviewed and we had also received the final versions of accepted abstracts. In view of the worsening crisis, a decision was made on 12 March that it would not be possible to hold the workshop in London on 30-31 March. Further events have confirmed that this was the only possible decision. Rather than an open-ended postponement, it was decided to hold a virtual workshop and to give the opportunity to authors to submit either a poster or narrated slides to be made available online on the SPPL2020 website (www.sppl2020.org) with open access. All registration costs were also reimbursed. A discussion forum was also added to the website to facilitate interchanges and discussions of individual papers. Although submissions are optional and may be delayed for some due to current restrictions in many countries, we hope that this vSPPL2020 will still give some of the opportunity for the dissemination of information and for discussions that we were looking forward to having in London.

This Book of Abstracts contains extended abstracts for 40 accepted papers from eleven countries worldwide that were due to be presented at the oral and poster sessions of SPPL2020 workshop as well as for the four invited talks and an opening talk.

Finally, we thank members of the SPPL2020 organising committee, the Economic and Social Research Council (ESRC) and Division of Psychology and Language Sciences for their financial sponsorship and, of course, all contributors. The workshop is organised under the aegis of ESRC project ES/P002803/1 on 'Speech masking effects in speech communication across the lifespan'.

Valerie HAZAN and Outi TUOMAINEN, Co-Chairs SPPL2020

SPPL2020 Organising Committee

Bronwen EVANS (UCL)

Valerie HAZAN (UCL) – Co-chair

Esther JANSE (Radboud University Nijmegen/Max Planck institute Nijmegen)

Merle MAHON (UCL)

Caroline NEWTON (UCL)

Linda TASCHENBERGER (UCL)

Outi TUOMAINEN (UCL) – Co-chair

Table of Contents

Invited talk 1

Oral session 2

Invited talk 2

Invited talk 3

Oral session 3

Poster session 2

Invited talk 4

Oral session 4

Speech masking effects in speech communication across the lifespan

Outi Tuomainen, Linda Taschenberger and Valerie Hazan
Department of Speech Hearing and Phonetic Sciences, UCL, London
o.tuomainen@alumni.ucl.ac.uk

This talk will give an overview of a three-year project that is investigating the impact of different types of maskers on speech communication in individuals aged 8 to 85 years. The degree to which communication with others is successful is often affected by the presence of noise and other voices in the environment (for a review, see Mattys et al., 2012). The impact of background noise on speech communication appears to be greater for certain age groups such as children and older adults (e.g., Gordon-Salant, 2005). Other groups such as non-native speakers or individuals with hearing impairments may be even more affected by poor communicative conditions. Part of this adverse effect is due to physical masking by interfering sounds (energetic masking EM) on the speech which needs to be understood; however, if the disrupting sound is meaningful speech (informational masking IM), this can also cause further difficulty. Previous work suggests that informational masking causes relatively more disruption for children and older adults than for young adults (e.g. Schoof and Rosen, 2014 for older adults) but these findings are based on laboratory tests that are far from realistic communication.

Although studies have investigated the impact of adverse conditions on speech communication in different age groups, this is the first study, to our knowledge, to take a lifespan view, with participants aged from 8 to 85 years using a common experimental design. Another important aspect of this study was that the impact of informational and energetic masking was investigated within a communicative task (Diapix, Van Engen et al., 2010) rather than from the perception of read words or sentences. Compared to read speech, speech recorded during this interactive problem-solving 'spot the difference' picture task carried out by pairs of talkers in a noisy background incorporates the dynamic adaptations that talkers make in speech communication to counter the effects of masking. In this task, we evaluate the impact of adverse conditions using measures reflecting speech communication efficiency rather than purely evaluating intelligibility. Finally, there is little evidence to date as to whether laboratory-based evaluations reflect the level of difficulty experienced in everyday life. This project aimed to relate evaluations of speech communication difficulties obtained from the diapix task and from background hearing tests with real-life ratings of difficulty collected in real time using a smartphone-based experiential sampling app over a two-week period with the same adult participants. Our research hypothesis was that masking that was primarily informational (here, three voices speaking in the background) would cause greater interference, e.g., as shown by greater dysfluencies or lower task efficiency, for some age groups such as children or older adults. We also predict that standardised tests may be poor predictors of communicative difficulty in real-life situations.

In Study 1, 114 individuals aged 8 to 85 years were recorded in age and sex-matched pairs while they carried out the diapix task using the DiapixUK picture sets (Baker and Hazan, 2011) in conditions varying in the informational (three voices in the background) and energetic masking (speech-shaped noise) present. A secondary task (pressing a bell when hearing a dog barking but suppressing when hearing a car horn honking) was added to make the task more cognitively demanding, thus reflecting real-life multitasking situations. Also, in order to give a more natural listening environment, we used Spatial Audio Simulation System software (Audio 3D) that mimics real room acoustics combined with head-related transfer functions in real-time. After completing each diapix task, both participants completed a paper-based questionnaire, answering four questions using an 11 point Likert scale. The questions were: 1. On average, how noisy did you experience the background noise you heard during the task? 2. Could you easily ignore the background noise? 3. Did you have to concentrate very hard to understand your partner? 4. Did you have to put in a lot of effort to understand your partner? Baseline measures of hearing (pure tone audiogram), speech perception (CCRM, BKB sentence intelligibility), cognitive function (tests of expressive vocabulary, letter-number sequencing, letter-digit substitution) and a standardised questionnaire of auditory disability (SSQ) were also collected. To establish SNR thresholds in IM/EM for simple speech materials, we used the adaptive CCRM test (described in Hazan et al., 2009 as 'WiNiCS test'); this was modelled on the Coordinate Response measure (Bolia et al., 2000), with either a 3 male-talker babble masker (CCRM-BABB) or speech-shaped noise masker (CCRM-SPSN), and a female voice as target. In Study 2, a self-selected group of 68 adult participants from this cohort were asked to report perceived communication difficulty and their listening environment on 6-7 occasions per day during a 2-week period using a smartphone-based app. Data from this study will be related to measures of communication effectiveness and SSQ data collected in Study 1.

A number of measures were obtained from these tests. For the diapix task in each condition, we extracted the following measures: transaction time, dysfluency (filled pause) rate, a discriminability measure (d') for the secondary task, ratings of perceived noise, interference, concentration and effort, and a number of acoustic-phonetic measures as in Hazan et al. (2016). From the CCRM, we obtained SNR thresholds for energetic and informational maskers; from the BKB sentence test, we obtained SNR thresholds for 67% intelligibility level for a male speaker heard in 8-talker babble. From the experiential sampling test, some measures showing the relation between perceived noise and effort were attained but these were calculated at group level as it proved difficult to obtain enough data per person for differing listening settings.

In this talk, we will present some results from our preliminary analysis of the complete data set.

References

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3), 761-770.

Bolia R. S., Nelson W. T., Ericson M. A., & Simpson B. D. (2000). A speech corpus for multitalker communication research. *Journal of the Acoustical Society of America*, 107, 1065–1066.

Gordon-Salant, S. (2005). Hearing loss and aging: new research findings and clinical implications. *J Rehabil Res Dev.*, 42, 9–24.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit? *Journal of Speech, Hearing and Language Research*, 52, 1510-1529.

Hazan, V. L., Tuomainen, O., & Pettinato, M. (2016). Suprasegmental Characteristics of Spontaneous Speech produced in Good and Challenging Communicative Conditions by Talkers aged 9 to 14 years old. *Journal of Speech, Language, and Hearing Research*, 59, S1596-S1607.

Mattys, S.L., Davis, M.H., Bradlow, A.R. & Scott, S.K. (2012). Speech recognition in adverse conditions: A review, Language and Cognitive Processes, 27, 953-978.

Schoof, T. & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience*, 6, 307.

Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M., & Bradlow, A.R. (2010). The Wildcat Corpus of Native- and Foreign-Accented English: communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53, 510-540.

Durations of words with various numbers of syllables across childhood

Mária Gósy
Research Institute for Linguistics, ELKHT and ELTE University, Budapest, Hungary
gosy.maria@nytud.hu

In spontaneous speech speakers regularly speed up and slow down their articulation depending on various factors, particularly the speaker's age. Temporal properties of words are influenced by universal (e.g., physiological including age), language-specific (e.g., phonetic realizations of phonemic differences of segments, frequency), and individual characteristics (e.g., speaker-related behaviour, emotion). A large number of studies discussed those factors that influence the duration of a word and the variability therein both in children and adults (Smith 1992; Bell et al., 2009).

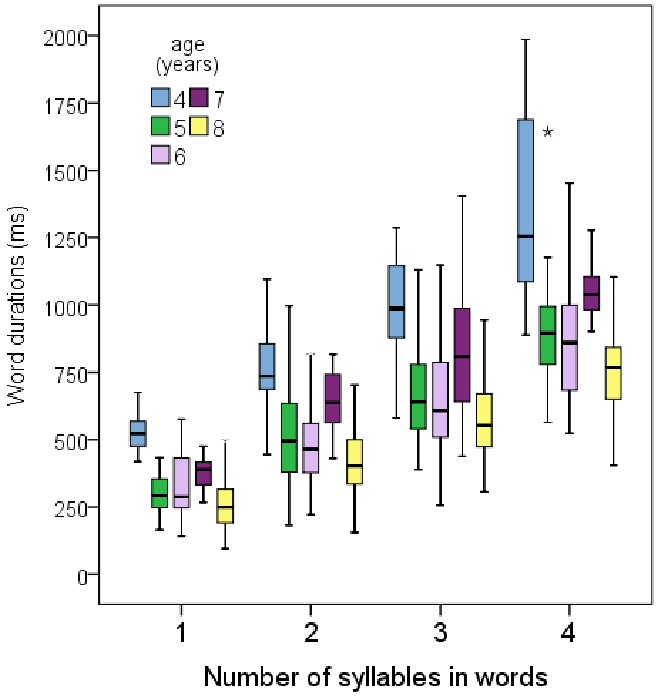
The retrieval of a word requires the selection of a syntactically, semantically and phonologically appropriate word with the corresponding articulatory gestures. All those processes that are involved in word production develop during language acquisition (Redford 2015). Grammatical knowledge and mastering of articulation are particularly responsible for word duration during language acquisition (Hulme et al. 1984). The acquisition of the morphology of words in an agglutinating language, like Hungarian, shows a relatively rapid process as children's linguistic and cognitive capacities grow (Bunta et al. 2016). The acquisition of temporal patterns reflects motor skill development where some patterns are stored as acoustically linked articulatory schema while others emerge during fluent speech practice (Redford 2015). Meaning and frequency of words also matter to the acquisition of temporal patterns. A number of studies confirmed that young children's spoken words tend to be longer and more variable than those of older children and adults (e.g., Smith 1992; Lee et al. 1999; Flipsen 2002; Tomasello 2003).

Investigations revealed that verbs were articulated faster than nouns (e.g., Hirsh-Pasek and Michnick Golinkoff 2006). Children acquiring Hungarian, which has a rich morphology, are able to use suffixes of nouns, verbs and adjectives by the age of 3 (Bunta et al. 2016). How do complex morphology of words, growing articulation skills of children and skilled access to the mental lexicon contribute to word durations and variance across ages 4 to 8? The goal of the present (cross-sectional) research was to show the changes of word durations in spontaneous speech of monolingual Hungarian-speaking children. We hypothesized that (i) the duration of words would be shorter as age increases, (ii) nouns and adjectives would be longer than verbs with the same number of syllables, and (iii) the durations of words would show a relatively constant increase with the increasing number of syllables irrespective of age.

Speech samples of 50 Hungarian-speaking children (ages between 4 and 8 years) – narratives of children's life, family, and hobbies of about 8 minutes each – were randomly selected from the GABI children's speech database (Bóna et al. 2019). There were 10 children in each age group (half of them girls). All children had normal hearing, and none of them had any speech defects. No known history of delayed onset of language acquisition were reported in any of the children. 6-year-olds started their schooling while 7-year-olds completed their first year of education. Speech samples were manually annotated according to part of speech (nouns, verbs and adjectives), number of syllables (from 1 to 4) and durations of words using Praat (Boersma and Weenink, 2014). No words were considered that were pronounced in the vicinity of a pause. The word boundaries were identified in the waveform signal and spectrogram display via continuous listening to the words. Markers were inserted at the closure and release of obstruents, at the onset and offset of voicing, and second formant information was considered following standard acoustic-phonetic criteria. A specific script was written to obtain the values automatically. To test statistical significance, General Linear Mixed Models were used (SPSS 20.0 version). A total of 3,460 words were selected to be analyzed. Although the number of words was different across ages, the ratio of words according to their lengths was similar across age groups.

The mean duration of all words was 656 ms, and the age-specific values decreased from the age of 4 up to the age of 8 with the only exception of those produced by 7-year-olds (918 ms, 623 ms, 585 ms, 646 ms, 508 ms, respectively). On average, seven-year-olds took longer to articulate the words than all the other children, with the exception of 4-year-olds. As expected, the more syllables the words contained the longer their durations were in all age groups. However, the differences in word duration depending on the word length showed age-specific patterns (Fig. 1). The temporal differences between the monosyllables and the disyllabic words and the disyllabic and tri-syllabic words were similar in 5- and 6-year-olds while there were significant differences between them in 4-year-olds and in schoolchildren. There were significant differences in durations between words of three and four syllables in all children.

Figure 1. *Word durations across ages and word lengths*



Word durations showed significant differences depending on the parts of speech (mean values of nouns: 667 ms, verbs: 621 ms, adjectives: 581 ms). Verbs turned out to be shorter than nouns in all age groups with the exception of those articulated by the 4-year-olds. 4-year-olds produced verbs longer (mean value: 1100 ms) than both nouns (mean value: 872 ms) and adjectives (mean value: 780 ms). Adjectives were the shortest in all age groups. Word lengths, however, influenced the durations of the analyzed parts of speech. For example, monosyllabic adjectives were longer than both nouns and verbs in 5-year-olds, and monosyllabic and tri-syllabic adjectives turned out to be longer than verbs in 8-year-olds. There were significant interactions among all analyzed factors.

Although the duration of words gradually decreased with age in four of the five age groups, the 7-year-olds' word articulation was exceptionally slow. This can be explained by their starting to learn reading and writing at school and presumably also the increasing number of newly acquired words at this age. We assumed that nouns would be longer than verbs with the same number of syllables, for which evidence has been found in the present study. Adjectives were pronounced relatively shortly by all children but it was not independent of word length. We suppose that these findings are in connection with the acquisition of both the various parts of words and the great variety of the corresponding suffixes. The durations of words did not show a constant increase with the increasing number of syllables which seems to be mainly a developmental factor including, for example, cognitive abilities, memory, attention span, grammatical knowledge, increasing individual vocabulary, and articulation skills (Redford 2015; Fletcher et al. 2016). All of them influence the temporal control over word durations in spontaneous utterances in children, several aspects of which were clearly evidenced in the case of 4-year-old and 7-year-old Hungarian-speaking children.

References

Bell, A., Brenier, J. Gregory, M., Girand, C. & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. Version 5.4.1. Retrieved from: http://www.praat.org.

Bóna, J., Vakula, T., & Váradi, V. (2019). GABI – Hungarian child language and speech database and information repository. *The Phonetician, 116*, 41–52.

Bunta, F., Bóna, J. & Gósy, M. (2016) HU-LARSP: Assessing children's language skills in Hungarian. In P. Fletcher, M. J. Ball, & D. Crystal (Eds.), *Profiling grammar. More languages of LARSP* (pp. 80–98). Bristol, Buffalo, Toronto: Multilingual Matters.

Fletcher, P., Ball, M. J., & Crystal, D. (Eds.). (2016). *Profiling grammar. More languages of LARSP*. Bristol, Buffalo, Toronto: Multilingual Matters.

Flipsen, P. (2002). Articulation rate and speech-sound normalization failure. *Journal of Speech, Language, and Hearing Research, 46*, 724–737.

Hirsh-Pasek, K., & Michnick Golinkoff, R. (2006). *Action meets words. How children learn verbs*. Oxford: Oxford University Press.

Hulme, Ch., Muir, C., Thomson, N., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241–253.

Lee, S., Potamianos, A. & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America, 105*, 1455–1468.

Redford, M. A. (2015). The acquisition of temporal patterns. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 379–403). New Jersey: Wiley, Blackwell.

Smith, B. L. (1992) Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America 91*(4), 2165-2174.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Changes in phonetic detail as a matter of discourse and aging: Evidence from a longitudinal study on French

Susanne Fuchs[1], Annette Gerstenberg[2], Laura L. Koenig[1,3,4]
[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft Berlin, Germany, [2]University of Potsdam, Germany
[3]Adelphi University New York, USA [4]Haskins Labs New Haven, USA

Variation in speech production of older speakers can have multiple origins which are due to biological, cognitive, social and discourse factors. Studying variation synchronically or diachronically, i.e. comparing speakers across generations or, instead, by means of longitudinal data over the lifespan may also reveal different results. For example, several studies (Linville 2001; Yuan et al. 2006) have reported a slower speech rate for older speakers, comparing across generations. However, in our own work (Gerstenberg et al. 2018), based on longitudinal data of German and French speakers, differences in speech rate and pauses were generally not found.

It has also been reported that older female speakers have lower fundamental frequencies (Linville 2001); however, Stathopoulos et al. 2011 found a rise after the age of 60.

Very little research investigates changes in the upper vocal tract and the respective articulation of older speakers. *"...There is a scarcity of research on the effects of aging in the oral and pharyngeal lumina and consequently an absence of any speech production models that address such aging effects …"* (Xue & Hao, 2003: 678). Filling this gap is crucial for understanding of speech communication in older speakers, and a precondition to scientific discussions about sound change, especially when comparing populations that differ in age. As long as the origin of variation is unclear, the factors contributing to aging effects are likewise unclear. In a study on younger and older speakers using an acoustically-based method to determine vocal tract length, Xue and Hao (2003) found no differences in vocal tract length, but considerable differences in oral cavity length and volume for older speakers. These morphological differences had an impact on the acoustics of vowel production and caused a significant lowering of formant frequencies, specifically of the first formant. Effects were vowel-specific and larger for low than for high vowels. These acoustic results, particularly in F1 lowering, have repeatedly been found, meaning that a biological origin might be plausible (Harrington, 2006; Winkler, 2009; Reubold et al. 2010; Hawkins & Midgley, 2005; Linville & Rens, 2001).

The aim of this study is to investigate phonetic properties of vowels in frequently-occurring words with important discourse functions in French. We take the advantage of the LangAge corpus, which consists of biographical interviews conducted with older speakers (>70 years) of French. From the LangAge longitudinal series, we selected female speakers recorded 10 years apart, 2005 and 2015, in a natural setting, with the same interviewer. So far, four out of ten female French speakers have been annotated: first speaker: a015: 78 (2005), 88y. (2015); second speaker a016: 71y. (2005), 81y. (2015); third speaker a026: 78y. (2005), 88y. (2015); fourth speaker: a049: 72y. (2005), 82y. (2015). None of them reported hearing problems. For the analysis we chose the discourse marker *alors* 'so', frequently used in turn-initial position, a fact that raises its functional comparability in the longitudinal perspective (in the token frequency list, it is at the 31st position after function words, auxiliaries and conjunctions). The other word is the hesitation marker or filler *euh* 'uh'. Both were chosen because they are frequently occurring tokens in the narrative corpus used by every speaker (*euh* is at the first position in the token frequency list). The marker *euh* is often considered to be a disfluency, in a class with repetitions, repairs, filled pauses and false starts which have received increasing interest in phonetics and cognitive science, because they represent the "complexity of spontaneous speech over the type of speech observed in the laboratory" (Shriberg, 2001: 167; see also Kairet 2019).

With increasing age, the marker *euh* shows a significant decrease in frequency of occurrence in the LangAge corpus, which sheds new light on the very notion of disfluency and raises interest in its property as a potential age-sensitive linguistic feature (Gerstenberg et al. 2018).

From these tokens, the corresponding vowels /a, o/ in *alors* and /2, 4/ (ASCII IPA) were labelled manually using Praat (version 6.1.08). Fundamental frequency was automatically extracted from the whole vowel and from the vowel midpoint within the range from 140-400 Hz. In some data, voices became creaky and f0 was only partially present or no data were obtained. Formant values were calculated from the vowel midpoint using the Praat default settings (maximum value = 5500 Hz, window length = 25 ms, number of formants = 5). Overall for each speaker between 127-192 tokens (=n) were taken into account. For all speakers we analysed 594 euh which were pronounced with an /2/, 172 with an /4/, 224 /a/ and 244 /o/ in alors (n= 1234 vowels).

Preliminary results from the four speakers reveal no consistent changes in f0 with age with different words in discourse (Figure 1). First formant frequencies (Figure 2) were higher at the older age, but only for all speakers in /2/ of *euh*. Finally, the F2 values (Figure 3) did not show age related differences in general. Differences in the different words may not only be a result of the different vowels, but also due to coarticulation (see e.g. the unaccented /a/ in *alors* shows a large variability in F1, see Figure 2), where *euh* can be considered as free of coarticulatory constraints, while the vowels in *alors* are heavily coarticulated and rather short in duration.

These preliminary findings are in general agreement with previous literature (especially Winkler, 2009; Reubold et al., 2010), and may reveal additional insights into the role of discourse markers. We suggest that discourse has a crucial function and the occurrence of a word in discourse can partially override age-related changes in f0 and vowel formants.

Figure 1: *Fundamental frequency during different vowels (subplots) for different speakers (x-axis). Boxes are colour coded according to the Time of recording (2005 in red, 2015 in green).*

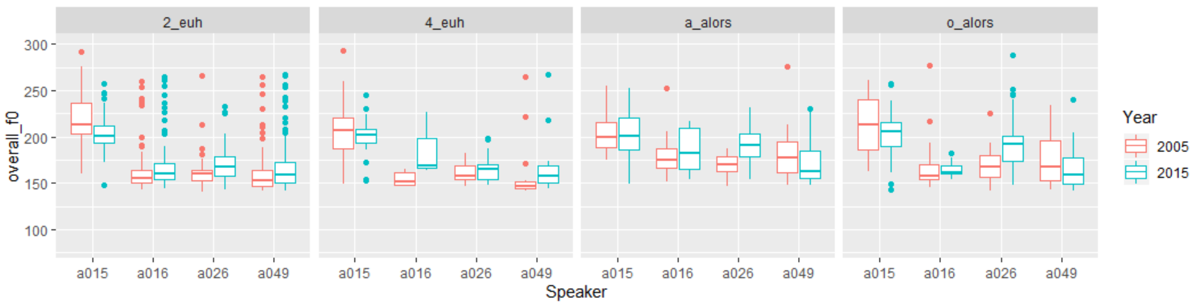
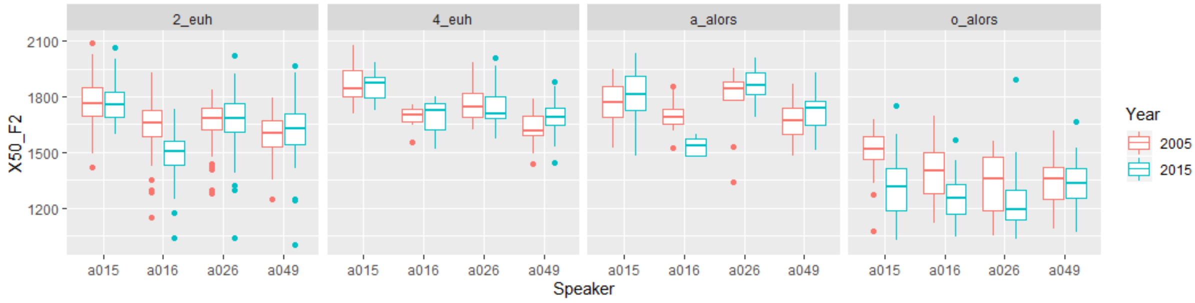


Figure 2: *Same as Figure 1 but for F1 (Hz) measured at vowel midpoint.*

Figure 3: *Same as Figure 1, but for F2 in Hz at vowel midpoint.*



References

LangAge corpus = www.langage-corpora.org

Gerstenberg, A., Fuchs, S., Kairet, J., Schroeder, J. & Frankenberg, C. (2018). A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. *Speech Prosody* 9, Poznan (Poland), 211-215.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, 34(4), 439-457.

Hawkins, S. & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35(02), 183-199.

Kairet, J. (2019). Silent pause duration and distribution in older womens' speech: a case study with within-speaker comparison. In O. Niebuhr, J. Neitsch, S. Berger, K. Fischer, J. Michalsky, S. Eisenberger, & M. Jelínek (Eds.), *Proceedings of the 1st International Seminar on the Foundations of Speech – Pausing, Breathing and Voice. University of Southern Denmark, Sønderborg, Denmark, December 1–3, 2019*. University of Southern Denmark, 61–63.

Linville, S. E. (2001). *Vocal Aging*. Australia, San Diego: Singular Thomson Learning.

Linville, S. E., & Rens, J. (2001). Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*, 15(3), 323-330.

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: a longitudinal analysis in adult speakers. *Speech Communication*, 52(7), 638-651.

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31(01). 153–169.

Stathopoulos, E. T., Huber, J. E. & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech Language and Hearing Research* 54(4), 1011–1021.

Winkler, R. (2009). *Merkmale junger und alter Stimmen: Analyse ausgewählter Parameter im Kontext von Wahrnehmung und Klassifikation*. Berlin.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *Journal of Speech, Language, and Hearing Research*, 46(3), 689-701.

Yuan, J., Liberman, M. & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation *Proceedings of Interspeech*, Pittsburgh (USA) 541–544.

When do neurophysiological correlates of word production change in the adult lifespan?

Tanja Atanasova [1], Giulia Krethlow [1], Raphael Fargier[2], Marina Laganaro[1]
[1]Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland
[2]Institute of Language, Communication and Brain - Aix-Marseille University, Laboratoire Parole et Langage, Aix-en-Provence, France

Introduction

Language is one of the most preserved cognitive functions in ageing, compared to other cognitive processes. Yet, age-dependent differences in word production performance can be observed in both accuracy and production latencies (Kavé et al., 2010; Kavé & Yafé, 2014; Stine-Morrow & Shake, 2009). Neurophysiological activity underlying word production has also been shown to vary between young and older adults, word production modifications being influenced by age-related changes (Valente & Laganaro, 2015; Mohan & Weber, 2018). As studies usually compare groups of young adults (20-30) to older adults it is unclear when such changes occur in the lifespan. We also want to take into consideration that at the neurobiological level brain development continues until the age of 35 and is certainly still in progress in adolescents. In this study we aimed to investigate the electrophysiological (EEG) and event-related (ERP) patterns underlying word production in a picture naming task, not only between the two extremities of the adult lifespan, but including intermediate age-groups and adolescents.

Method

High-density EEG was recorded on 120 French native speakers aged 14 to 80 years-old divided into six age-groups (14-16, 17-18, 20-30, 40-50, 60-70, 70-80). Participants were asked to overtly name, as fast and as precise as possible, 120 pictures corresponding to modal names for which several psycholinguistic variables were available from two French databases (Alario & Ferrand, 1999; Bonin et al., 2003). For each trial Reaction Times (RT) and Accuracy were calculated.

Results

Behavioral results show that only the participants of the oldest group (70-80) display slower production latencies relative to the other groups. By contrast, ERP microstates differ in adolescents and young adults as compared to all other older groups in a time-window associated to lexical-semantic processes (after the P1 component). Hence, distinct behavior is observed only in the oldest group of adults, while a different pattern appears in the neurophysiological data with the three youngest groups differing from all other groups.

Discussion

Results show that behaviorally only the 70-80 years-old (long with the young adolescent) differ from the rest of the adult groups. However, a different pattern emerges on ERP microstates: only the response-locked first Map follows the behavioral trend, in the stimulus-locked and in particular in the N170-like time-window adolescents and the youngest adults present distinct neurophysiological patterns than all other adult groups. This time-window has previously been associated with semantic processes in picture naming.

This unforeseen pattern observed in 20-30 years-old may reflect the ongoing maturation of young adults' brain wiring (Lebel & Beaulieu, 2011). However, another possible interpretation could consist of the fact that either changes in behavioral and brain activity start in the 40s but without behavioral decline until the age of 70.

Thus, it questions the relevance of investigating mental processes mostly with undergraduates (20-30 years old) to study cognitive processing.

References

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531-552.

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 158-167.

Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging*, *25*(3), 719.

Kavé, G., & Yafé, R. (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: independence or interdependence of skills? *American Journal of Speech-Language Pathology, 23*, 36–45.

Lebel, C., Beaulieu, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *Journal of Neuroscience, 31*, pp. 10937-10947.

Mohan, R., & Weber, C. (2018). Neural activity reveals effects of aging on inhibitory processes during word retrieval. *Aging, Neuropsychology, and Cognition*.

Stine-Morrow E., & Shake, M. (2009). Language in aged persons. *Encycl Neurosci 5*:337–342.

Valente, A., & Laganaro, M. (2015). Ageing effects on word production processes: an ERP topographic analysis. *Language, Cognition and Neuroscience, 30*:10, 1259-1272.

No evidence of a high variability benefit in phonetic vowel training for children

Gwen Brekelmans[1], Bronwen G. Evans[2], Elizabeth Wonnacott[1]
[1]Department of Language and Cognition & [2]Department of Speech and Hearing Sciences
University College London, United Kingdom
gwen.brekelmans.15@ucl.ac.uk bronwen.evans@ucl.ac.uk e.wonnacott@ucl.ac.uk

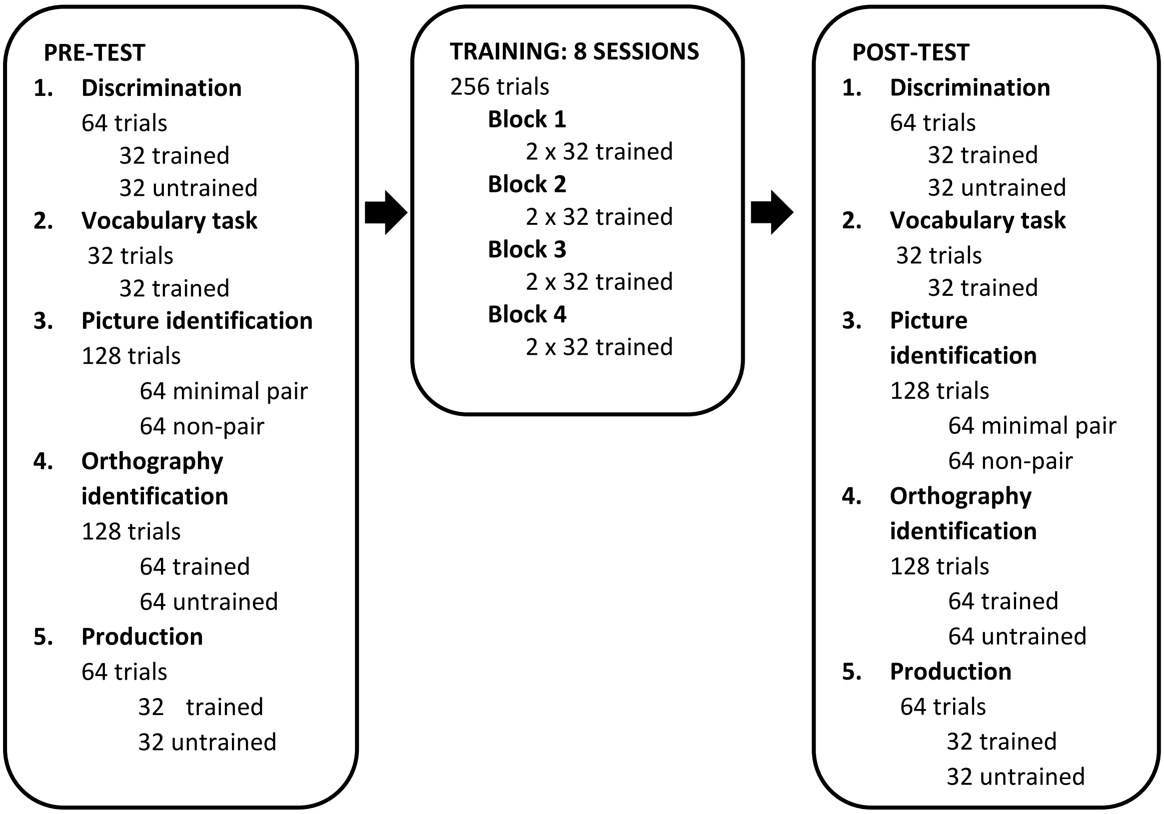
Acquiring an L2 speech contrast that does not exist in the L1 is often difficult. However, since seminal studies in which Japanese learners were successfully trained on the English /r/-/l/ contrast (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991) using high variability phonetic training (HVPT), this method has become a well-established standard paradigm. Their method of training learners on specific non-native phoneme contrasts critically used high variability (HV) input using multiple talkers and phonetic contexts (while earlier attempts which had used low variability (LV) input had proved unsuccessful). Since then, HV phonetic training has been successfully applied in many adult studies (e.g. Nishi & Kewley-Port, 2007), and has more recently gained ground in similar studies with children (Giannakopoulou, Uther, & Ylinen, 2013; Shinohara & Iverson, 2013)**.** However, only two studies to date (Evans & Martín-Alvarez, 2016; Giannakopoulou, Brown, Clayards, & Wonnacott, 2017) have directly investigated the effect of variability in the input for training children, and they found mixed results The current study aims to further investigate the effect of variability on phonetic training for children.

The present study investigates how children of two age groups differ in their L2 speech learning using the HVPT paradigm, and whether they both show the expected HV over LV advantage. In a two-week training study, two groups of Dutch learners of English, 7-8 year-olds and 11-12 year-olds, were taught English vowel contrasts.

On the basis of initial one-day phonetic training experiments where we found no evidence for an HV benefit, a computerised training paradigm was developed in which participants identified the correct member of a minimal pair; crucially, this training program used meaningful pictures in training in addition to orthography, to make the task more akin to vocabulary learning and to boost learning in general. As part of the training study, children were trained on four Standard Southern British English (SSBE) phoneme contrasts: /uː/-/ʊ/, /e/-/æ/, /ʌ/-/ɒ/, and /iː/-/ɔː/. These contrasts were chosen because the first three are notoriously difficult for Dutch learners, while the latter served as a baseline measure of learning. The main manipulation was whether the children received HV (4 talkers) or LV (1 talker) input in training. Participants received eight sessions of training of half an hour each, spread over two weeks.

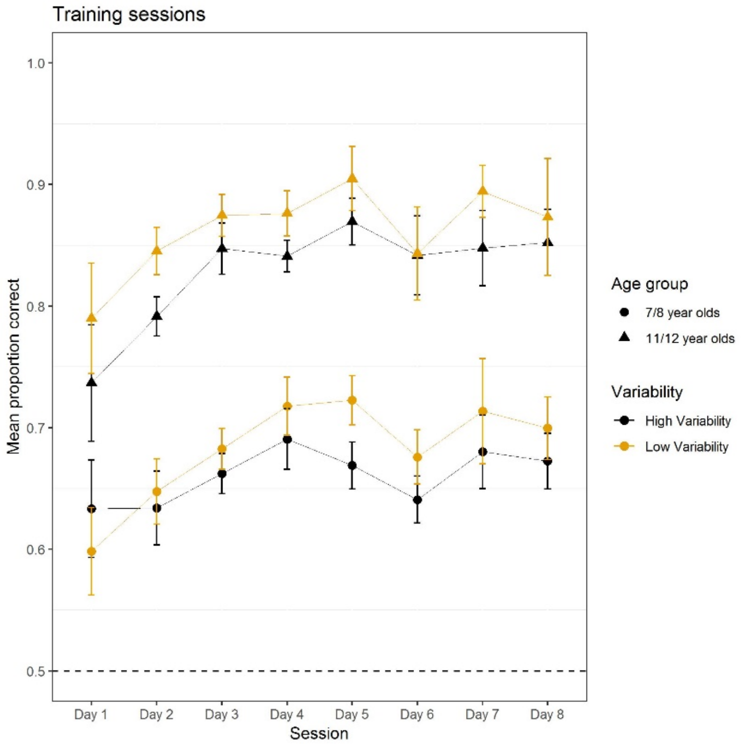
Potential effects of variability input in training were investigated using a pre/post-test design in which children completed a battery of tests examining their learning from pre- to post-test. Additionally, all tests measure generalisation to untrained talkers as novel talkers only were used in pre/post-test, as well as measuring generalisation abilities by testing children on both trained and novel items. A two-alternative forced-choice minimal pair picture identification task was used to measure phoneme identification ability, while phoneme discrimination was tested using a three-alternative forced-choice oddity task. Vocabulary was measured through a familiarity rating and translation of meaning, and the phoneme production task involved repeating trained and novel words to measure any improvements in vowel production accuracy. See Figure 1 for a detailed overview of the testing procedure.

Figure 1. *Overview of the procedure of training and the pre/post-tests, including the number and type of experimental trials per task. Task numbering shows the order in which the tasks were conducted. Pre- and post-test were each administered over 2 sessions, with the production task administered separately from the other tasks due to time constraints.*



Results show that overall, and contrary to the notion of greater plasticity in younger learners, older children not only outperformed younger children at pre-test, they also improved more across training (see Figure 2), and showed more improvement as a result of training in the discrimination, minimal pair picture identification and vocabulary tasks. In fact, although younger children showed improvements during training itself, they made no improvement at all on any of the pre/post-tests, suggesting that they do not generalize their learning to untrained talkers. Interestingly, even the 11-12 year-olds, who did improve on most tasks, did not show improvement on the task which tested whether they were able to make use of the orthography during training, with evidence of the null. Critically, there was no evidence of a HV benefit anywhere i.e. even for the 11-12 year-olds, who did show learning, and Bayes Factor analyses even found evidence for the null in some tests. These results suggest there is a trade-off between task complexity and potential variability benefits.

Figure 2. *Accuracy results for 7-8 year olds and 11-12 year olds over 8 sessions of Training, comparing accuracy for HV versus LV training input. The error bars indicate 95% CI, and the dashed line indicates chance level.*



References

Evans, B. G., & Martín-Alvarez, L. (2016). Age-related differences in second-language learning?A comparison of high and low variability perceptual training for the acquisition of English /i/-/ɪ/ by Spanish adults and children. *New Sounds*. Aarhus, Denmark.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209. https://doi.org/10.7717/peerj.3209

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, *29*(2), 201–218. https://doi.org/10.1177/0265659012467473

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1242–1255. https://doi.org/10.1121/1.408177

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Nishi, K., & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech Language and Hearing Research*, *50*(6), 1496. https://doi.org/10.1044/1092-4388(2007/103)

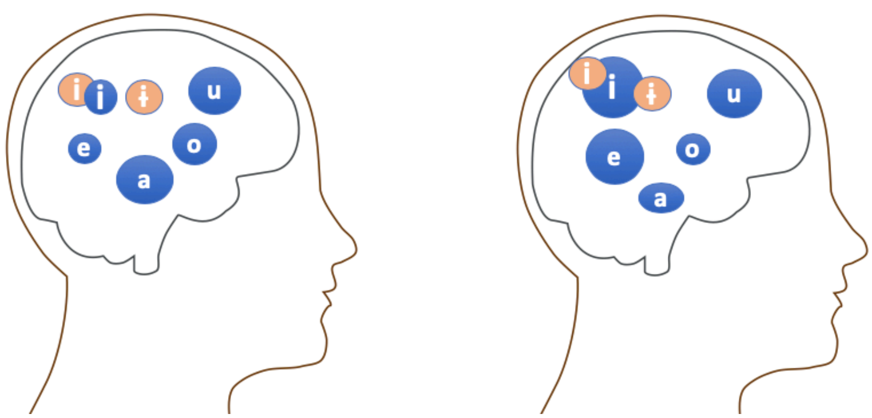
Shinohara, Y., & Iverson, P. (2013). Computer-based English /r/-/l/ perceptual training for Japanese children. *Proceedings of Meetings on Acoustics*, *19*, 060049–060049. https://doi.org/10.1121/1.4800136

The effect of first language perception on the discrimination of nonnative vowel contrasts: Investigating individual differences

Vita Kogan
University of Barcelona, Department of Modern Languages and Literatures and English Studies
vitavkogan@gmail.com

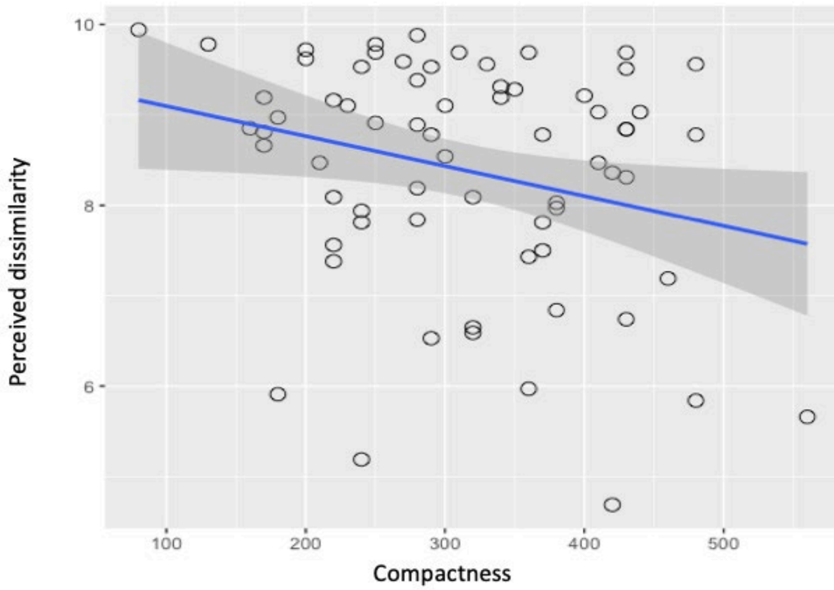
Previous research has shown that native (L1) phonetic categories function as a filter that removes specific acoustic cues for non-native sounds and cause "a perceptual accent" (Best & Taylor, 2007; Flege, 1995). Yet, some individuals are remarkably successful at perceiving non-native L2 sounds (e.g., Bongaerts, Planken, & Schils, 1995). This suggests that perceptual ability is a subject of high inter-learner variability. Whereas L1 background has been the focus of much second language acquisition research to explain such variability, little attention has been paid to the role of individual differences within the same L1 perception. We argue that not only individuals with various L1s are equipped differently for the task of non-native perception, but also individuals with the same L1 vary in how their native categories are represented in perceptual space. Such variability is observable in measures of compactness of L1 phonetic categories (Kartushina & Frauenfelder, 2014), and its effects on non-native perception can be assessed by relating the degree of compactness to the perceived dissimilarity between novel contrasting sounds. We hypothesized that compact L1 categories give an initial advantage in distinguishing non-native contrasts (Figure 1).

Figure 1. *The compactness of L1 vowel categories (in blue) affects the perception of a novel contrast (in pink). The individual on the left is more likely to discriminate between non-native /i/ and /ɨ/ than an individual on the right.*



Sixty-eight Spanish monolinguals participated in the present study. The degree of compactness of their native category /i/ was measured through a goodness-of-fit rating task, where participants listened to synthesized variants of the Spanish /i/ vowel (differing in F1, F2 or both) and rated them as either good or bad exemplars of their internal representation of the category /i/ on an intuitive scale. These ratings provided an individual /i/ compactness index for each participant that was related to the individual perceived dissimilarity score for the novel Russian contrast /i - ɨ/. The results obtained confirmed the hypothesis. Even though L1-based individual differences in perception were small, compactness of the L1 category /i/ contributed significantly to the participants' ability to perceive the acoustic distance between the Russian /i/ and /ɨ/ (Figure 2).

Figure 2. *The relationships between category compactness and perceived dissimilarity between two non-native sounds.*



These findings suggest that having more compact vowel categories might facilitate the process of category formation for unfamiliar sounds. The ability to discern subtle phonetic differences provides an advantage in the acquisition of L2 phonetics and phonology, but not only: Kachlicka, Saito, and Tierney (2019) argue that precise audition can predict L2 lexical and grammatical skills, and might be a gate into language-learning.

References

Best, C. T., & Tyler, M. D. (2007). Non-native and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. The age factor in second language acquisition, 30-50.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore, MA: York Press.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, *192*, 15-24.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246.

The impact of background noise on speech output in people with aphasia

Luisa Zenobi-Bird, Nadia Tetrault, Carolyn Bruce & Caroline Newton†
Department of Language and Cognition, University College London, United Kingdom
†caroline.newton@ucl.ac.uk

Background

Individuals with aphasia report particular difficulties communicating in adverse conditions such as in the presence of background noise. While research has explored comprehension difficulties in noise (Rankin, Parker, Newton & Bruce, 2014), little is known about the effects of background noise on the acoustic-phonetic features of speech output in people with aphasia. Studies on speech production in neurotypical individuals have demonstrated that adverse conditions elicit adaptations known collectively as Lombard speech, including increased intensity, increased pitch and reduced speech rate, which benefit intelligibility (e.g. Cooke, King, Garnier & Aubanel, 2014; Garnier & Henrich, 2014).

The aim of the study was to determine whether people with aphasia make Lombard speech adaptations, and whether type of noise affects features of speech output. Lombard speech requires a range of higher-level cognitive skills as well as lower-level control capabilities (Vogel, Fletcher & Maruff, 2014) and may therefore be particularly challenging for individuals with aphasia who often present with attentional and information processing deficits in addition to language difficulties (Murray, 2012).

Methods & Procedures

Participants were 19 people with aphasia and 19 neurotypical age- and gender-matched adults. The individuals with aphasia were at least 6 months post-stroke, and had a range of types and severity of aphasia (the majority with mild anomic aphasia). Control participants had no known neurological difficulties. Participants were recruited for the study if they reported that they met the following inclusion criteria: (a) have English as their primary language, (b) have no significant hearing difficulties, (c) and no visual impairments that would prevent them from accessing the task stimuli.

The stimulus used for speech production was an A4 Diapix scene (Baker & Hazan, 2011) depicting a beach. To obtain speech produced with communicative purpose, the participants were told that their recording would be played to another individual with a blank scene who would be tasked with drawing the picture based on the description from the participant's recording.

Speech samples were recorded in three listening conditions: quiet, competing speech by a single speaker and cafeteria noise. These maskers were presented binaurally through headphones. The masker conditions were counter-balanced to control for order effects. Fourteen of the participants with aphasia also recorded a picture description in a fourth condition comprising two competing speakers.

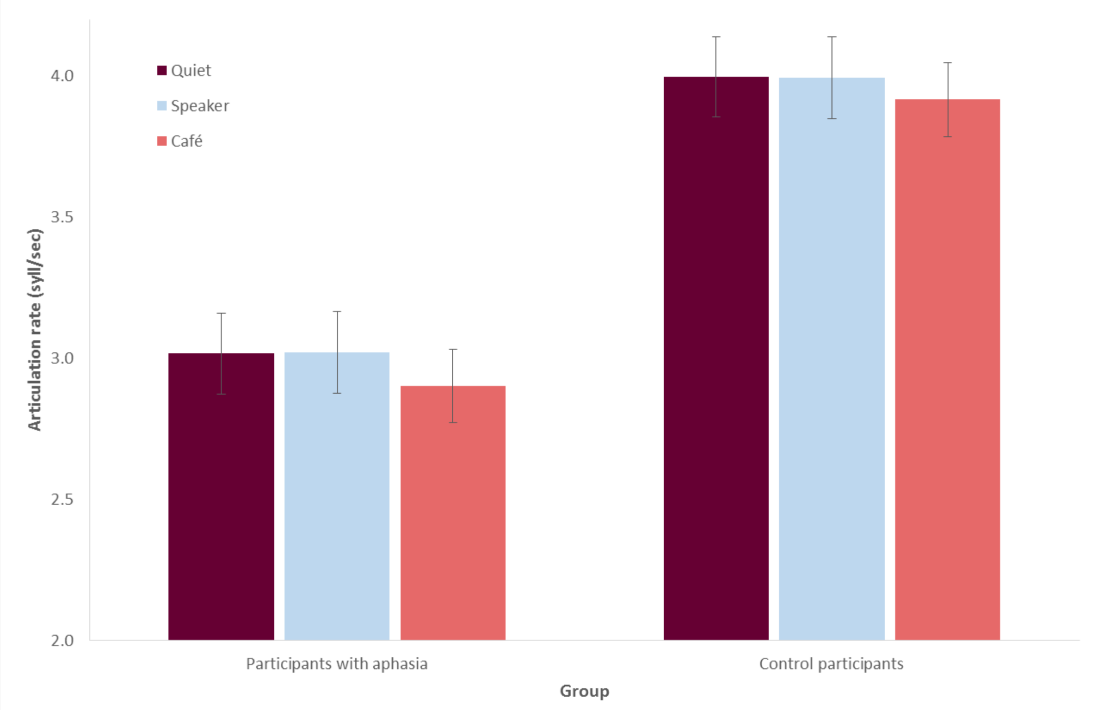
The impact of noise on the participants' speech output was assessed using the following acoustic-phonetic measures: median F0, intensity, articulation rate and pause frequency. Praat scripts were used to extract the median F0 and intensity values following the steps detailed in Hazan et al. (2016). The articulation rate (defined as number of syllables per second) was derived using the qdap package in R. Pause frequency was calculated with Praat as number of pauses over 300ms in duration divided by number of words, making use of the coded silences. These measures were compared across the three conditions and between the two groups using mixed analysis of variance.

Results

Across intensity and F0, there was a significant effect of speaking condition but no effect of group and the group x condition interaction was not significant. People with aphasia therefore exhibited patterns of speech adaptations similar to those displayed by the neurotypical group: elevated intensity and F0 in the two noise conditions with adaptations greater in the cafeteria noise condition.

The aphasic participants had a significantly slower articulation rate than the control participants but, as with the previous two measures, there was no difference between in the groups in terms of the effect of the masker conditions: rate was significantly slower in the context of the café masker than in either of the other conditions (Figure 1).

***Figure 1**. Mean articulation rate by condition and group (error bars show +/- 1 standard error)*



The listening conditions did not have a significant effect on pause frequency in either group. For the 14 participants who also completed the dual speaker condition, there were no differences between this and any other condition for any of the measures.

Despite the café condition eliciting the greatest adaptations to output, the aphasic group reported finding speaking and organizing their thoughts was most difficult in the single competing speaker condition.

Discussion

The results of this study provide evidence that individuals with aphasia are able to make some of the adjustments associated with the Lombard Speech. This suggests that they have access to the higher- and lower-level resources required to make these adaptations. Further research is required to determine the effect of the acoustic-phonetic adjustments on intelligibility and to examine the impact of noisy conditions on language structure. It is possible that by allocating resources to improving intelligibility there is even less processing capacity for ordering ideas and constructing sentences.

Though they were able to make adjustments, the participants with aphasia found the conditions demanding. Future research could also examiner whether they are able to maintain the Lombard speech adjustments for longer periods, and whether resilience to challenging conditions can be increased in this group.

References

Baker, R. & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*(3), 761–770.

Cooke, M., King, S., Garnier, M. & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*(2), 543–571.

Garnier, M. & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, *28*(2), 580–597.

Hazan, V., Tuomainen, O., & Pettinato, M. (2016). Suprasegmental characteristics of spontaneous speech produced in good and challenging communicative conditions by talkers aged 9-14 years. *Journal of Speech, Language & Hearing Research, 59*(6), S1596–S1607.

Murray, L. L. (2012). Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech Language Pathology, 21*(2), S51–S64.

Rankin, E., Newton, C., Parker, A. & Bruce, C. (2014). Hearing loss and auditory processing ability in people with aphasia. *Aphasiology*, *28*(5), 576–595.

Vogel, A. P., Fletcher, J. & Maruff, P. (2014). The impact of task automaticity on speech in noise. *Speech Communication*, *65*, 1–8.

Speech fluency assessments for adults and children: Self-report scales design and development

Roa'a Alsulaiman & Peter Howell
Division of Psychology and Language Sciences, University College London, United Kingdom
p.howell@ucl.ac.uk

Introduction

Fluency problems such as stuttering usually onset between 2 and 7 years with 80-90% of onsets before age 6 (DSM-V; American Psychiatric Association, 2013). Fluency difficulties also have a negative impact on the quality of life and anxiety (Craig, Blumgart & Tran, 2009) and, consequently, such issues are included as criteria in the recent DSM-5 diagnostic criteria for stuttering. DSM-5 also places children on a continuum spanning between extremely fluent and highly disfluent speech. Available instruments for assessing stuttering do not place children on a continuum based on their fluency nor do they include assessment of affectual issues like anxiety. The aim of the present research was to develop assessment protocols that fulfil these requirements. Furthermore, forms meeting DSM-5 criteria appropriate for: 1) adults; and 2) children were created.

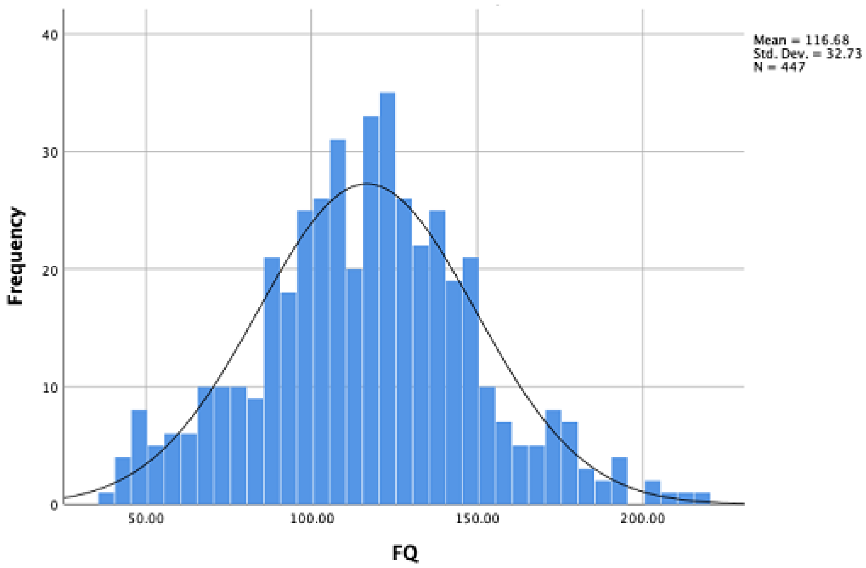
Method

A fluency questionnaire (FQ) was designed that incorporated DSM-V criteria by: 1) placing fluent and dysfluent speakers on a continuum; and 2) including affective components of speech dysfluency such as communication difficulty, quality of life and anxiety. There were 50 questions representing 5 topics, each of which had 10 questions: 1) General Knowledge about Dysfluent Speech, 2) Reactions to Dysfluent Speech, 3) Communication in Social Situations, 4) Quality of Life, and 5) Coping Strategies for Dysfluent Speech. Topic 1 concerned the speaker's perception of fluency and the naturalness of speech. Topics 2 and 3 involved different aspects of affect such as self-confidence as well as the difficulties speakers have when communicating in different settings including at work and in one-to-one conversation. Topic 4 addressed how speech dysfluency interferes with the speaker's satisfaction about their ability to participate in different relationships, their ability to achieve aspirations and their overall wellbeing. Topic 5 addressed coping manoeuvres that dysfluent speakers use to avoid anticipated dysfluency moments such as making non-verbal body movements. Questions were answered by 447 adult participants online using six-point Likert scales. Seventy-four participants were disfluent, and 373 were fluent.

Results

The scores on FQ were approximately normally distributed. This is confirmed by Figure 1 which shows raw scores and a normal distribution fitted to them.

Figure 1: *Histogram with normal distribution curve fitted showing the distribution of raw scores on the FQ, the mean and the standard deviation.*



Independent sample t-tests indicated that there were statistical differences between the FQ scores of fluent and dysfluent speakers; t(445) = 7.805, $p \leq .001$. This confirmed that FQ is appropriate for distinguishing the two fluency groups. Several reliability and validity checks were then performed to assess how effective FQ was in comparison to standard measures of speech fluency and social anxiety. To determine whether FQ questions measured a unidimensional latent construct, Cronbach's α was run on the 447 participants. Cronbach's α coefficient for the FQ was .942, which indicated a high level of internal consistency and a highly reliable scale. To assess criterion validity, raw scores on the Wright and Ayre Stuttering Rating Profile (WASSP; Wright & Ayre, 2000), the Brief Fear of Negative Evaluation Scale (B-FNES; Leary, 1983) and the 'Social Interaction Anxiety Scale' (SIAS; Heimberg, Mueller, Holt, Hope, & Liebowitz, 1992) were compared against FQ. Three separate linear regression models were fit between FQ and one of the three measures. Table 1 shows that FQ scores correlated positively with all scales and therefore was a significant predictor of scores on WASSP, FNES and SIAS. Hence, FQ incorporates assessment of affect (topics 2 and 3), satisfaction ratings and coping mechanisms.

Table 1: *Single linear regression models between the FQ and the other 3 questionnaires used.*

Questionnaire	F	Df	*p*-value	R	R^2
WASSP	380.989	1,445	.000	67.9%	46.1%
SIAS	289.127	1,445	.000	62.8%	39.4%
FNES	164.987	1,445	.000	52.2%	27.1%

Development of child form

A principal component analysis (PCA) was run to check the internal consistency of the questionnaire and to reduce the number of items. The goal was to use the adult form to derive a valid, short and objective self-report test appropriate for use with children that relates to the adult form. Initially, a robust six factor solution emerged where the six factors explained 59% of the variance in the data. Only the first five factors were retained because they each have Eigen value of greater than one. The sixth factor's Eigen value was just over one and only one item loaded strongly on it (>.5) which indicated a weak and unstable factor. The first component accounted for 35.59% of variance within the data and the entire five components accounted for a total of 56.73% of variance in the data. To establish what each of the five components represented and which items measure which factors, the component matrix was examined looking at factor loadings (i.e., Pearson correlations between the items and the components). The criterion for item selection was when the item showed a minimum correlation coefficient of .6 with the component. Following the analysis, it was apparent that 6 items from topic 2 on reactions to dysfluent speech (e.g. "I am afraid to express myself when not speaking fluently") loaded strongly on the first component. The second component was exclusively related to communication in social situations with 4 items loaded strongly on it. The 4 items that loaded strongly on the third component were all related to quality of life when not speaking fluently. The fourth component was exclusively related to the general knowledge of dysfluent speech but only three items loaded strongly. Only two items from quality of life loaded on the fifth component, suggesting that this factor might be redundant. Coping mechanism items did not load onto any of the five components and could be excluded in the child version. Overall, the PCA analyses indicated criterion validity for three subscales; reactions to dysfluent speech, communication in social situations and quality of life. It appears that all three factors relate to the negative consequences of dysfluent speech such as the difficulty to succeed due to not speaking fluently. The fourth factor could probably be strengthened in future versions by revising items with low loadings.

To generate the child form, all 20 questions that resulted from PCA were entered into a logistic regression (LR) model. LR is used to model the probability of correct classification of a binary outcome variable (in this case: fluent or dysfluent). The full model was statistically significant ($\chi 2$(19, N = 447) = 169.630, $p < .001$) indicating that the model was able to distinguish between fluent and dysfluent speakers. The Hosmer and Lemeshow test of goodness of fit also suggested that the model was a good fit to the data (p = .164 >. 05). Results of this model revealed the most significant and nearly significant questions. Those were tested again in a new model that was run with 6 independent variables from the previous model. These questions, which appeared to be the strongest predictors concerned reaction to dysfluency, communication difficulties as well as knowledge of dysfluent speech.

The full model fit to the data was statistically significant ($\chi^2(5, N = 445) = 152.842$, $p <.000$) indicating that the reduced model was able to distinguish between fluent and dysfluent speakers. Wording of those items were changed to make them comprehensible by children. Testing with 30 children is underway and the relationship between their fluency level as assessed by Riley's (1994) Stuttering Severity Instrument and the child form of FQ will be reported at the conference.

Discussion

This work presented a new instrument for assessing speech fluency and examined its psychometric properties. A second aim concerned developing a child version of FQ by assessing and keeping useful items to massively reduce the length of the questionnaire making it suitable for testing children who have limited understanding and attention. This would serve our wider goal of developing instruments to assess speech dysfluency in preschool aged children. FQ was administered on a large sample of adult fluent and dysfluent speakers and appeared to be a specific indicator of speech dysfluency. FQ underwent rigorous statistical analysis and in all instances the results went in the predicted direction; supporting high reliability, high internal consistency, and adequate validity level. FQ is still under development and in need of further validation. In ongoing work, there is an Arabic version of FQ has been adapted culturally. It is hoped that use of all forms of FQ will be useful for clinicians and researchers to facilitate their s assessment of speech fluency by providing them with a short and easy-to-use instrument that meets DSM-5 criteria.

Acknowledgment

This research is supported by King Saud University (whose support is gratefully acknowledged)

References

Association, A. P., & others. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of fluency disorders*, *34*(2), 61-71.

Heimberg, R. G., Mueller, G. P., Holt, C. S., Hope, D. A., & Liebowitz, M. R. (1992). Assessment of anxiety in social interaction and being observed by others: The Social Interaction Anxiety Scale and the Social Phobia Scale. *Behavior therapy*, *23*(1), 53-73.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*(3), 371-375.

Riley, G. (1994). Stuttering severity instrument for young children (SSI-3). *Stuttering Severity Instrument for Young Children (SSI-3)*.

Wright, L., & Ayre, A. (2000). *WASSP: Wright & Ayre Stuttering Self-Rating Profile*. Winslow.

Disfluency patterns in speech of children with and without autism spectrum disorder

Charlotte Bellinghausen[1]*, Mária Gósy[2]*, Reinhold Rauh[3*], Bernhard Schröder[1], Thomas Fangmeier[4], Ludger Tebartz van Elst[4], Andreas Riedel[4]
[1]Institute of German Studies, University of Duisburg-Essen [2]Research Institute for Linguistics, Budapest
[3]Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg [4]Department of Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg
*charlotte.bellinghausen@uni-due.de, *gosy.maria@nytud.hu, *reinhold.rauh@uniklinik-freiburg.de

The goal of the current paper is to give an overview of studies investigating the role of disfluencies in speech development with focus on children with autism spectrum disorder (ASD). In this context, we focus on the production of disfluencies. We aim at comparing different studies from a multi-disciplinary perspective, i.e. from linguistics, medicine, and psychology. Our goal is to explore the nature of the occurrence of disfluencies in children's speech with and without ASD by discussing frequency and types of disfluency.

Autism spectrum disorder (ASD) is according to the diagnostic criteria of DSM-5 (American Psychiatric Association, 2013) and ICD-11 (World Health Organization, 2018) a neurodevelopmental disorder with severe impairments in the two domains of social communication and restrictive repetitive behaviors/interests. It can be found "… in at least 1% of the child population" (NICE, 2011: 4) and there is a difference in occurrence with respect to gender (3,5:1; boys: girls) (cf. Biscaldi et al., 2012: 500 referring to Fombonne (2005)). With respect to speech processing it can be said that processing of syntax and semantics are intact in high-functioning ASD, i.e. in ASD without intellectual impairment, whereas the main difficulties are found in pragmatics. Riedel et al. (2014) presented the Freiburg Questionnaire of linguistic pragmatics which is used for diagnosis, i.e. for the understanding of metaphors.

As Scaler Scott (2015: 240) points out, in addition to problems with language and communication, other speech issues have also been observed in ASD, e.g. difficulties with articulation, voice and fluency. Our aim is to give an overview of studies investigating disfluencies in the speech production of children with high-functioning ASD. The term of disfluency refers to various phenomena that are inserted into spontaneous utterances with diverse functions such as revealing speech planning, execution difficulties, or signaling discourse interactions.

Disfluencies in speech development – typically developing children

Soderstorm and Morgan (2005) empirically investigated in a preference test whether children by age of 10 months differentiate between fluent and disfluent speech samples using the headturn preference procedure. The study showed that for the 10-month olds mean looking time did not differ when listening to fluent vs. disfluent speech. Results for 22-month olds suggested a tendency for longer looking time for the fluent speech. Kidd et al. (2011) conducted an eyetracking study with children in the ages of 1;8-2;8 in order to see whether young children used disfluencies for the prediction of speaker's intended referents. Results showed that children by the age of 2;4-2;8 reliably attended to disfluencies of speech and used this information to infer speaker intention even before the object was labelled. Furthermore, there is evidence that this ability manifests in the age of 2.

Hudson Kam and Edwards (2008) tested the use of *uh* and *um* for children aged 3;4-4;11. Children's speech was elicited by means of asking help from the children for story-telling of the experimenter. The scenario was to describe stories in books. The assumption was that a pause succeeding a delay marker is longer than a pure pause. Results did not show significant difference in length. The fluency of typically developing children's utterances is influenced by various factors depending also on age (e.g. for Turkish see Furman, Özyürek, 2007). The most common disfluencies during language acquisition are silent pauses, filled pauses and repetitions from the age of about 2 years (Ambrose & Yairi, 1999). Studies show that children encounter more disfluencies as they get older because of their relatively long and grammatically more complicated utterances (e.g., Neuberger, Gósy, 2014). Filled pause *uh* appears earlier in children's speech than *um* (Kestenbaum, Williams, 1992). In some languages, the majority of the filled pauses contains a neutral vowel as opposed to a nasal also in children's speech without any functional difference (e.g. for Hungarian see Horváth, 2014). In a case study with a Dutch child investigated between 1;10 and 2;7 articulation reasons for *um*-like hesitations were found (Taelman–Durieux, 2009). In the study of Hudson Kam and Edwards (2008) children between 3-4 years used longer pauses following *uh* and *um* as opposed to pauses not following *uh* and *um*, but there was no difference between the pauses following *uh* and *um*.

Disfluencies in speech development – children with ASD

MacFarlane et al. (2017) collected speech data of 4-8 year old children in three groups: i) children with ASD, ii) children with specific language impairment, iii) a control group. The authors showed that children with ASD produce a higher ratio of content to filler disfluencies than typically developed children. In contrast, there was no significant difference with respect to relative frequencies of repetitions, revisions, and false starts. Scaler Scott et al. (2014) investigated the nature of disfluencies for children with average age of 11 years. Speech samples were elicited during an expository discourse task also for three groups of children: i) children with high-functioning ASD, ii) stuttering children and iii) a control group. Results showed that disfluencies produced by the children with ASD differed qualitatively and quantitatively from disfluencies produced by the stutterers and also were characterized by a higher frequency of word-final disfluencies.

In other studies it was tested whether children with high-functioning ASD use the filled pauses *um* and *uh* differently than typically developing children. Lunsford et al. (2010) tested children from 4-8 years. They were recorded while having a conversation, describing a picture or book and playing. Results showed that the children with ASD use *um* less often than the typically developed children. Also, fewer pauses were produced after the *um* in the group with ASD and pauses were shorter after *um.* The subjects in the study by Irivine et al. (2016) were aged between 8-21 years and were instructed to describe paintings while simultaneously tapping a keyboard. Results suggested significant differences for the *um* rate, but not for the *uh* rate. It was concluded that *um* may rather than *uh* serve as a listener-oriented signal due to the problems in linguistic pragmatics in ASD. Gorman et al. (2016) conducted a similar study by testing younger participants aged between 4-8 years. Speech data were collected during play, picture description, story-telling, and conversation. Again, the autistic children used *um* less often than the control group did. McGregor and Hadden (2018) showed also for 7-15 year olds with ASD that *um* was used less often in the ASD group than in the control group.

In our current project we test whether disfluencies functioning as prosodic indicators of uncertainty are perceived by adult listeners with and without ASD. In Bellinghausen et al. (2019) different degrees of intended uncertainty were generated by different combinations of *rising intonation*, *fillers*, and *pauses* using the articulatory speech synthesizer VocalTractLab (2017). Neurotypical adult listeners were asked to judge the degree of uncertainty of the utterance. Results suggest evidence for a superadditive principle of the cues: the more uncertainty cues are activated, the higher the perceived level of uncertainty. In current work we are presenting the material to adult listeners with and without ASD to test if there are differences in perception. In our future work we are also planning to present the material to children with and without ASD.

Conclusion

The studies presented here suggest that there are differences in disfluency production comparing children with ASD to neurotypical children. The studies reported here provide evidence for the age of >=4 years that there are in particular differences between the use of *uh* vs. *um* as filled pauses. The finding of the functional difference in use between *um* and *uh* in children with ASD which is not characteristic of typically developing children might be interpreted as a candidate of a linguistic marker for diagnosis of certain subtypes of ASD.

References

Ambrose, N. G., Yairi, E. (1999). Normative disfluency data for early childhood stuttering. In: *Journal of Speech, Language, and Hearing Research,* 42, 895–909.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edition: DSM-5. Washington, DC: American Psychiatric Publishing.

Bellinghausen, C., Fangmeier, T., Schröder, B., Keller, J., Drechsel, S., Birkholz, P., Tebartz van Elst, L., Riedel, A. (2019). On the role of disfluent speech for uncertainty in articulatory speech synthesis. In: *Proceedings of DiSS 2019: The 9th Workshop on Disfluency in Spontaneous Speech*, *DiSS 2019*, Budapest, Hungary, 39-42.

Biscaldi, M., Rauh, R., Tebartz van Elst, L., Riedel, A. (2012). Autismus-Spektrum-Störungen vom Kindes- bis ins Erwachsenenalter. Klinische Aspekte, Differenzialdiagnose und Therapie. In: *Nervenheilkunde*, 31(7-8), 498-507.

Dilling, H., Mombour, W., Schmidt, M.H. (2015). (eds.). *Internationale Klassifikation psychischer Störungen: ICD-10 Kapitel V (F) - Klinisch-diagnostische Leitlinien* (10. Auflage). Bern: Hans Huber.

Fombonne, E. (2005). The changing epidemology of autism. In: *J Appl Res Intellect*, 18(4): 281-94.

Furman, R., Özyürek. A. (2007). Development of interactional discourse markers: Insights from Turkish children's and adults' oral narratives. In*: Journal of* Pragmatics, 39(10), 1742–1757.

Gorman K., Olson L., Hill A.P., Lunsford R., Heeman P.A., van Santen J.P. (2016). Uh and um in children with autism spectrum disorders or language impairment. In: *Autism Research*, 9(8), 854–865, doi: 10.1002/aur.157.

Horváth,V. 2014. *Hezitációs jelenségek a magyar beszédben*. / Hesitatiion phenomena in Hungarian speech/ Budapest, ELTE Eötvös Kiadó.

Hudson Kam, C.L., Edwards, N.A. (2008). The use of *uh* and *um* by 3- and 4-year-old native English-speaking children: Not quite right but not completely wrong. In: *First Language* 28, 313–327.

Irvine C.A., Eigsti I.M., Fein D.A. (2016). Uh, um, and autism: Filler disfluencies as pragmatic markers in adolescents with optimal outcomes from autism spectrum disorder. In: *Journal of Autism and Developmental Disorders*, 46(3), 1061–1070, doi: 10.1007/s10803-015-2651-y.

Kestenbaum, C., Williams, D. T. (1992) (eds.). Handbook of clinical assesment of children and adolescent. NYU Press: New York.

Kidd, C., White, K. S., Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. In: *Developmental Science*, 14, 925-934.

Lunsford, R., Heeman, P.A., Black, L., Santen, J. van (2010). Autism and the use of fillers: differences between 'um' and 'uh. In: *Proceedings of the 5th Workshop on Disfluency in Spontaneous Speech*, Tokyo, 107-110.

MacLachlan, B. G., Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. In: *Journal of Speech and Hearing Disorders,* 53, 2–7.

Mac Farlane, H., Gorman, K., Ingham, R., ...van Santen, J. (2017). Quantitative analysis of disfluency in children with autism spectrum disorder or language impairment. In: *PloS One*, 12, 3, e0173936.

McGregor, K., Hadden, R. (2018). Brief Report: ‚Um' Fillers Distinguish Children With and Without ASD. In: *Journal of Autism and Developmental Disorders*. DOI: 10.1007/s10803-018-3736-1.

Neuberger, T., Gósy, M. (2014). A cross-sectional study of disfluency characteristics in children's spontaneous speech. In: *Govor*, 31(1), 3-28.

NICE (2011). *Autism diagnosis in children and young people. Recognition, referral and diagnosis of children and young people on the autism spectrum*. Clinical Guideline no. 128. London: National Institute for Health and Clinical Excellence.

Riedel, A., Suh, H., Haser. V., Hermann, I., Ebert, D., Riemann, D., Bubl, E., van Elst, L.T., Hölzel, L.P. (2014): Freiburg Questionnaire of linguistic pragmatice (FQLP): psychometric properties on a psychiatric sample. In: *BMC Psychiatry*, 14: 374.

Scaler Scott, K.S. (2015). Disfluency in Autism Spectrum Disorder. In: *Procedia – Social and Behavioural Sciences*, 193, 239-245.

Scaler Scott, K.S. Tetnowski, J. A., Flaitz, J., Yaruss, J. S. (2014). Preliminary study of dysfluency in school-age children with Autism. In: *International Journal of Language and Communication Disorders*, 49(1), 75-89.

Soderstrom, M., Morgan, J. L. (2005). Disfluency in speech input to infants? The interaction of mother and child to create error - free speech input for language acquisition. In: *Procedings of the Disfluency in Spontaneous Speech Workshop*, Aix -en-Provence, France, 157-162.

Taelman, H., Durieux, G., Gillis, S. 2009. Fillers as signs of distributional learning. In:. *J. of Child Language*, 36/2, 323-335.

Vocal Tract Lab (2017), www.vocaltractlab.de [last retrieved: 20th of February 2020].

World Health Organization. (2018*). International classification of diseases for mortality and morbidity statistics* (11th Revision),

https://icd.who.int/browse11/l-m/en [last retrieved: 20th of February 2020].

Yaruss, J. S., Newman, R. M., Flora, T. (1999). Language and disfluency in nonstuttering children's conversational speech. In: *Journal of Fluency Disorders* 24, 3, 185–207.

Theory of mind related changes in speech production in preclinical stages of Alzheimer's Dementia

Dagmar Bittner[1] & Johannes Schröder[2]
[1]Leibniz-Centre for General Linguistics Berlin, [2]University of Heidelberg, Germany,
bittner@leibniz-zas.de

The abilities to infer other peoples' mental states, thoughts, expectations and feelings belongs to the key aspects of an individual's social cognition. In psychological and linguistic research these abilities are sampled under the term ‚theory of mind' (ToM). While research on ToM-abilities as well as on working memory and executive functions concerns all ages in psychology, linguistic research focussed ToM-related language nearly exclusively in specific contexts as there are child development and language use in autistic and aphasic speakers. Findings on changes in ToM-abilities in the later ages (see Castelli, Pini, Alberoni, Liverta-Sempio, Baglio, Massaro, Marchetti & Nemni, 2011) suggest, however, to take ToM into account when investigating changes in language use and language knowledge over the life span. This specifically holds for research on pathological changes in language use due to cognitive diseases like dementia. So far, changes in language use in older adults-healthy aging or suffering from cognitive diseases-are typically seen as related to losses in working memory and attention span. Without denying the impact of these factors we will show, however, that there are preclinical changes in people developing Alzheimer's Dementia (AD) that cannot be explained by mere losses in working memory and attention span. We will hypothesize instead that the pattern of changes in the use of pronouns, modal verbs, and connectors we are reporting on is related to early, typically unnoticed changes in ToM-abilities.

The focus of our (ongoing) study is on pronoun use in preclinical and initial stages of AD. Several studies reported an increase in pronoun use in mild and moderate stages of AD (e.g., Almor, Kempler, MacDonald, Andersen & Tyler, 1999; Bucks, Singh, Cuerden & Wilcock, 2000 on English; and Tönjes, 2012; Wendelstein 2016 on German). Recent studies suggest that the increase is present already in the preclinical stage of AD (e.g., Jarrold, Peintner, Wilkins, Vergryi, Richey, Gorno-Tempini & Ogar, 2014; Frasera, Meltzerband & Rudzicz, 2016 on English; Schecker 2010; Wendelstein 2016 on German). However, the types of pronouns investigated vary from study to study. It even varies which elements of single pronoun types were considered. Concerning personal pronouns for example, studies on English spontaneous speech investigate the complete range of personal pronouns. Same studies on German vary with respect to the inclusion of 1st and 2nd person pronouns (*I, we – you*). Gress-Heister (2003) and Wendelstein (2016) included it, Tönjes (2012) excluded it. They are also typically excluded from studies based on experimental data and data elicited by story completion or picture description. 1st and 2nd person pronouns are qualitatively different from 3rd person pronouns. They are exclusively deictic. Changes in the use of these pronouns cannot be related to changes in the accessibility or activation of an underlying concept or the appropriate lexem. Studies on German further vary with respect to inclusion of the so-called D-pronouns (*der, die, den* etc., see below). Gress-Heister (2003) and Schecker (2010) excluded this type, Tönjes and Wendelstein included it but in different ways. Finally, most studies include other types as e.g. possessive pronouns (*my, your, his, her*) or interrogative pronouns (*where, what, who, whom*) causing further variation in the investigated samples of pronouns. Explanations of the observed increases in pronoun use under AD also differ. Almor et al. (1999) reject the *semantic knowledge impairment hypothesis* (see Hier, Hagenlocker, & Shindler, 1985) and propose that losses in working memory lead to increased use of semantically empty words including pronouns. Wendelstein (2016), who found no differences between ADs and CTRs in lexical parameters, suggests early deficits in pragmatic constraints on pronoun use.

Given this state of the art we investigated pronoun use in spontaneous speech production of 10 participants of the German ILSE-cohort born in 1930-32. All 10 subjects took part in the first 3 intervention waves of the ILSE-study. Each intervention wave included a semi-structured biographical interview and a battery of medical and social screenings. All 10 subjects were cognitively healthy at the first intervention wave (T1), that is at age 62 to 64. At the third intervention wave (T3) at age 73 to 75, 5 of the 10 subjects were diagnosed with mild AD (AD-group) while the other 5 subjects stayed cognitively healthy (CTR-group). Each subject of the AD-group was matched with one subject of the CTR-group for age, gender, years of education, and place of living (participants of the ILSE-study are recruited in two towns in Germany: Leipzig and Heidelberg). Analyses were performed on the language data from the biographical interviews conducted at T1 (preclinical stage of AD) and T3 (mild stage of AD). We selected the AD's first 300 3rd person pronouns (irrespective of pronoun type) from each interview. Analyses are, thus, based on 1500 data points for each group and point in time (AD-T1, AD-T3, CTR-T1, CTR-T3 = 6000 data points in total). Data are coded for type of pronoun (12 types), number (singular/plural), type of clause (main/sub), position of pronoun (preverbal/postverbal), referent of pronoun (family member, other human, inanimate), type of finite verb in the clause containing the pronoun (lexical, modal, copula, auxiliary).

Group comparison revealed a significantly higher proportion of *D-pronoun* (*der, die, das, dem, den, des*) at T1 and T3 in the AD-group than in the CTR-group. Proportions of all other pronoun types do not significantly differ except of *propositional das* and *propositional es*, that is of *das/es* pronouns referring to propositions (*das/es war schön* 'this was beautiful'). Proportion of these two types is significantly lower at both T1 and T3 in the ADs than in the CTRs. This difference increases from T1 to T3. There is no group difference in the distribution of D-pronoun over syntactic contexts (pre- vs. postverbal position; main- vs. sub-clause) but in the preferred types of referential relations of *D-pronoun* and *personal pronoun*. CTRs prefer *personal pronoun* over *D-pronoun* in references to family members. ADs, on the contrary, prefer *D-pronoun* over *personal pronoun* in reference to family members. In addition, ADs show an increased use of *personal pronoun* in reference to other humans compared to CTRs. Analyses of verb use in clauses containing *propositional das* revealed a lower proportion of copula-constructions in ADs than CTRs. Taking clauses containing *D-pronoun* and *propositional das* together we noticed a mere lack of conjunctive verb forms, epistemic modals and some connectives as e.g. *ob* 'if' in the AD's data. Closer analyses of these latter findings are currently running.

Summing up, we identified pathological changes in the use of two pronoun types, *D-pronoun* and *propositional das*, in a very preclinical stadium of AD. The changes do not affect information-structural constraints on the use of *D-pronoun* - tested here for postverbal position and occurrence in sub-clauses. Also, structures of clauses containing *propositional das* are intact. The observed changes rather reduce i) information on the emotional state of the speaker towards people and events referred to and ii) information the hearer would need to establish more detailed representation of referents and events. The former is evidenced by the replacement of *personal pronoun* by *D-pronoun* in references to family members and the lower proportion of clauses like *das ist schön* 'it is beautiful', *das war eine schwere entscheidung* 'it was a difficult decision' etc. The latter is evidenced by the lower proportion of clauses like *das ist ein dorf bei Mannheim* 'it is a village near Mannheim'. Finally, the mere lack of conjunctive forms, epistemic modals, and connectors like *ob* 'if' in the data of the AD's results in reduced specification of factivity/non-factivity and external sources of information. In sum our findings suggest that language use in the preclinical stage of Alzheimer's Dementia undergoes pathological changes due to changes in ToM-related abilities.

References

Almor, A., Kempler, D., MacDonald, M. C., Andersen E. S., & Tyler L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language* 67, 202–227.

Bucks R.S., Singh, S., Cuerden, J. M., & Wilcock, G. C. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance. *Aphasiology* 14, 71-91.

Castelli, I., Pini, A., Alberoni, M., Liverta-Sempio, O., Baglio, F., Massaro, D., Marchetti, A., & Nemni, R. (2011). Mapping levels of theory of mind in Alzheimer's disease: A preliminary study, Aging & Mental Health, 15(2), 157-168.

Frasera, K. C., Meltzerband, J. A., & Rudzicz, F. (2016). Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease* 49, 407–422.

Gress-Heister, M. (2003). Sprachverarbeitung bei Altersdemenz am Beispiel pronominaler Formen. In: Fiehler R; Thimm C. *Sprache und Kommunikation im Alter*. Wiesbaden: Westdeutscher Verlag.

Hier, D. B., Hagenlocker, K., & Shindler, A. G. (1985). Language disintegration in dementia: Effects of etiology and severity. Brain and Language, 25(1), 117–133.

Jarrold, W., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M. L., & Ogar, J. (2014). Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology*, pp. 27-36.

Tönjes, M. (2012). *„So jetzt geben wir das da rein." Formen der Referenz bei Demenzpatienten*. Tübingen: Stauffenburg. (Stauffenburg Linguistik 68).

Wendelstein, B. (2016). *Gesprochene Sprache im Vorfeld der Alzheimer-Demenz: Linguistische Analysen im Verlauf von präklinischen Stadien bis zur leichten Demenz*. Heidelberg: Universitätsverlag Winter.

Phonological and semantic control in bilingual word production across the adult lifespan

Marco Calabria
Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
mcalabria@uoc.cat

Introduction

The studies that have investigated the bilingual language production across the adult lifespan have found mixed results for the effect of aging on lexical retrieval. For instance, no effect of aging was found on naming latencies when three age groups of bilinguals (young, middle-aged and elderly) were compared in language switching tasks (Calabria et al., 2015), but it was for semantic fluency tasks (Ivanova et al., 2016). This may suggest that the effect of aging on lexical retrieval may be partially modulated by the type of task and potentially the level of processing involved.

To further explore this hypothesis, the present study investigated the aging effects on the language control components of lexical retrieval when modulated by semantic (Experiment 1) and phonological (Experiment 2) contexts. To do so, we used cycling naming tasks as a measurement of the competition processes during lexical selection. Naming latencies were compared between young and older adults for the two languages.

Methods

Two age groups of Catalan-Spanish bilinguals took part in Experiment 1 (20 young adults, age range= 21-26; 20 older adults, age range = 52-57) and Experiment 2 (20 young adults, age range= 21-27; 20 older adults, age range = 51-56). Language history and dominance were determined by means of a questionnaire administered to the participants and an interview with them. Participants were considered early bilinguals as, on average, they were first regularly exposed to their non-dominant language at 6 years of age. Moreover, they reported equal amounts of Catalan and Spanish usage and thus would be considered balanced bilinguals.

In Experiment 1, stimuli consisted of two sets of 16 pictures (one set for each language) with 8 exemplars from 4 semantic categories (animals, vegetables, kitchen tools, and furniture). In Experiment 2, stimuli consisted of two sets of 16 different pictures (one set for each language) with 8 exemplars of 4 groups, each that sharing a phonological overlap for the initial segment of the word (first phoneme of first two phonemes, Spanish: me-, co-, ha-, ra-; Catalan: po-, ga-, ca-, na-). In both experiments participants were required to name 8 blocks of pictures: 4 blocks containing semantically or phonologically related items and 4 blocks containing unrelated items.

The experimental software used for the tasks was DMDX and performances (naming latencies and accuracy) were analysed off-line with Checkvocal. Each trial included the following elements: a fixation point presented for 750 ms followed by the picture to be named, which appeared for up to 2000 ms or until a response was provided.

Results

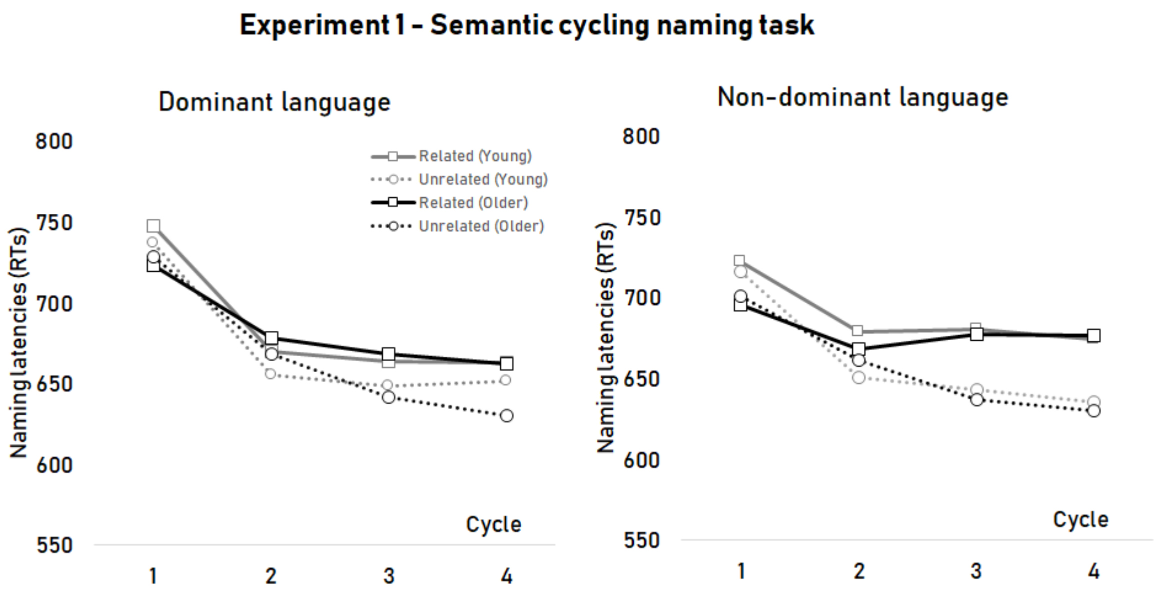
In Experiment 1, the results showed a significant effect of relatedness ($F(1, 38) = 16.68$; $p < .01$), indicating that in the semantically related condition participants were slower than in the unrelated condition ($F(3, 114)= 6.53$; $p < .05$). Moreover, the difference in naming latencies increased across repetitions especially in older adults. This result was explained by an advantage of older (38 ms) over young adults (5 ms) in the last three repetitions for the unrelated items. No differential language effects were found in any age group.

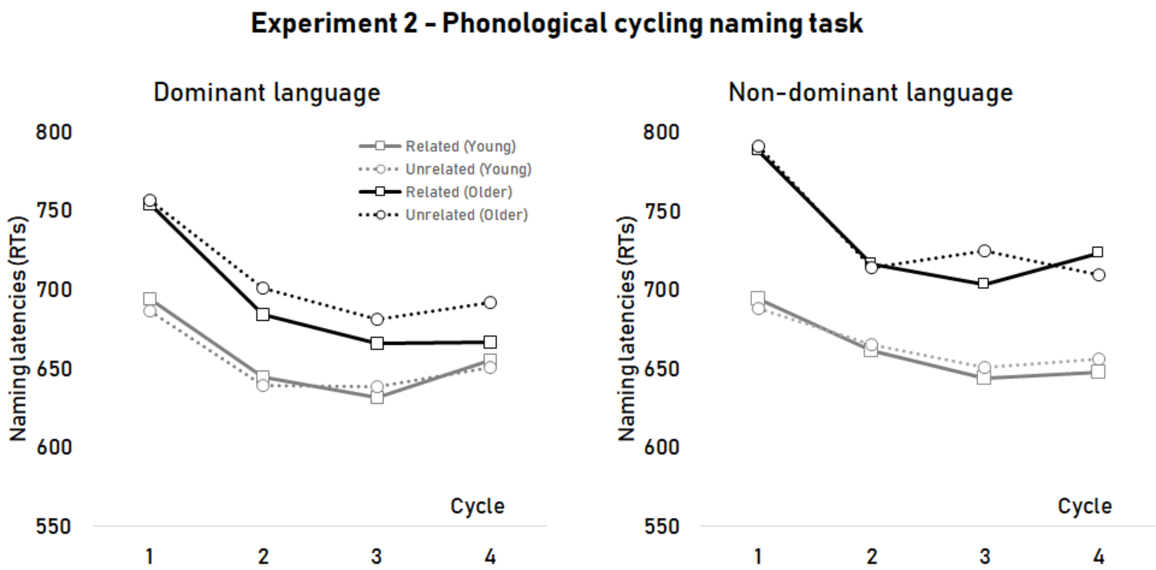
In Experiment 2, it was found a significant effect of repetitions ($F(3, 114) = 45.12$, $p < .01$), suggesting that participants get faster from the first to the fourth presentation of the picture. Moreover, the repetition effect was larger in older than young adults (Group x Repetition interaction: F (3, 114) = 8.47, *p*< .01). Finally, older adults were overall slower in naming compared to young adults F (1, 28) = 6.05, *p*< .05). However, no differential language effects were found in any age group.

In a further analysis, RT distributions were fitted to an *ex-Gaussian distribution*, which leads to an estimation of both the normal and the exponential distribution. The use of the ex-Gaussian analysis is that it has been previously applied successfully to understand differences between groups of participants (e.g., Spieler et al., 1996; Tse et al., 2010). The estimation gives a value for two parameters of interest, namely mu (μ) and tau (τ), and for each participant per condition (related vs. unrelated) and per group.

In Experiment 1, the results for the estimated values of mu it was found an effect of relatedness (F (1, 39) = 6.32, *p* < .05), but no significant effects for tau values. In Experiment 2, the results for the estimated values of mu did not show significant effects. However, for tau values it was found a significant interaction between Group and Relatedness (F (1, 26) = 6.10, *p*< .05), suggesting a facilitation in naming for phonologically related items for young adults, but an interference for older adults.

Figure 1: *Naming latencies (ms) of the cycling naming tasks as a function of languages and groups of participants*





Discussion and conclusions

We found three main results that are relevant in the context of bilingualism and lexical retrieval across lifespan. First, phonological and semantic control processes related to lexical selection are language-independent, as measured by similar effects across languages for the two naming tasks. Second, lexical retrieval is sensitive to aging effects when modulated by phonology, but not by the semantic context. Third, ex-Gaussian analyses showed that there was age group difference for those parameters that measure control processes but only for phonology. This may suggest that the engagement of language control in lexical selection and retrieval depends on level of linguistic processing involved.

References

Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism*, *18*(1), 67-78.

Ivanova, I., Murillo, M., Montoya, R. I., & Gollan, T. H. (2016). Does bilingual language control decline in older age? *Linguistic Approaches to Bilingualism*, *6*(1–2), 86–118.

Spieler, D. H., Balota, D. A., and Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 461–479.

Tse, C. S., Balota, D. A., Yap, M. J., Duchek, J. M., and McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology,* 24, 300–315.

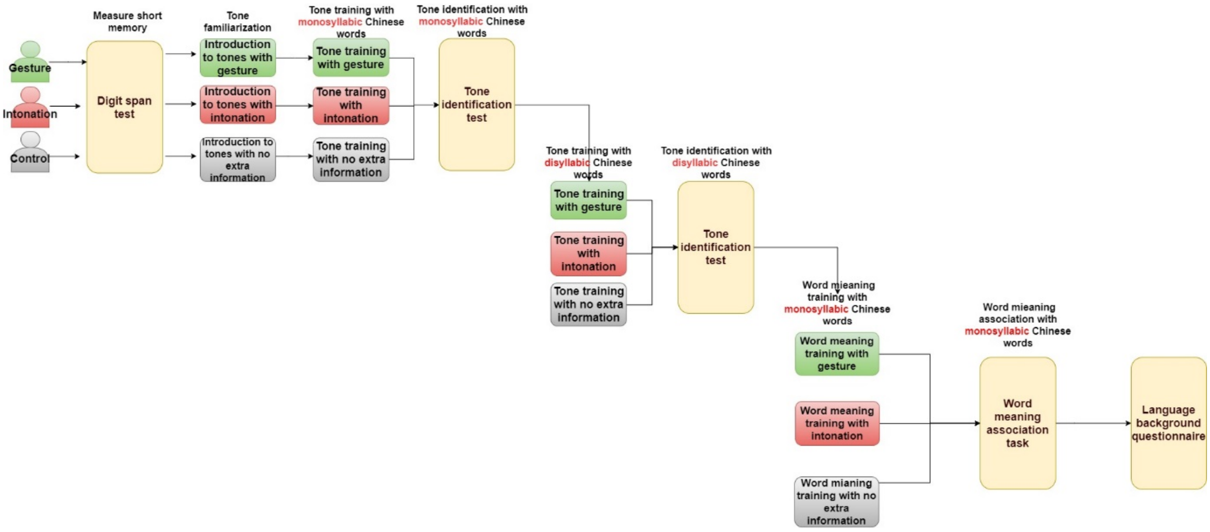
Teaching Mandarin tones to native English speakers: Tone-mimicking hand gestures vs assimilation to English intonational categories

Siyu Chen[1], Laurence White[2], Maria Arche[1] & Claire Monks[3]
[1]School of Humanities and Social Sciences, University of Greenwich, UK
[2]School of Education, Communication and Language Sciences, Newcastle University, UK
[3]School of Human Sciences, University of Greenwich, UK
s.chen@greenwich.ac.uk

We compared the effect of two training methods on English speakers' learning of Mandarin lexical tones: 1) observation of hand gestures that mimic tone contours; 2) assimilation of Mandarin tones to English intonational categories. As suggested by McNeill (2008), hand gestures may help second language (L2) learners visualize abstract L2 features and build up a physical sense of those features. More specifically, tone-mimicking gestures are held to be effective in enabling English-speaking learners of L2 Mandarin to discriminate between words on the basis of the four principal Mandarin lexical tones (Morett & Chang, 2015). This cross-modal support may be limited, however, as L2 learners still need to interpret the visual information into an appropriate auditory format. Moreover, a possible auditory-only method of lexical tone training for native speakers of non-tonal languages is suggested by consideration of the Perceptual Assimilation Model for L2 learners (PAM-L2; Best & Tyler, 2007). This model proposes that L2 learners assimilate L2 sounds to existing L1 categories at points of relative similarity in the L1 and L2 phonological inventories (Best & Tyler, 2007). Thus, a within-modality training method that encourages assimilation between novel L2 tone features and existing L1 intonational features might be more beneficial than a method requiring cross-modal mapping.

In a between-subjects design, 75 native British English speakers with no experience of any tonal language participated the study. Participants were randomly assigned to one of the three training conditions: gesture observation (GO), tone assimilation (TA), baseline control (BC) and the experiment was performed in the following sequence (see also Figure 1).

Figure 1: *Schematic representation of the experiment procedure in the three tone training conditions*



1. Participants completed a digit span test to assess their short-term memory span and then watched a video to introduce the four Mandarin tones using the (pinyin) syllable "a".

2. Participants watched a tone training video of 36 monosyllabic Chinese words, according to their assigned training condition (see below) and took a tone identification test with 24 monosyllabic Chinese words (12 pre-trained and 12 new).
3. Participants watched another tone training video of 24 disyllabic Chinese words, according to the training condition, and then took a tone identification test with disyllabic Chinese words (12 pre-trained and 12 new).
4. Participants watched a word-meaning training video with 12 monosyllabic words that were tested in the tone identification test and then performed a word meaning identification test with the 12 pre-trained words.
5. Participants completed a language background questionnaire.

Training videos in the three conditions were designed to be as similar as possible except for the following key distinctions. In the gesture observation (GO) condition, the instructor in the videos presented the Mandarin stimuli using her hand to mimic the corresponding lexical tone pitch contour (Figure 2). In the tone assimilation (TA) condition, Mandarin tones 1, 2, 3 and 4 were explicitly compared to English hesitation, question, uncertainty, and statement intonation contours respectively, with each tone and its corresponding English intonation contour demonstrated before the presentation of the training videos. In the baseline control (BC) condition, the videos were presented in a similar way as in the GO and TA conditions, except that the instructor did not perform pitch-mimicking hand gestures and the English intonation comparisons were not discussed. All groups received passive training, i.e., they listened and watched videos without imitating the tones.

Figure 2: *Still illustrations from the tone training video in the GO condition*

A one-way by-subjects ANOVA shows no differences in digit span results between groups, $F(2, 74)=0.903$, $p=.405$; thus short-term memory capacity was comparable across conditions. The accuracy rate models were constructed with glmer in the R package lmerTest (Kuznetsova et al., 2015). Participants and items were treated as random effects on the intercept and slope. Model comparisons show improvements in model fit after adding fixed effects of condition: monosyllabic tone identification: $\chi2(2) =8.481$, $p<.01$; disyllabic tone identification: $\chi2(2) =6.14$, $p<.05$.

Planned pairwise comparisons show that performance accuracy on the tone identification test for monosyllabic Chinese words was higher for both the GO condition (70%, $p<.05$) and the TA condition (74%, $p<.01$) than for the baseline (59%). However, higher accuracy on the tone identification test for disyllabic Chinese words was only observed, relative to baseline (44%), with the TA group (63%, $p<.05$). Furthermore, the TA group's disyllabic performance was close to being significantly better than the GO group (47%, $p=.056$). A similar pattern of numerical variation in accuracy rates was found in the monosyllabic word meaning identification task (GO: 67%, TA: 69%, BC: 65%), but differences were not statistically robust.

We conclude that both gesture observation and tone assimilation are helpful for English-speaking learners of Mandarin to learn L2 lexical tones at the initial stage. However, when the lexical tone combinations became more complicated, such as in the training of disyllabic Chinese words, a within-modality training method (TA) may be more helpful than a cross-modality method (GO) for the learning of L2 tones. Further research could explore whether this advantage extends to more complex tone-combinations such as those found in idioms or sentences.

References

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M.J. Munro (Eds.), Language Experience in Second Language Speech Learning: In Honor of James Emil Flege (pp. 13-34). Philadelphia: John Benjamins.

McNeill, D. (2008). Gesture and Thought. Chicago: University of Chicago Press.

Morett, L.M., & Chang, L.Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. Language, Cognition and Neuroscience, 30(3), 347-353.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2015). R package 'lmertest' version 2.0.

High-variability Phonetic Training and training Set Size considerations: Are L2 learners able to learn more than five L2 vowel categories at a time?

Cristina Aliaga-Garcia
University of Barcelona, Spain
cristinaaliaga@ub.edu

A vast literature has documented how high-variability phonetic training (HVPT) leads to significant and robust gains in learner's perception and/or production of new L2 speech sounds. In the field of L2 pronunciation, HVPT has constituted a consolidated area of research that holds great promise in pronunciation training, even with those non-native features that seem highly resistant to FL instruction (Iverson, Pinet, & Evans, 2012). However, full understanding of the suitability of HVPT in the FL instruction settings requires additional research (Barriuso & Hayes-Harb, 2018). With regards to the efficacy of HVPT for L2 vowel learning, for language pedagogy and curriculum designers, there are several considerations regarding the *training set size*: does efficient learning require a *full set of vowels* (varying in degree of difficulty) or a *small subset* of difficult vowels (assuming that L2 learners are able to generalize their learning to untrained vowel categories) (Nishi & Kewley-Port, 2008)? The present study aims to (1) assess whether L2 learners are able to learn *more than five L2 vowel categories at a time*, and are able to generalize their learning to untrained words and new talkers, which seems of particular relevance to real-world L2 learning and the application of HVPT in real teaching settings. It also aims to (2) assess whether L2 perception-only or production-only HVPT methods seem more efficient for improving both L2 vowel perception and production (Aliaga-Garcia, 2017).

A total of 84 bilingual Catalan-Spanish adult learners of English were divided into two experimental groups and a control group (N=20), and all were pre- and post-tested on identification of natural vowels, identification of duration-manipulated vowels, vowel discrimination, and a delayed repetition task. 64 participants were trained on the 11 English vowel monophthongs through two different audiovisual HVPT methods consisting of ten 60-minute sessions over five weeks: identification (ID) training or articulatory (ART) training (Nishi & Kewley-Port, 2003). Whereas ID training required them to focus on the relevant audiovisual cues for more accurate vowel identification, ART training required them to focus on the critical cues for more accurate vowel articulation. In both cases, auditory feedback provided assistance to correct trainees' identification or to change erroneous articulations.

This paper discusses and compares remarkable effects of perception- and production-based HVPT methods, as well as the generalization of gains to new words, talkers and contexts in the perception and production domains.

References

Aliaga-Garcia, C. (2017). The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan/Spanish learners of English. *Unpublished doctoral thesis.*

Barriuso, T., Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal*, v30, 1, 177-194.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English *Applied Psycholinguistics, 33 (01),* 145-160.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech Language and Hearing Research, 51(6),* 1480-1493.

Change of speaking fundamental frequency under a decade in young male speakers

Tekla Etelka Gráczi[1,2], Valéria Krepsz[1], Anna Huszár[1], Alexandra Markó[2,3]
[1] MTA Research Institute for Linguistics, Budapest, Hungary, [2] MTA–ELTE „Lendület" Lingual Articulation Research Group, Budapest, Hungary, [3] Department of Applied Linguistics and Phonetics, Eötvös Loránd University, Budapest, Hungary
graczi.tekla.etelka@nytud.hu

Fundamental frequency (f0) varies across the speaker's life span because of the physiological changes due to primary and secondary aging (Busse, 2002), but further factors may also contribute to the value of the change of speech characteristics, like the speech style, emotional background, sociolinguistic factors, etc. Most and heaviest aging-induced changes in the physiology appear in elder ages, however, the changes start in early adulthood, around the age of 30-35 years (Lalley, 2013). The change of fundamental frequency characteristics along the subject's age is a well-studied phenomenon, however, the longitudinal and cross-sectional analyses have brought contradictory results in a few questions. Cross-sectional analyses showed that f0 declined in male speakers steadily from 4 to 50 years of age and then began to steadily rise (e.g., Stathopoulos, Huber, & Sussman, 2011, summarized data: Baken, 2005). Fouqet and her colleagues' results (2016), however, showed relatively stable f0 values in 10 young adult subjects in 7 years periods following puberty.

In studies using a lifespan of 29 to 35 years interval starting from the 20s or 30s of the subjects, the results for the change in f0 are contradictory. Harrington, Palethorpe and Watson (2007) and Reubold, Harrington, and Kleber. (2010) found decreased f0 values in the elder ages, while Decoster & Debruye (2000) did not find a uniform change in this measure, however, the values typically decreased in their results, too. The typical life span in adult longitudinal studies override 20 years with a few exceptions, however, changes can be detected in a shorter period, as well. Verdonck-de Leeuw & Mahieu (2004) showed a change in f0 after five years, but their speakers were men over the age of 50, and the changes in psychic and smoking effects were significant along with age, while Fouqet et al. (2016) reported stable f0 after puberty. Most longitudinal studies involve a low number of speakers, reading aloud or prepared speech (e.g., Christmas speech by Elisabeth II in Harrington et al., 2007 and Reubold et al., 2010), while others use sustained vowels and text fragments (Verdonck-de Leeuw & Mahieu 2004).

The present study focuses on the longitudinal aspect of the speaking fundamental frequency in spontaneous speech (SP), reading aloud (RA) and interpreted speech (ISP) in young male speakers in a period of a decade (10-11 years). Our question was if we can detect a difference in the f0 in male subjects in the period of a decade, and if this change can be traced back in RA, ISP and SPSP, representing three distant points of a style scale.

13 male subjects of a Hungarian longitudinal corpus in preparation were selected (Gráczi et al., 2020). Their age at the first recording was between 19 and 40 years (mean: 27 ys, sd: 5 ys), and 29 and 50 years at the second recording (mean: 38 ys, sd: 5 ys). The first samples were recorded for the BEA database (Neuberger et al., 2014), while the second ones were recorded for the Longitudinal corpus. All participants are native speakers of Hungarian without any speech or hearing deficits. Samples of three speech styles were selected for analysis: (i) a spontaneous speech task ("interview", henceforth: SP) in which the subjects talk about their job, family, hobbies in a quasi-monological way, (ii) reading aloud (RA) of a scientific outreach text and (iii) the spontaneous summarization of a historical anecdote (ISP).

The speech samples were labeled manually in Praat (Boersma & Weenink, 2019). F0 analysis was run on each speech interval that appeared between two consecutive pauses or silences of any length and any type. Speech intervals that included noise, laughter or simultaneous speech were ignored.

F0 was measured in 50 ms running windows with 25 ms overlaps in Praat (Boersma & Weenink, 2019). The f0 minimum and maximum were set to 50 to 300 Hz, while the further settings of Praat were used in their default values. The values outside the 1.5 * of the IQR range of the f0 in a given speech sample (one speech style in one age for each speaker) were eliminated. The f0 mean was calculated in Hz, and the f0 range of the entire speech sample in semitones by Hmisc package (Harrel & Dupont, 2017). The range was calculated in two ways: an global value for the given speech sample, and a local for each interpausal unit in the given speech sample. Linear mixed models were used (Bates et al., 2015) in R (R Core Team, 2018). Speakers were set as random intercept. 8 models were built: 4 with one factor with/without random slope, 1 with two factors without and 2 with random slope for one of these, 1 with both. The best model was chosen based on the AIC number (Akaike, 1974) with the help of ANOVA (Kuznetsova, Brockhoff, & Christensen, 2017). The f0, the overall f0 range and the local f0 range were set as dependent variables. In the case of the overall f0 range, model (viii) was not built as the number of random effects cannot be equal to the number of observations.

The mean f0 was lower by at least 5% in the case of 8 speakers in ISP (mean change: 10.7%), 6 in RA (mean change: 11.4%) and 4 in SPSP (mean change: 14.8%). A decrease of at least 5% was found in 3 speakers in ISP (mean change: 19.4%), 4 in RA (mean change: 16.7%) and 2 in SPSP (mean change: 14.4%). Whose f0 mean increased or decreased in SPSP by at least 5%, it also showed the similar tendency in the other speech styles. The best fitting model was the one including both age and speech style with random slopes for both factors, however, only the speech style had significant effect on the f0 ($F(2, 12.156)= 8.713$, $p=0.004$).

The global f0 range decreased by at least 5% in 2 speakers in ISP (mean change: 12.4%), in 6 in RA (mean change: 23.3%) and in 12 in SPSP (mean change: 27.9%), and increased at least 5% in 9 speakers in ISP (mean change: 32.6%), in 6 in RA (mean change: 33.7%) and in 1 in SPSP (41.7%). The best fitting model was the one including both age and speech style as factors with a random slope only for age. The interaction of age and speech style had a significant effect on the overall f0 range ($F(2, 47.998)=8.890$, $p<0.001$).

The local f0 range decreased by at least 5% in 2 speakers in ISP (mean change: 22.7%), in no one in RA and in 6 in SPSP (mean change: 14.9%), and increased at least 5% in 9 speakers in ISP (mean change: 31.9%), in 11 in RA (mean change: 68.8%) and in 7 in SPSP (35.9%). The best fitting model was the one including both age and speech style with random slopes for both factors, however, only the speech style had significant effect on the f0 ($F(2, 12.372)= 4.6788$, $p=0.031$).

The f0 mean and f0 distribution and their change was largely dependent on speech style. The difference between the two speech styles necessarily did have an effect on the results, as SPSP is either less formal both towards the addressee and in its topic, than ISP and the RA. Also the speech planning processes are different across the speech styles. Therefore the effect of this factor cannot be neglected. When analyzing changes in speech parameters, also further factors have to be considered that play a large role in the interspeaker variability: Results of longitudinal studies on formants (Harrington Palethorpe, & Watson, 2000) and on speaking rate (Quené, 2013) also suggested that as reasons behind the changes in the speech characteristics not only the subject's age-related physiological changes but also further factors (e.g., sociolinguistic factors, like the addressee) have to be taken into consideration. Also a decade is a short time period in that large changes are not expected, as these would possibly signal pathological issues. In the present results the changes were diverse in certain sense, however speaker identification may be affected by changes undergoing in a decade in the subject's speech characteristics. A corpus is under development that will include at least 40 subjects recorded twice with a decade's span (Gráczi et al., 2020). The first recordings are from the BEA database (Neuberger et al., 2014). The second, follow-up recording are being carried out under the same circumstances.

This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. FK-128814.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19/6. pp. 716–723.

Baken, R.J. (2005). The Aged Voice. A New Hypothesis. *J. of Voice* 19/3, pp. 317–325.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, pp. 1–48.

Boersma, P., Weenink, D. (2019). *Praat: doing phonetics by computer.* http://www.fon.hum.uva.nl/praat/download_win.html. (Downloaded: 2019. október 1.),

Busse, E. W. (2002). General Theories of Aging. In: Copeland, J. R. M., Abou-Saleh, M. T., Blazer, D. G. (eds.) *Principles and Practice of Geriatric Psychiatry, Second Edition.* Willey-Blackwell

Decoster, W., Debruyne, F. (2000). Longitudinal Voice Changes: Facts and Interpretation. *J. of Voice* 14(2), 184–193

Fouquet M, Pisanski K, Mathevon N, Reby D. (2016). Seven and up: individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood. *R. Soc. opensci.* 3: 160395

Gráczi, T.E., Huszár, A., Krepsz, V., Száraz, B., Damásdi, N., Markó, (A. 2020). Longitudinális korpusz magyar felnőtt adatközlőkről. [Longitudinal corpus of Hungarian adult speakers]. In: *Magyar Számítógépes Nyelvészeti Konferencia 2020.* pp. 103–114.

Harrel, F. E. & Dupont, C. (2017) Hqmisc: Harrell Miscellaneous. R package version 3.16-0.

Harrington, J., Palethorpe, S. & Watson, C. I. (2007). Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Proc. of Interspeech 2007*, pp. 2753–2756.

Harrington, J., Palethorpe, S., Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature* 408. 927.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82. pp.1–26.

Lalley, P. M. (2013). The aging respiratory system-Pulmonary structure, function and neural control. *Respiratory Physiology & Neurobiology* 187, 199–210

Neuberger T., Gyarmathy D., Gráczi T. E., Horváth V., Gósy M., Beke A. (2014). Development of a large spontaneous speech database of agglutinative Hungarian language. In: *Proc. of the 17th International Conference, TSD 2014*, September 8-12, 2014., Brno. pp. 424–431.

Quené, H. (2013). Longitudinal trends in speech tempo: The case of Queen Beatrix. *JASA* 133, EL452. 452–457

R Core Team (2018). *R: A Language and Environment of Statistical Computing.* R Foundation for Computing, Vienna.

Reubold, U., Harrington, J., Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication* 52, 638–651

Russell A., Penny L., Pemberton C. (1995). Speaking fundamental frequency changes over time in women: a longitudinal study. Journal of Speech, Language and Hearing Research 38, 101–109.

Stathopoulos, E.T., Huber, J. E. & Sussman, J. E. (2011). Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4–93 Years of Age. J*. Speech Lang. Hear. Res.h* 54, pp. 1011–1021.

Verdonck-de Leeuw, I. M., Mahieu, H. F. (2004). Vocal aging and the impact on daily life: a longitudinal study. *J. Voice* 18(2), 193–202

The effect of the partner's age on backchanneling behaviour

Valéria Krepsz [1], Dorottya Gyarmathy[1], Csilla Ilona Dér[1,2], Ágnes Hámori[1], Viktória Horváth[1]
[1] MTA Research Institute for Linguistics, Budapest, Hungary
[2] Károli Gáspár Református Egyetem

Does our partner's age influence the way we interact in conversations? Numerous studies on spoken language have documented the changes of speech production across the lifespan (e.g. Mortensen et al. 2006). The speech production of the older adults may differ, in e.g. temporal characteristics of spontaneous speech, the characteristics of pauses, and other factors such as the realization of overlapping speech, the development of turns, and the realization of backchannels (BCs) (Zellner-Keller 2006; Bóna 2014).

The BCs signals or BCs responses (BCs) have been considered an important phenomenon of discourse, based on Yngve's distinction between the speech produced by the person holding the turn and the listener who occupies the 'back' channel of communication (Yngve 1970). Their research has developed into an important field in the analysis of conversations (e.g. Ward–Tsukahara 2000; Deng 2009). In studies on BCs there are two main theoretical approaches: formal models investigate the linguistic form or structure of these signals, while functional descriptions focus on their function in conversations, however less is known about the phonetic characteristics of BCs and about their relationships to the functional aspects. According to a previous study (Hámori et al. 2019), BCs responses can be ordered in main types based on their form and phonetic realization; this types are not distinct classes with clear boundaries but gradable categories around some prototypes (Taylor 2008). They constitute a continuum, depending on length, lexicality and the specificity of semantic content of the BCs response. At one end of the continuum are short, non-lexical, non-verbal BCs signals, such as humming ("hm") or sighing; in the middle are longer verbal and more lexical responses such as "yeah", "yes" and "okay" or "right"; the other end is constituted by fully lexical words or multi-word BCh constructions such as "oh dear" or "oh my God". The form (length, lexicality etc.) varies according to the actual function of backchanneling in the conversation. The use and realization of BCs can be influenced by many factors, e.g. by the environment (e.g. noise circumstances), the current physical and mental state of the speaker, biological and sociolinguistic factors (e.g. age, gender) and the communicative and social goals of the speaker and the conversation partner(s). It is important to highlight, that in conversation, the speech and verbal behaviour of the speaker is inseparable from the social context of the speech situation and from the person of other participants. It can be an especially significant in conversations with older people, where both speech accommodation ("elderspeak") (Coupland et al. 1988; Kemper 1994) and politeness phaenomena (Culpeper et. al. 2017) often play a major role. However, there are few research results on effect of age or of the partner's age on the speakers' BCs responses. Some studies (Gould–Dixon 1993; Geertzen 2015) found that younger adults produce more BCs than older adults in various circumstances, indicating explicitly the active monitoring of the partner's production. Other result shows that the total length of the conversation and the ratio of overlapping speech were influenced by age of the experimental person (S1 speaker) (Bata 2009).

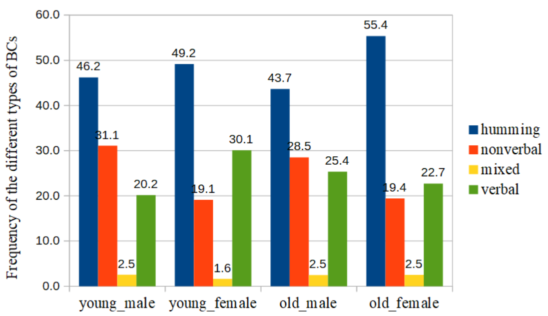
Based on the previous research, the questions may arise (i) to what extent the age of the partner influence the backchanneling behaviour (ii) whether the characteristics of BCs can be considered as individual factors. The present study examines only one speaker's BCs patterns addressed to listeners of different ages to reduce the number of influencing characteristics, and to focus on the age factor. The aim of our study is the multi-dimensional analysis of BCs in the spontaneous conversations with regard to the participants' age, at the very first time in Hungarian language. Our main questions were: i) what are the BCs' main phonetic characteristics, ii) is there a difference in the realization of BCs of the interviewer depending on the partners' age, iii) is there some special pragmatic role of BCs (esp. politeness) which correlate with the differences in phonetic realization. We investigated these questions in a functional pragmatic-sociolinguistic approach (Verschueren 1999; Culpeper et al. 2017).

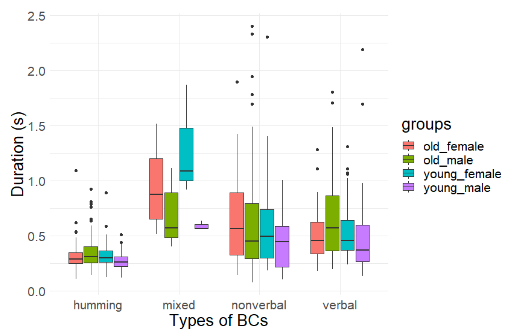
20 conversations (about 250 min., average duration: 13,5; min: 7 min.; max: 23,5 min.) of the Conversational Corpus (Horváth et al. 2019) were analysed and annotated using Praat software (Boersma–Weenink 2014). The recordings were quasi-spontaneous due to the fact that current topic of the conversation was added at the beginning of the recording. The participants had no previous time to plan the speech production, and the execution was carried out depending on the other’s opinion and utterances. Three people participated in each conversation: the interviewer (who coordinated the recording itself), the fieldworker (the colleague of the interviewer, the third participant in the conversations) and the subject. The interviewer (Int) and the fieldworker (S2) were the same in each case. The S1 speakers were different in the conversations (5 young males, 5 young females ages between 20–35 years, mean: 28 years, and 5 old males, 5 old females ages between 60–75 years, mean: 69 years, with typical hearing thresholds). In this study, the Int was the experimental person. The Int is a monolingual Hungarian female speaker working in a linguistics field. She was 28 years old at the beginning of making the recordings. The time between the first and last recording is about 10 years.

The BCs were analyzed in the Int’s speech in terms of the S1 speaker’s age on one hand. On the other hand, the realization of BCs was analyzed with regard to the global structure of the conversations. The aspects of our analysis were: i) form and phonetic characteristics of the BCs (verbal or nonverbal like laughing; duration; frequency; whether they were produced in the partner’s silent pause or it was an overlapping etc.) and ii) main pragmatic functions. The realization of BCs’ was analyzed with regard to the global structure using the following method: The internal text units were annotated whether they were rather narrative or conversational parts of the conversations (cf. Hutchby – Woofitt 2006). The duration of these internal text units were divided into 5 equal sections using a Praat script. The occurrence of BCs was analyzed in these equal parts. Statistical analysis will be carried out using SPSS 20.0 and R softwares (GLMM and nonparametric tests).

The results of the complex (pragmatic, structural and phonetic) analysis provide new information on the role of BCs in managing interaction and on characteristics of Hungarian language conversations. Preliminary results showed that most of the BCs, regardless of age, were humming (e.g. *ühüm*, *hm*), with fewer but almost equal proportions of verbal types and laughter. In terms of age, preliminary results showed a significant difference in the BCs addressed to older and younger speakers. The interviewer (Int) produced significantly more BCs in conversations with older speakers than with young people. There was a difference not only in the number of BCs, but also in their types in terms of the speakers’ age and gender. More humming was detected in the conversations with young people while more verbal types occurred in the conversations with older people (Fig 1). The higher number of BCs as well as different types of BCs for older speakers may be a feature of elderspeak but can also be considered to be a signal of honour and part polite behavior.

Figure 1: *Frequency of the different types of BCs (on the left) and their duration (on the right) depending on the S1 speaker’s age and gender*





The structure of the conversation also influenced the occurrence of BCs: more BCs were found at the end of the conversational units than in the initial parts. Furthermore, more BCs were detected in the narrative units than in the dialogic units. The explanation of the results can be propounded by the theory of speech accommodation and by the pragmatic theory of politeness as well.

The research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. NKFIH-K-128810.

References

Bata S. 2009. A társalgás fonetikai jellemzőinek alakulása a beszédpartnerek életkorának függvényében. [Phonetic properties of conversations depending on the participant's age.] In Váradi Tamás (szerk.): III. Alkalmazott Nyelvészeti Doktorandusz Konferencia. Bp. MTA NyTI. 3–13.

Boersma, D. & Weenink, P. 2014. Praat: doing phonetics by computer. Version 5.4.

Bóna, J. 2014. A spontán beszéd sajátosságai az időskorban [Features of spontaneous speech by older speakers]. ELTE Eötvös Kiadó, Budapest.

Coupland, N., Coupland, J., Giles, H. & Henwood, K. 1988. Accommodating the elderly: Invoking and extending a theory. Language and Society Vol. 17, Issue 1. March 1988. 1-41.

Culpeper, J., Haugh, M. & Kádár, D. Z. 2017. The Palgrave Handbook of Linguistic (Im)politeness. Palgrave.

Deng, X. 2009. Listener response. In: D'Hondt, S., Ostman, J. & Verschueren, J. (eds) The Pragmatics of Interaction. Amsterdam - Philadelphia: John Benjamins. 104–125.

Geertzen, J. 2015. Exploring age-related conversational interaction. https://www.semanticscholar.org/paper/Exploring-age-related-conversational-interaction-Geertzen/62385d55d359c17694122829f8af8affd93b12be

Gould, O. N. & Dixon, Roger A. 1993. How we spent our vacation: Collaborative storytelling by young and old adults. Psychology and Aging, 8(1). 10–17.

Hámori, Á., Dér, C. I., Gyarmathy, D., Krepsz, V. & Horváth, V. 2019. Backchannels: An old-new phenomenon bringing new perspectives to discourse analysis. Poster presented on the 16th International Pragmatics Conference (IPRA), Hong Kong, 9-14. June 2019.

Horváth, V., Krepsz, V., Gyarmathy, D., Hámori, Á., Bóna, J., Dér, Cs. I. & Weidl, Zs. 2020. Háromfős társalgások annotálása a BEA-adatbázisban: elvek és kihívások. [Annotating tryadic conversations of Hungarian Spontaneous Speech Database: principles and challenges.] Nyelvtudományi Közlemények (megjelenés alatt/in press).

Hutchby, I. & Wooffit, R. 2006. Conversational analysis. Principles, practises and applications. Polity Press, Cambridge.

James, D. & Clarke, S. 1993. Women, men, interruptions: a critical review. In Tannen, Deborah (ed.): Gender and conversational interaction. New York – Oxford, OUP. 231–280.

Kemper, S. 1994. Elderspeak: Speech accommodations to older adults. Aging, Neuropsychology, and Cognition, 1(1), 17–28.

Mortensen, L., Meyer, A. S. & Humphreys, G. W. 2006. Age-related effects on speech production: A review. Language and Cognitive Processes, 21(1-3): 238–290.

Taylor, J. R. 2008. Prototypes in cognitive linguistics. In Robinson, P. – Ellis, N. C. (eds.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge. 39-65.

Yngve, V. 1970. On getting a word in edgewise. In Campbell, M. A. et al. (eds.): *Papers from the Sixth Regional Meeting [of the] Chicago*. Chicago Linguistic Society. 567–577.

Verschueren, J. 1999. *Understanding Pragmatics.* London – New York –Sydney – Aukland. Arnold.

Ward, N. &Tsukahara, W. 2000. Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics 32. 1177–1207.

Zellner-Keller, B. 2006. Ageing and speech prosody, Speech Prosody – 3th International Conference, May 2–5, Dresden, Germany, Proceedings.

Effects of f0 level, speech rate and articulation rate on age perception

Marina Rohloff
Institute for German Studies, Carl von Ossietzky University Oldenburg, Germany
marina.rohloff1@uol.de

The aging process affects every phase of human speech production. Besides the cognitive skills which are required in the planning process, initiation, phonation and articulation are all subject to age-related changes (Braun & Fiebris, 2009, p. 142). These changes have an impact on multiple acoustic measures like speech and articulation rate (Hoit & Hixon, 1987, Braun & Fiebris, 2009), f0 mean and median (c.f. Stathopoulos et al., 2011, Harnsberger et al., 2008, Hollien & Shipp, 1972, Ferrand, 2002, Linville, 2001) and on measures of voice quality like harmonics to noise-ratio (HNR), jitter and shimmer (Stathopoulos et al., 2011, Linville, 1988).

The acoustic changes allow listeners to estimate the age of a person relatively precisely (Braun & Cerrato, 1999, Schötz, 2005). Previous studies show that age perception is mainly affected by the average speaking fundamental frequency and the speech and articulation rate (Brückl, 2007, Harnsberger et al., 2008, Skoog Waller et al., 2015, Skoog Waller & Eriksson, 2016). In addition, the results of Huntley et al. (1987) and Shipp and Hollien (1969) suggest that the listener age might affect the age perception since they found a varying precision in the estimations depending on the listener age.

The aim of this study is to determine whether f0 median, speech rate and articulation rate affect age perception and, if so, whether age and sex of the recipient have an influence on the age perception depending on these acoustic variables.

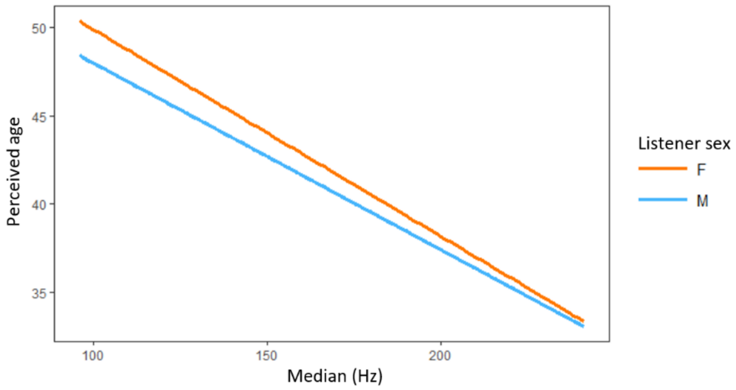
To answer these questions an online survey was developed, containing two different speech samples of 50 native speakers of the northern standard variety of German. The speakers were selected from three different age groups. The first group covered the age-span from 20 to 39 years (average age: 27.85 years), containing 10 male and 10 female speakers. The age of the second group ranged from 40 to 59 years (average age: 52.33 years) with 5 male and 10 female speakers. The last group, consisting of 9 male and 6 female speakers, ranged from 60 to 85 years (average age: 67.47 years).

One of the two speech samples was a section from a picture-based story telling task while the second one was a sustained vowel /a/. After a short demographic questionnaire, the participants of the online survey were asked to judge age and sex of each speaker per speech sample. 165 people of varying age (from 15 to 71 years old) participated in the survey. Each age group of speakers was estimated most precisely by listeners of the same age group while overall for the first group (20 to 39 years) the estimations show the smallest discrepancies to the real age. There was no difference in the estimation precision regarding the speaker or listener sex.

The data were analyzed by performing separate linear mixed effect models for each acoustic variable using the lme4 (Bates et al., 2015) and nlme (Pinheiro et al., 2018) packages in R (R Core Team, 2018) with perceived age as a fixed factor and speakerID and listenerID as random factors. As dependent variables f0 median, speech rate, articulation rate, perceived sex, listener age and listener sex were used as well as the interactions between the acoustic variables and perceived sex, listener age and listener sex.

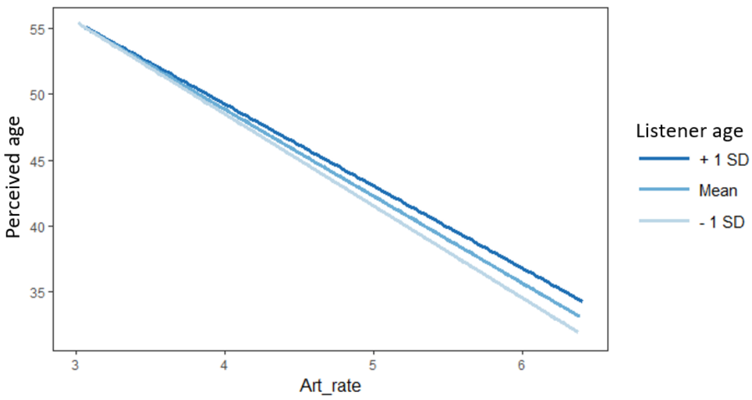
The results indicate for both task types that the f0 median correlates with the age estimation in a way that a rising f0 median tends to correspond with a decreasing age estimation. In addition, female recipients seem to estimate the age in the story telling task slightly higher than the male recipients depending on f0 median (see Figure 1). Regarding the vowel task, effects were found for the interaction between f0 median and listener sex and for f0 median and perceived sex. Posthoc tests with subsets revealed that only male listeners estimating a perceived female voice and female listeners estimating a perceived male voice reached a level of significance for f0 median. Both followed the general tendency described above.

Figure 1: *Interaction between f0 median and listener sex for the story telling task.*



While speech rate showed no significant effects, the articulation rate is associated with the age estimation. The estimations tend to decrease with a rising articulation rate independent of listener age. While this tendency is true over all listeners, an increase of the listener age results in slightly higher estimations in speakers with a higher articulation rate (see Figure 2). There was no effect for listener sex or speaker sex.

Figure 2: *Interaction between articulation rate (in syllables per second) and listener age for the story telling task.*



The general tendencies are in line with the results of Brückl (2007), Harnsberger et al. (2008) and Skoog Waller and Eriksson (2016) and fit the acoustic development of an aging voice. In addition, it could be shown for f0 median that listener sex (and for the vowel the perceived sex) influences the age estimation and should be considered while working with such factors. This might be explainable through a higher familiarity with the voice of the own sex, but then listeners should be able to estimate the age of a speaker with the same sex more precisely, which is not the case for the data in this study.

The results for the articulation rate indicate firstly that older listeners tend to estimate the age of a speaker slightly higher than younger listeners and secondly that this effect becomes more obvious with the increase of the articulation rate. The explanations of these two effects and their connection remain a topic for further research.

References

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

Braun, A. & Cerrato, L. (1999): Estimating speaker age across languages. In: Proceedings of ICPhS (Vol. 99). S. 1369-1372.

Braun, A., & Fiebris, S. (2009). Phonetic cues to speaker age: A longitudinal study. In: Grewendorf, G. & Rathert, M. (eds.), Formal Linguistics and Law. Berlin: de Gruyter. S. 141-162.

Brückl, M. (2007): Women's Vocal Aging: a Longitudinal Approach. In Eighth Annual Conference of the International Speech Communication Association. S. 1170-1173.

Ferrand, C. T. (2002): Harmonics-to-Noise Ratio: An Index of Vocal Aging. In: Journal of voice, 16(4). S. 480-487.

Harnsberger, J. D., Shrivastav, R., Brown Jr, W. S., Rothman, H., & Hollien, H. (2008): Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. In: Journal of voice, 22(1). S. 58-69.

Hoit, J. D. & Hixon, T. J. (1987): Age and Speech Breathing. In: Journal of Speech, Language, and Hearing Research, 30(3). S. 351-366.

Hollien, H. & Shipp, T. (1972): Speaking Fundamental Frequency and Chronological Age in Males. In: Journal of speech and hearing research, 15(1). S. 155-159.

Huntley, R., Hollien, H., & Shipp, T. (1987): Influences of listener characteristics on perceived age estimations. In: Journal of Voice, 1(1). S. 49-52.

Linville, S. E. (1988): Intraspeaker variability in fundamental frequency stability: An age-related phenomenon? In: The Journal of the Acoustical Society of America, 83(2). S. 741-745.

Linville, S. E. (2001): Vocal Aging. San Diego: Singular Thompson Learning.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team (2018): nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-137. URL: https://CRAN.R-project.org/package=nlme.

R Core Team (2018): R: A language and environment for statistical computing. Vienna: R Foundation for Statistical computing. URL: https://www.R-project.org/. Version: 1.1.442.

Schötz, S. (2005): Stimulus Duration and Type in Perception of Female and Male Speaker Age. In: Ninth European Conference on Speech Communication and Technology. S. 529-532.

Shipp, T. & Hollien, H. (1969): Perception of the Aging Male Voice. In: Journal of Speech and Hearing Research, 12(4). S. 703-710.

Skoog Waller, S. & Eriksson, M. (2016): Vocal Age Disguise: The Role of Fundamental Frequency and Speech Rate and Its Perceived Effects. In: Frontiers in psychology, 7. Art. 1814.

Skoog Waller, S., Eriksson, M., & Sörqvist, P. (2015): Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. In: Frontiers in psychology, 6. Art. 978.

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011): Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4-93 Years of Age. In: Journal of Speech, Language, and Hearing Research, 54(4). S. 1011-1021.

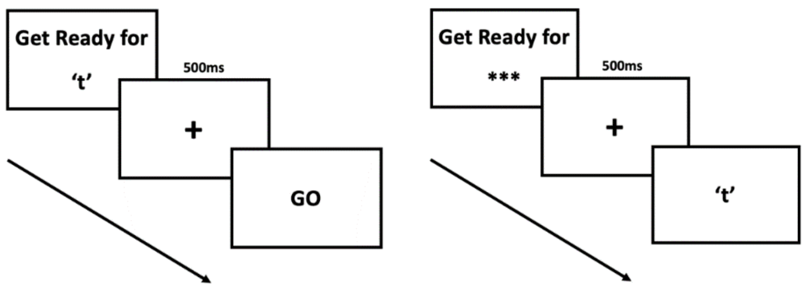
Decomposing age effects on speech motor planning

Chen Shen & Esther Janse
Centre for Language Studies, Radboud University, The Netherlands
{c.shen|e.janse}@let.ru.nl

Apart from well-known age effects on word finding, speech production in cognitively healthy older adults may be affected by age-related declines in the planning and execution of speech movements (Tremblay et al., 2018). To follow up on these speech movement results, our study attempted to pinpoint which aspect of speech motor planning may be susceptible to age-related decline. To this end, following related attempts to identify the locus of the impairment in apraxia of speech (Maas & Mailend, 2012), we used a speeded production task in which participants produced a target nonword in two conditions (prepared vs. unprepared). In the 'prepared' (or, simple) condition, participants produced the nonword after having prepared it in advance and having waited for the go signal. In the 'unprepared' (or, choice) condition, participants could only start preparing target nonword production (from a choice of three known alternatives) upon receiving a critical-information cue. Age groups were expected to differ in their production latencies in both conditions, but we aimed to find out whether age differences increased or decreased going from prepared (simple) to unprepared (choice) conditions. Hence, this simple versus choice reaction-time paradigm can be used to investigate whether age groups differ primarily in speech production processes that *cannot* be pre-programmed (i.e., those processes evident from simple-condition RT, such as retrieving and unpacking of motor programs from a buffer), or in production processes that *can* be pre-programmed (i.e., internal preparation, as evident from the difference between choice and simple RT).

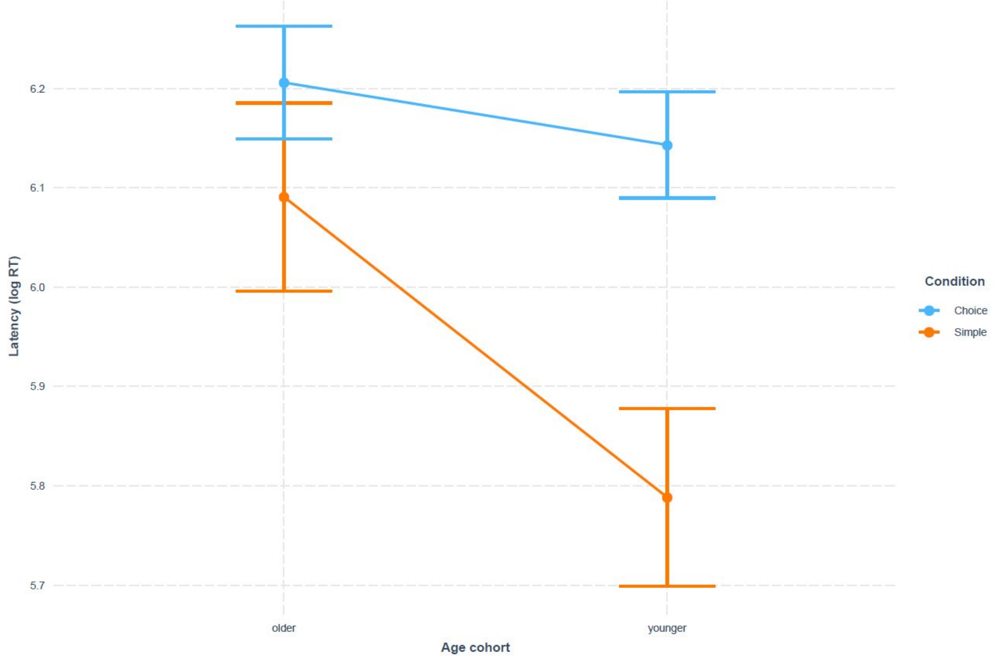
Speech production latency data were collected from a younger adult cohort (N = 30; aged 18 to 32 years) and an older adult cohort (N = 25, aged 65 to 77 years). Participants were asked to produce monosyllabic and disyllabic nonsensical target words that are composed of simple consonant-vowel syllables (i.e., /tu/, /ka/, and /bi/ as monosyllabic targets, and /tuka/, /kabi/ and /bitu/ as disyllabic targets). In the simple condition, speech onset was prompted by first visually presenting the initial phoneme of the target nonword, followed by a visual GO cue accompanied by a beep tone after 500 ms of preparation time. In the choice condition, instead of receiving the initial phoneme, participants were presented with a '***' sign. Speech onset was prompted 500 ms later by visually presenting the initial phoneme of the target nonword as a critical-information cue, accompanied by a beep tone (see Figure 1 below). Trials were blocked by target length and condition, with each block being preceded by 12 (monosyllabic condition) or 18 (disyllabic condition) familiarisation trials respectively.

Figure 1. *Illustration of the nonword production task in the 'Simple Condition' (left) and 'Choice Condition (right)*



Speech onset latency was measured as the period of time (in ms) between the onset of the beep tone (which served as the cue for participants to start speaking) and the onset of their speech. Speech onset latency data was analysed (log-transformed, for correct and valid responses only) showing effects of Condition (faster responses in the simple than choice condition) and of Target length (faster onset latencies for monosyllabic targets than disyllabic targets). Critically, there was an Age cohort by Condition interaction, such that the RT difference between Age cohorts was larger for the simple condition than for the choice condition (Figure 2). These data suggest that younger and older adults may differ more on aspects of speech production that cannot be pre-programmed (retrieving and sequencing of motor movements) than on internal speech planning aspects.

***Figure 2**. Interaction effect of Age cohort and Condition on speech onset latency (log-transformed)*



References

Maas, E., & Mailend, M. L. (2012). Speech planning happens before speech execution: Online reaction time methods in the study of apraxia of speech. *Journal of Speech, Language, and Hearing Research,* 55(5), S1523-S1534.

Tremblay, P., Deschamps, I., Bédard, P., Tessier, M. H., Carrier, M., & Thibeault, M. (2018). Aging of speech production, from articulatory accuracy to motor timing. *Psychology and Aging*, 33(7), 1022-1034.

Effects of Age, Sex and Education on Speakers' Preference of Filler Types in Japanese

Michiko Watanabe[1], Yuma Shirahata[2]
[1]National Institute for Japanese Language and Linguistics, [2]The University of Tokyo
watanabem@ninjal.ac.jp

Introduction

Fillers (filled pauses) are intrinsic in spontaneous speech. They are considered to reflect online speech planning processes. UHs and UMs have been most studied among English fillers. While there are arguments about their functions, Fruehwald (2016) claims that filler selection is a sociolinguistic variable. Fruehwald found that UM is replacing UH and that the change is being led by women. There are various filler types in Japanese (NINJAL 2006). Is the choice of filler types in Japanese also affected by sociolinguistic factors? This paper addresses this question. We examined effects of speakers' age, sex and education level on filler rates in 1628 informal presentation speeches, called "Simulated Public Speaking (SPS)", in "The Corpus of Spontaneous Japanese (CSJ)".

Method

Each presentation lasts between 10 and 15 minutes. The presentations contain 186,869 fillers in total. We first examined the total filler rate per word. Then, we analyzed the rate of five most common filler types in the corpus, ANOO, EE, MAA, EETO and SONO. EE and EETO exclusively function as fillers, whereas the others have functions as lexical words. EE is frequent in presentation speeches but rare in conversations. EETO is common in conversations as well as presentation speeches. ANO and SONO are demonstrative determiners like "that" and "the" respectively, with lexical accents. They are pronounced with flat F0 contours as fillers. MAA is used as interjection expressing surprise by women. MAA also has usage as adverb, immediately preceding estimation or evaluation. We calculated filler rates per word for each presentation. Then, we estimated statistical models for filler rates, using a generalized linear mixed model, with maximum likelihood estimation of variance components. We included speaker's age, sex and education level as fixed effects predictor variables and individual difference as a random variable. The speakers' age ranges from late teens to late 60s. 892 presentations are given by female speakers, and 736 by male speakers. Education level is categorized into three, junior high or high school graduate, university-educated and post graduate school graduate. We included interaction between the factors as a predictor variable when they were likely to have a significant effect. We used 'lmer' function in 'lme4' package and 'MuMIn' and 'lmerTest' packages running under R version 3.5.3. A model for total filler rate is expressed as follows: Total filler rate ~ age + sex + education + (1|speakerID)

Results

For total filler rate, main effects of age, sex and education were significant (Table 1). Older speakers use more fillers than younger speakers. Male speakers use more fillers than female speakers. University graduates use more fillers than high school graduates. As for ANOO, main effect of sex and an interaction between age and sex factors were significant (Table 2). While female speakers' use of ANOO increases with age, male speakers' use of ANOO falls after early 40s. As for EE, main effects of age, sex and education factors were significant (Table 3). Opposite to the total tendency, younger speakers use more EEs than older speakers. Male speakers use more EEs than female speakers. Speakers with university or higher degrees use more EEs than high school graduates. As for MAA, age and sex factors were significant (Table 4). Younger speakers use more MAAs than older speakers. Male speakers use more MAAs than female speakers. About EETO, only age factor was significant

(Table 5). Opposite to the total tendency, younger speakers use more EETOs than older speakers. About SONO, age and sex factors were significant (Table 6). Older speakers use more SONOs than younger speakers. Male speakers use more SONOs than female speakers. Table 7 shows results summary.

As a general tendency, older speakers use more fillers than younger speakers, and male speakers utter more fillers than female counterparts. Word retrieval ability is reported to start declining in one's forties (Sasanuma 1991). It is likely that lexical access takes older speakers longer and causes them to use more fillers than younger speakers. It is interesting to note that ANOO and SONO are more frequently used by older speakers, whereas EE, EETO and MAA are more often uttered by younger speakers. Inoue (2015) found that older speakers use more ANOOs than younger speakers in the Okazaki Survey conducted in 2008, which is in accordance with our results. ANO and SONO are determiners immediately preceding nouns or noun phrases. Speakers occasionally prolong ANO or SONO to buy time when they cannot timely retrieve nouns or noun phrases, as English speakers prolong "the" as "theee" when they hesitate. Such prolongations are likely to have caused ANO and SONO to have a function as fillers and suggest why older speakers who tend to need longer time for word retrieval use more ANOOs than younger speakers. This study indicates that sociolinguistic factors are relevant to speakers' choice of filler types.

Table 1 *Results of mixed-effects regression for total filler rate*

*(Signif. Codes for Table 1 - 6: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1)*

Variable		Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)		0.031	0.004	609	7.187	2E-12	***
Age		0.0003	0.0001	578.8	3.181	0.00155	**
Sex	Female	0					
	Male	0.025	0.003	584.9	8.992	<2e-16	***
Education	High school	0					
	University	0.025	0.003	668.5	3.112	0.00194	**
	Post graduate	0.013	0.008	625.5	1.691	0.09134	.

Table 2 *Results of mixed-effects regression for ANOO*

Variable		Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)		-0.0004	0.002	602.2	-0.19	0.850	
Sex	Female	0					
	Male	0.011	0.003	595.8	3.204	0.001	**
Sex:Age	Female	0.0005	0.000	587.7	8.083	0.000	***
	Male	0.0001	0.000	581.4	0.782	0.434	

Table 3 *Results of mixed-effects regression for EE*

Variable		Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)		0.009	0.002	596.7	4.206	0.00003	***
Age		-0.0001	0.0000	566.7	-2.083	0.038	*
Sex	Female	0					
	Male	0.013	0.0014	573.7	9.184	<2e-16	***
Education	High school	0					
	University	0.004	0.001	645.6	3.139	0.002	**

Post graduate	0.009	0.004	618.7	2.257	0.024	*

Table 4 *Results of mixed-effects regression for MAA*

Variable		Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)		0.010	0.001	600.5	6.753	3.4E-11	***
Age		-0.0001	0.000	583.8	-2.044	0.041	*
Sex	Female	0					
	Male	0.011	0.001	591.3	10.36	<2e-16	***

Table 5 *Results of mixed-effects regression for EETO*

Variable	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.006	0.001	591.3	8.579	<2e-16	***
Age	-0.0001	0.000	581.5	-5.069	0.000	***

Table 6 *Results of mixed-effects regression for SONO*

Variable		Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)		0.001	0.001	615.9	2.124	0.034	*
Age		0.0000	0.000	589.1	2.785	0.006	**
Sex	Female	0					
	Male	0.002	0.000	600.9	4.621	4.68E-06	***

Table 7 *Results summary*

Filler type	Age	Sex	Education
Total fillers	older > younger	male > female	univ. grad > high school grad
ANOO	Female: older > younger	male > female	n.s.
EE	younger > older	male > female	univ. and post-grad > high school grad
MAA	younger > older	male > female	n.s.
EETO	younger > older	n.s.	n.s.
SONO	older > younger	male > female	n.s.

This research is supported by JSPS KAKENHI, Grant Number 18K00559.

References

Fruehwald, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, 22(2), Article 6. https://repository.upenn.edu/pwpl/vol22/iss2/6. Download

Inoue, F. (2015). Late adoption of honorifics – "River Character" change in the Okazaki Survey of honorifics – ". *NINJAL Project Review*, 5(3), 98-107. Retrieved January 10, 2020, from http://doi.org/10.15084/00000779

National Institute for Japanese Language and Linguistics (NINJAL). (2006). *Construction of the Corpus of Spontaneous Japanese*. https://pj.ninjal.ac.jp/corpus_center/csj/en/document.html

Sasanuma, S. (1991). Aging and communication abilities. *The Journal of The Japan Society of Logopedics and Phoniatrics* 32, 203-205. https://www.jstage.jst.go.jp/article/jjlp1960/32/2/32_2_203/_article/-char/ja

How the impact of noise changes with development

Rochelle Newman
Department of Hearing & Speech Science, University of Maryland, USA
rnewman1@umd.edu

The speech input children receive is a critical factor in their language development, but much of the speech they hear occurs in noisy or multi-talker environments. Yet children must rely on these degraded signals to attempt to learn their native language. The language learning systems of the brain evolved in what were presumably far quieter ambient environments than present-day settings, where noise from traffic, television, and electronic devices are ubiquitous. Yet studies suggest children are generally more disadvantaged by noise and other distractors than adults (Fallon, Trehub, & Schneider, 2000; Newman, 2005), leading to a catch-22: young children who are still learning their language have a greater need for understanding speech in the presence of maskers, but are simultaneously less equipped to do so.

In this talk, I summarize work suggesting that children are not only more sensitive to background noise than are adults, but also that they process speech in noise quite differently. For example, infants and young children are particularly disadvantaged when the background consists of a single talker (Newman, 2009), whereas adults show the opposite pattern, performing better with a single-talker distractor than multi-talker babble (Bronkhorst & Plomp, 1992; Drullman & Bronkhurst, 2004; Pollack & Pickett, 1958; Simpson & Cooke, 2005). This difference is particularly striking when we consider the environments in which children live and learn: The home environment may frequently have a single other person talking in the background, either from the TV or radio, or a sibling or parent in an adjacent room. Indeed, the average US child is exposed to nearly 4 hours of background TV per day; for infants and toddlers, this amount increases to 5.5 hours per day (Lapierre, Piotrowski, & Linebarger, 2012). These sources of noise can distract children from other activities (Schmidt, Pempek, Kirkorian, Lund, & Anderson, 2008), and appear to be detrimental for children's cognitive development (Wachs, 1986).

Moreover, young children do not appear to make use of many of the cues that adults regularly use to separate different streams of speech. For instance, young children do not appear to benefit when a background speaker differs from a talker in sex (Leibold, Buss, & Calandruccio, 2019; Newman & Morini, 2017). Similarly, we have found that while adults and children aged 32 months show better performance when a noise masker is played in a different ear than the target, children aged 24 months showed no effect of spatial separation. In a third set of studies, children identified words in the presence of noise that either overlapped the speech range (centered on 1kHz) or did not (centered on 6 kHz). Children showed no impact of frequency range overlap, whereas adults showed stronger performance when the noise was distant from the speech in frequency range. Thus, in all three cases, acoustic cues that adults use to help them separate different sounds in their environment do not appear to be used by young children.

Finally, we present research examining the relationship between performance listening in noise and more domain-general attention skills. We find that among adults, individuals who show better speech perception in noise also perform better on visual selective attention tasks, providing evidence for the role of domain general selective attention skills on speech perception in noise. We are in the midst of testing whether these relationships also hold true for young children.

These various results all paint a picture in which children are not simply poorer at listening in noise than are adults, but process speech in noise in a qualitatively different manner. As a result, children are likely to be hampered in their ability to understand and learn from speech in environments that adults would not find noisy or distracting enough to be troublesome. Understanding how noise impacts children's ability to correctly interpret the speech they are hearing is critical to understanding how speech perception changes across the lifespan.

References

Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America, 92*(6), 3132-3139.

Drullman, R., & Bronkhurst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *Journal of the Acoustical Society of America, 116*(5), 3090-3098.

Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America, 108*(6), 3023-3029.

Lapierre, M. A., Piotrowski, J. T., & Linebarger, D. L. (2012). Background television in the homes of US children. *Pediatrics, 130*(5), 839-846.

Leibold, L. J., Buss, E., & Calandruccio, L. (2019). Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch. *Ear & Hearing, 39*(5), 935-945.

Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology, 41*(2), 352-362.

Newman, R. S. (2009). Infants' listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics, 71*(4), 822-836.

Newman, R. S., & Morini, G. (2017). Effect of the relationship between target and masker sex on infants' recognition of speech. *Journal of the Acoustical Society of America, 141*(2), EL164-169.

Pollack, I., & Pickett, J. M. (1958). Masking of speech by noise at high sound levels. *Journal of the Acoustical Society of America, 30*(2), 127-130.

Schmidt, M. E., Pempek, T. A., Kirkorian, H. L., Lund, A. F., & Anderson, D. R. (2008). The effects of background television on the toy play behavior of very young children. *Child Development, 79*(4), 1137-1151.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N* (L). *Journal of the Acoustical Society of America, 118*(5), 2775-2778.

Wachs, T. D. (1986). Noise in the nursery: Ambient background noise and early development. *Children's Environment Quarterly, 3*(1), 23-33.

Consonant production and early word-form processing are linked at 14 months of age

Irene Lorenzini & Thierry Nazzi
Integrative Neuroscience and Cognition Center, Université de Paris, CNRS
irene.lorenzini@parisdescartes.fr

Background

Speech perception and production are linked in adults (Skipper et al., 2017), which raises questions about the development and function of this neurocognitive link. Pioneering studies have highlighted a precocious influence of sensorimotor information on speech perception (Yeung & Werker, 2013; Bruderer et al., 2015; Choi et al., 2019). Furthermore, infants' perceptual processing has been shown to evolve in correspondence with the emergence of production as, at this stage, infants display a perceptual advantage for the speech sounds they can produce versus those they cannot yet produce. This mechanism has been documented in non-word recognition (DePaolis et al., 2010), audiovisual matching (Vilain et al., 2019) and word-learning (Majorano et al., 2019). While the direction(-s) of the perception/production coupling still has to be fully understood (Nazzi & Gonzalez-Gomez, 2012), the shared conclusion is that production and perception develop in tandem. We explored whether similar patterns exist for familiar word-form processing.

Methods

Two groups of healthy, typically-developing French-learning monolinguals (11- and 14-month-olds, N = 32 each) were recruited for this study. Each group was internally subdivided (i.e. median-split) into high- and low-Producers, based on a detailed parental questionnaire collecting the set of consonants that each infant was able to produce.

Differently from previous studies (e.g. DePaolis et al., 2010), our stimuli were not selected based on individual production (i.e. consonants that the participant does or does not yet produce), but along the broader distinction between easy-to-articulate (early-acquired) vs difficult-to-articulate (late-acquired) consonants (i.e. plosives vs fricatives). Using the Head-Turn Preference Procedure, the groups heard two lists, each containing 10 familiar words composed of easy- vs difficult-to-articulate consonants.

No familiarization phase was included. Word frequency, vowel context and syllabic length were varied within but balanced across lists. All words were spoken in mild Infant-Directed Speech by a French-native female speaker.

The degree of familiarity of the words used was determined following the methodology in Poltrock & Nazzi (2015) and verified *a posteriori* for each participant through a parental interview. This revealed that only a minority of the infants were not familiar with all the words given (31% at 11 months: Mean Unfamiliar Words = 4.5; 25% at 14 months: Mean Unfamiliar Words = 3) and that these subjects mostly belonged to the low-Producers groups (6 out of 7 at 11 months; 7 out of 10 at 14 months).

Results

At 11 months, an ANOVA on orientation times with Producer (high- versus low-Producer) and List (easy- versus difficult-to-articulate consonants) revealed a main effect of Producer ($F = 6.7$; $p = .01$; $\eta_p^2 = .18$), with longer orientation times in high producers. This signals a broad association between perception and production abilities whose underlying reasons need further clarification.

At 14 months, no main effect was found, but the Producer x List interaction was significant ($F = 8.1$; $p = .008$; $\eta_p^2 = .21$), due to high-Producers' longer orientation times towards the lists containing difficult-to-articulate consonants ($p = .005$). This is consistent with the fact that these infants were

starting to produce the difficult-to-articulate sounds. Conversely, the lack of Producer x List interaction at 11 months is consistent with the fact that fricatives were overall not produced in this cohort.
This study is the first investigation, to the authors' knowledge, to test the speech perception/production association with real familiar words. Our results agree with previous literature, describing a gradual emergence of this connection.

References

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. Proceedings of the National Academy of Sciences, 112(44), 13531-13536.

Choi, D., Bruderer, A. G., & Werker, J. F. (2019). Sensorimotor influences on speech perception in pre-babbling infants: Replication and extension of Bruderer et al. (2015). Psychonomic bulletin & review, 26(4), 1388-1399.

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? Infant Behavior and Development, 34(4), 590-601.

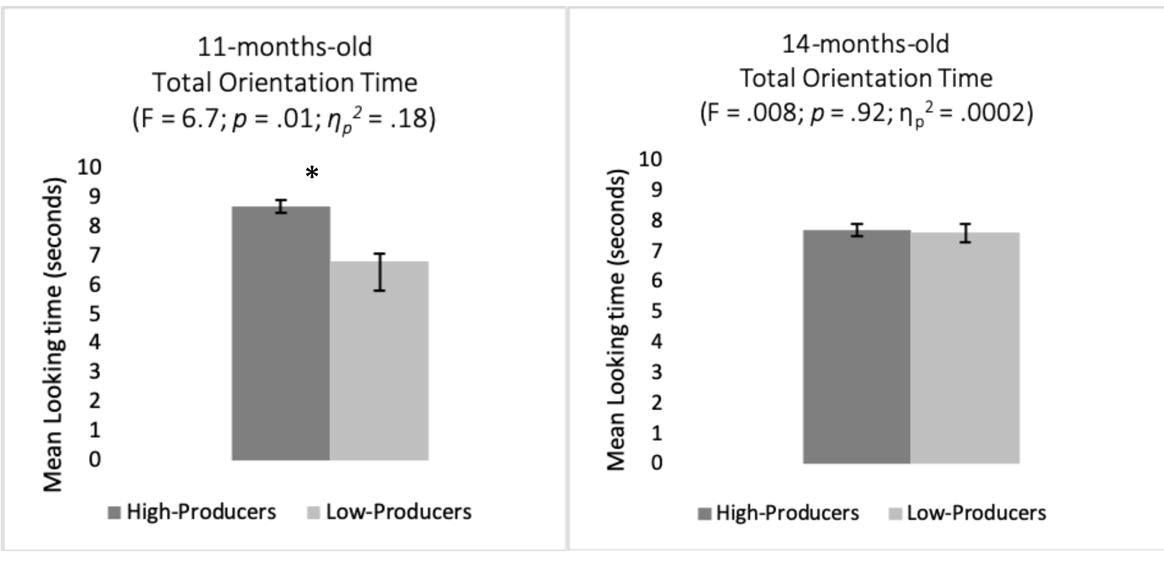
Gonzalez-Gomez, N., & Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy*, *17*(5), 498-524.

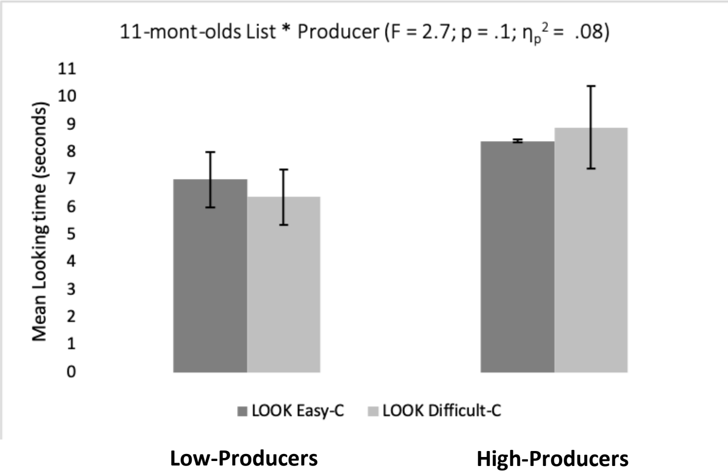
Majorano, M., Bastianello, T., Morelli, M., Lavelli, M., & Vihman, M. M. (2019). Vocal production and novel word learning in the first year. Journal of child language, 46(3), 606-616.

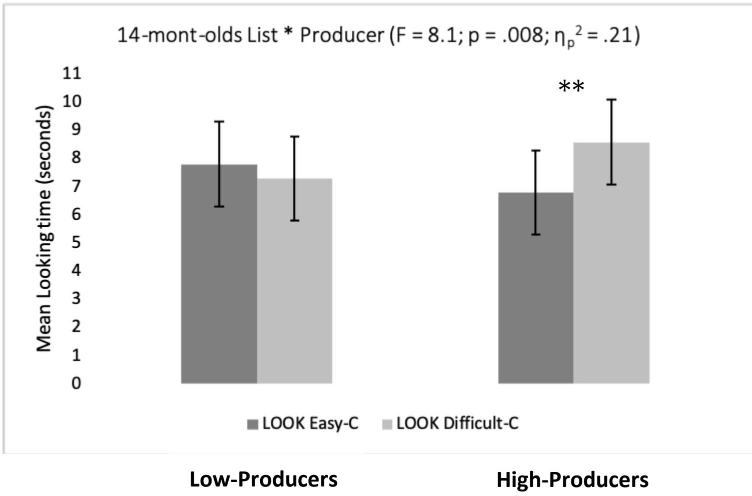
Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. Brain and language, 164, 77-105.

Vilain, A., Dole, M., Lœvenbruck, H., Pascalis, O., & Schwartz, J. L. (2019). The role of production abilities in the perception of consonant category in infants. Developmental science, 22(6), e12830.

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. Psychological Science, 24(5), 603-612.




11-mont-olds List * Producer (F = 2.7; p = .1; η_p^2 = .08)
Mean Looking time (seconds)
11
10
9
8
7
6
5
4
3
2
1
0
LOOK Easy-C
LOOK Difficult-C
Low-Producers
High-Producers


14-mont-olds List * Producer (F = 8.1; p = .008; η_p^2 = .21)
Mean Looking time (seconds)
11
10
9
8
7
6
5
4
3
2
1
0
**
LOOK Easy-C
LOOK Difficult-C
Low-Producers
High-Producers

Acquiring segmental and suprasegmental phonology of vowel length: development across the first year of life

Kateřina Chládková[1], Nikola Paillereau[2] & Šárka Šimáčková[3]
[1]Institute of Psychology, Czech Academy of Sciences, Prague, Czech Republic
[2]Institute of Phonetics, Charles University, Prague, Czech Republic
[3]Palacký University Olomouc, Czech Republic
k.chladkova@uva.nl

The speech signal is composed of numerous acoustic cues, some of which can be more relevant at the level of individual sound segments, and others can have greater relevance over entire phrases. An acoustic cue that can have linguistic function both at the segmental and the suprasegmental level is duration. Across many languages, duration serves as a cue to linguistic rhythm, through variations in speech rate, phrase-final, lexical-stress, or sentence-focus related lengthening (White et al. 2012). At the same time, in some languages, such as Czech, Finnish, and Japanese, duration cues contrasts between phonemically short and long vowels or consonants. Partly due to the multitude of linguistic roles that it can have, duration is particularly interesting from a developmental perspective. We do not yet know at what age children acquire phonemic length (see e.g. Tsuji & Cristia 2014) and how the acquisition of segmental length phonology interacts with the acquisition of suprasegmental patterns. The research presented here aims to help reveal how the phonemic and prosodic length functions are acquired.

Durational variations are perceptually salient from early on in a child's development. Durational information is veridically transmitted to the intrauterine environment (Graven & Browne 2008); fetuses indeed seem to learn prosodic temporal patterns through exposure and immediately after birth, infants show language-specific processing of temporal variations in rhythm (Granier-Deferre et al. 2011, Ramus 2002, Abboub et al. 2016). Relatively little is known about the early development of duration at the level of phonemes. Duration-cued segmental contrasts are considered perceptually highly salient, meaning that they could be acquired at a rather early age (Burnham, 1976). Empirical studies on perceptual acquisition of vowel length are rare and do not converge in their findings (Sato et al. 2010, Minagawa-Kawai et al. 2007).

We aimed to find out when infants learning a language with phonemic vowel length acquire adult-like sensitivity to length contrasts at the segmental level and how this ability interacts with their processing of durational variations at the suprasegmental level, i.e. rhythm. Our questions were tested with infants exposed to Czech. In speech production, Czech adults realize length contrasts through duration, phonemically long vowels are almost twice as long as their short counterparts (Paillereau & Chládková 2019). The prosodic function of duration in Czech is less clear: duration marks phrase boundaries but not lexical (fixed-position) stress (Dankovičová 1997, Skarnitzl & Eriksson 2017); in terms of the traditional rhythm metrics, Czech seems unclassifiable (Dankovičová & Dellwo 2007). For the acquisition of Czech durational patterns, one might thus predict that with developing length contrasts at the phonemic level, sensitivity to temporal rhythm variations will decline (which aligns with adult speech production in languages like K'ekchi and Hungarian, Berinstein 1979, Vogel et al. 2015). Alternatively, infants may simultaneously develop sensitivity to both phonemic length and language-specific temporal rhythm (in line with a formal description of many languages with quantitative contrasts, Lunden et al. 2017), and rely on durational rhythm cues that had not yet been identified in the speech of Czech adults.

Below we report two experiments. In Experiment 1, we traced infants' perceptual discrimination of the native /fɛ/-/fɛː/ length contrast (relative to the native /fɛ/-/fa/ vowel quality contrast). In Experiment 2, we tested infants' discrimination of typical and atypical temporal rhythmical patterns. In both experiments, 4 groups of infants were tested: 4-, 6-, 8-, and 10-month olds. Since durational cues are most likely amongst the first ones available (or, audible) to infants, we hypothesized that Czech infants will show a strong sensitivity to phonemic vowel length from the first age tested, i.e. from 4 months (relative to vowel quality, which typically develops at about 6 months), and that this sensitivity might get enhanced as a result of further perceptual attunement to native-language phonemes. For prosodic duration we, too, expected a strong sensitivity at the early age, 4 months, but hypothesized an age-related decline due to a relatively weak relevance of temporal cues in Czech (adult) prosody.

In Experiment 1, 79 infants were tested and after exclusion due to fussiness, experimenter error, or failure to habituate, data from ten 4-month olds, eight 6-month olds, seven 8-month olds, and six 10-month olds were analyzed. In a central fixation paradigm, infants were tested on their discrimination of a vowel length contrast, /fɛ/-/fɛː/, and on a spectral contrast, /fɛ/-/fa/. Each trial contained 16 repetitions of naturally produced syllables and was 13 s long. After habituating to [fɛ]-syllables, infants were presented with three types of test trials in a pseudorandomized order. The test trials contained alternations in terms of vowel duration [fɛ]-[fɛː], spectral change [fɛ]-[fa], or no change [fɛ]-[fɛ], i.e. control trials. We assessed the total looking time in each test trial type (duration, spectrum, control) and compared it to the average looking time of the last two habituation trials. Statistical inferences were done using linear mixed-effects models with Trial type and Age as fixed factors, and per-participant random intercept. There was a trending main effect of Trial type, showing that the duration test trials yielded longer looking times than the last two familiarization trials (by 1 s, $t = 1.789$, $p = .077$). Trial type interacted with Age: in older infants (8- and 10-month olds), spectral test trials yielded longer looking times than the last two familiarization trials (by 1.6 s, $t = 2.003$, $p = .049$).

In Experiment 2, 59 infants were tested, and after exclusion, we analyzed eight 4-month olds, thirteen 6-month olds, nine 8-month olds, and eight 10-month olds. The stimuli were well-formed Czech sentences containing low-frequency content words produced in infant-directed speech by three women. The three women imitated 30 naturally-produced model sentences with typical Czech rhythm, and the same 30 edited sentences with atypical rhythm. The atypical rhythm was realized through altering the duration ratios between stressed and unstressed syllables. Ten different trials were created, 5 trial per rhythm type, each trial containing 9 different sentences (3 from each speaker). Infants were tested in a central fixation paradigm, without habituation, and with pseudo-randomized order of the 5 typical-rhythm and the 5 atypical-rhythm trials (average trial length was 23 s). The dependent measure was first look duration during the typical-rhythm and the atypical-rhythm trials (averaged across trials of the same type). Linear mixed-effects models had Accent and Age as fixed effects, per-participant random intercepts, and per-trial-order random slopes for Accent. Accent interacted with Age, showing that at 4-months, infants looked significantly longer to the typical than to the atypical rhythm (by 2.1 s, $t = -2.30$, $p = .028$). Data collection in both experiments is ongoing (estimated completion by the time of the conference with n=14 per age in each experiment).

To sum up, we found that Czech infants between 4 and 10 months discriminate changes between a short and a long vowel duration that cues segmental phoneme identity. The overall sensitivity to segmental duration was not found to differ across development (unlike the sensitivity to spectral contrasts that seemed to rise between month 6 and 8). At the suprasegmental level, the 4-month olds showed reliable discrimination of typical and atypical temporal variations, preferentially listening to the typical renditions of their native-language rhythm. We did not find any rhythm-specific listening at later ages.

In conclusion, early in development, i.e. at 4 months, Czech infants are sensitive to durational changes at both the segmental and the suprasegmental level. However, the subsequent development of duration sensitivity seems to take different paths. Whereas the sensitivity to duration at the level of individual segments remains robust throughout development, perhaps reflecting early-established and maintained short and long phoneme categories, the sensitivity to duration at the suprasegmental level seems to decrease within the 1st year, perhaps reflecting the fact that suprasegmental variations in duration are phonologically little relevant in the infants' native language. Future studies in our lab will examine how durational changes affect Czech infants' perception of lexical stress and how the duration sensitivity interacts with their growing vocabulary.

References

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. Brain and Language, 162, 46–59.

Burnham, D. K. (1986). Developmental loss of speech perception: Expoure to and experience with a first language. Applied Psycholingistics 7, 201– 240.

Dankovičová, J. (1997). The domain of articulation rate variation in Czech. Journal of Phonetics, 25 (3), 287 - 312.

Dankovičová, J., & Dellwo, v. (2007). Czech Speech Rhythm and the Rhythm Class Hypothesis, In: Proceedings of ICPhS, Saarbrucken, 1241-1244.

Granier-Deferre, C. Ribeiro, A., Jacquet, A.-Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science* 14 (2), 336–352.

Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn and Infant Nursing Reviews*, *8*(4), 187–193.

Lunden, A., Campbell, J., Hutchens, M., & Kalivoda, N. (2017). Vowel-length contrasts and phonetic cues to stress: An investigation of their relation. Phonology, 34(3), 565–580.

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural Attunement Processes in Infants during the Acquisition of a Language-Specific Phonemic Contrast. *The Journal of Neuroscience*, *27*(2), 315.

Paillereau, N., & Chládková, K. (2019). Spectral and temporal characteristics of Czech vowels in spontaneous speech. *Acta Universitatis Carolinae— Philologica 2/2019, Phonetica Pragensia*, 77–95.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition* 2, 85-115.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, 46(1), 106.

Skarnitzl, R., & Eriksson, A. (2017). The Acoustics of Word Stress in Czech as a Function of Speaking Style. In: Proceedings of Interspeech, 3221-3225.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191.

Vogel, I., Athanasopoulou, A., & Pincus, N. (2015). Acoustic properties of prominence in Hungarian and the Functional Load Hypothesis. In Katalin E. Dekany, K. E. Kiss & B. Suranyi (eds.). Approaches to Hungarian 14. Amsterdam: John Benjamins, 267–292.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language, 66*(4), 665–679.

Modeling early phonetic development as representation learning

Yevgen Matusevych[1], Thomas Schatz[2], Naomi H. Feldman[2], Sharon Goldwater[1]
[1] University of Edinburgh, Edinburgh, UK; [2] University of Maryland, College Park, USA
yevgen.matusevych@ed.ac.uk

Infants' phonetic perception changes in the first year of life. For example, at the age of 6–8 months, English-learning and Japanese-learning infants are equally able to discriminate between English sounds [ɹ] and [l] (Kuhl et al., 2006). By the age of 10–12 months, the two groups diverge, showing attunement to the phonetic contrasts present in their input language. Similar results have been reported for other languages (Werker & Tees, 1984; Bosch & Sebastián-Gallés, 2003; Tsao et al., 2006, etc.). Despite the abundance of empirical results, the mechanisms of early phonetic learning are still not well-understood. Many theoretical accounts explaining early phonetic learning have been proposed (Kuhl & Iverson, 1995; Best, 1994; Werker & Curtin, 2005, etc.), but until recently no computational models could explain how the input to which infants are exposed leads to observed crosslinguistic differences in discrimination. A recent study by Schatz et al. (2019) presented the first such model, based on an algorithm used for unsupervised speech learning in engineering applications. Their model was able to correctly predict the crosslinguistic pattern of [ɹ]–[l] discrimination observed in infants. The success of this model raises the question of whether other algorithms for unsupervised learning from speech may lead to equally good, or even better, models of infants' early phonetic learning.

In this study we apply the approach of Schatz et al. (2019) to test five computational models on three data sets of infant phone discrimination from different languages. Doing so allows us to gain insight into the kinds of representations and learning mechanisms that infants are likely to employ. We run three sets of simulations, corresponding to the results of behavioral experiments with infants (see Table 1):

1. Kuhl et al. (2006): infants learning English or Japanese tested on English [ɹ]–[l].
2. Tsao et al. (2006): infants learning Mandarin or English tested on Mandarin [ɕ]–[t͡ɕʰ].
3. Bosch & Sebastián-Gallés (2003): infants learning Catalan or Spanish tested on Catalan [e]–[ɛ].

In each experiment, 8–12-month-olds were more successful in discriminating the test contrast if they had been learning the corresponding language (English, Mandarin, and Catalan, respectively).

We consider five computational models: the one introduced in Schatz et al. (2019) and four neural network models inspired by existing work in unsupervised speech representation learning. These models show high performance in low-resource speech technology applications, making them a good starting point for modeling unsupervised infant learning. In high-level terms, the models differ along several dimensions, as summarized in Table 1. Three of the models learn representations at the level of speech frames (i.e., 25-ms-long chunks of speech commonly used in automatic speech recognition), while two learn to encode word-sized units of variable length as vector representations of fixed length (somewhat similar to semantic word embeddings). Finally, three models are strictly unsupervised, while two others rely on top-down guidance from known wordforms—based on evidence that even 6–8-month-olds can segment and recognize some wordforms (e.g., Jusczyk & Aslin, 1995). The five models we use here are existing implementations that were developed for processing speech with minimal supervision: (1) Dirichlet process Gaussian mixture model (DPGMM: Chen et al., 2015; Schatz et al., 2019); (2) autoencoder (AE: Kramer, 1991); (3) correspondence autoencoder (CAE: Kamper et al., 2015); (4) autoencoding recurrent neural network (AE-RNN: Chung et al., 2016); (5) correspondence-autoencoding recurrent neural network (CAE-RNN: Kamper et al., 2019). Table 2 summarizes the basic characteristics of each model.

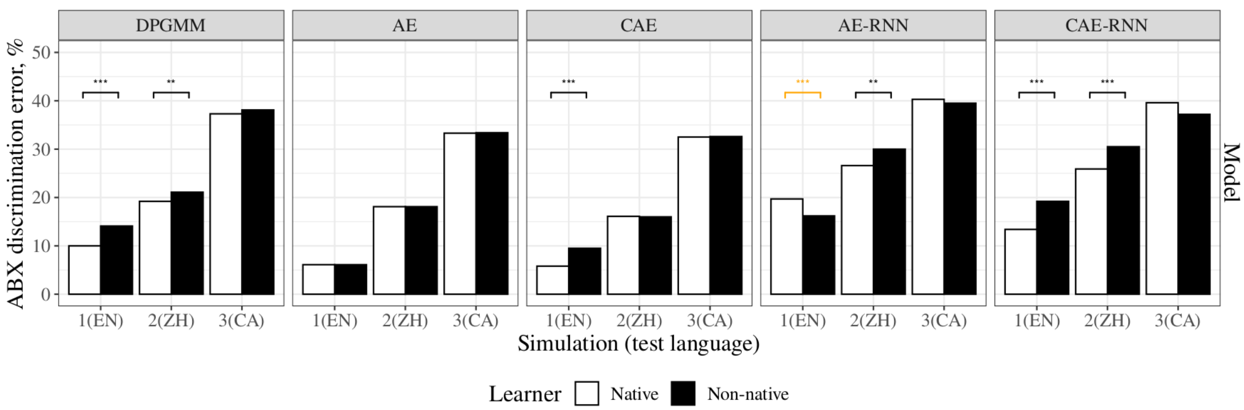
Table 1. *Training and test conditions.*

Simulation	Test language	Training language	Listener type
1	English	English	Native
		Japanese	Non-native
2	Mandarin	Mandarin	Native
		English	Non-native
3	Catalan	Catalan	Native
		Spanish	Non-native

Table 2. *Computational models used in our study.*

Model	Top-down guidance	Representation type
DPGMM	No	Frames (25 ms)
AE	No	Frames (25 ms)
CAE	Yes	Frames (25 ms)
AE-RNN	No	Word-sized units
CAE-RNN	Yes	Word-sized units

Figure 1*. ABX error rates of the five native and non-native models in the three discrimination tasks: English (EN) [ɹ]–[l], Mandarin Chinese (ZH) [ɕ]–[t͡ɕʰ], and Catalan (CA) [e]–[ɛ].*



We train each of the five models on a sample of a speech corpus of either the test language (to simulate a native speaker) or another language (as informed by the corresponding study with human infants, to simulate a non-native speaker).[1] At the end of learning, we evaluate the models on the machine ABX task (Schatz et al., 2013), which allows us to test their ability to discriminate between pairs of phones, similarly to the tests carried out with infants, such as conditioned head turn. To replicate the crosslinguistic patterns found in experimental data, the "native" version of the model is expected to show a significantly lower error rate on the test contrast than the "non-native" version of the model.

Figure 1 shows the discrimination error rates for the target phones. Significance of the differences between a given pair of computational models ("native" vs. "non-native") was computed using two-tailed ANOVA tests on the predicted values of mixed-effects models fitted to the error rates of the two computational models in question. Out of seven significant crosslinguistic differences found, six are in the expected direction, and one (the AE-RNN on the English contrast, the orange bracket in Figure 1) is in the wrong direction. Looking at these results, we observe that three out of five models (DPGMM, CAE, and CAE-RNN) show the infant-like crosslinguistic pattern of discrimination for English, three out of five models (DPGMM, AE-RNN, and CAE-RNN) show such pattern for Mandarin, and no models show significant differences for Catalan. The Catalan contrast is particularly difficult to discriminate even for "native" models: the corresponding discrimination error rates vary between 26.6% (CAE) and 38.2% (AE-RNN). We additionally tried to filter out very short phones (shorter than 80 ms) from the test data: the

[1] Speech data is preprocessed using standard speech recognition methods: we slice each segment into 25-ms-long frames and extract mel-frequency cepstral coefficients and their first and second derivatives from each frame.

models showed lower error rate on this data, but there was still no difference between the "native" and the "non-native" model.

Using computational modeling on realistic input, we compared five algorithms of representation learning in their ability to predict three crosslinguistic differences in phone discrimination observed in infants. In our study, Schatz et al.'s DPGMM model showed the infant-like crosslinguistic pattern of discrimination both for English [ɹ]–[l] and for Mandarin [ɕ]–[t͡ɕ^h], making it clear that the earlier result was not specific to a particular English contrast. Moreover, a second model of representation learning, the CAE-RNN, also made correct predictions on the same two contrasts. This result supports the idea that models learning representations directly from unsegmented natural speech can correctly predict some of the infant phone discrimination data. Based on the results of this study, the DPGMM and the CAE-RNN show promise as models of early phonetic learning, although their failure to capture the effect on the Catalan contrast warrants further investigation to determine whether the failure is a problem with the models or with the amount and/or quality of the training and the test data.

References

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman and H. C. Nusbaum (eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 233–277). MIT Press: Cambridge, MA.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46, 217–243.

Chen, H., Leung, C. C., Xie, L., Ma, B., & Li, H. (2015). Parallel inference of Dirichlet process Gaussian mixture models for unsupervised acoustic modeling: A feasibility study. In *Proceedings of INTERSPEECH* (pp. 3189–3193).

Chung, Y. A., Wu, C. C., Shen, C. H., Lee, H. Y., & Lee, L. S. (2016). Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder. In *Proceedings of INTERSPEECH* (pp. 765–769).

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.

Kamper, H. (2019). Truly unsupervised acoustic word embeddings using weak top-down constraints in encoder-decoder models. In *Proceedings of ICASSP-IEEE* (pp. 6535–3539).

Kamper, H., Elsner, M., Jansen, A., & Goldwater, S. (2015). Unsupervised neural network based feature extraction using weak top-down constraints. In *Proceedings of ICASSP* (pp. 5818–5822).

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, *37*, 233–243.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, F13–F21.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). York Press: York, England.

Schatz, T., Feldman, N., Goldwater, S., Cao, X., & Dupoux, E. (2019). Early phonetic learning without phonetic categories – Insights from large-scale simulations on realistic input. PsyArXiv [Preprint.] https://doi.org/10.31234/osf.io/fc4wh

Schatz, T., Peddinti, V., Bach, F., Jansen, A., Hermansky, H., & Dupoux, E. (2013). Evaluating speech features with the minimal-pair ABX task: Analysis of the classical MFC/PLP pipeline. In *Proceedings of INTERSPEECH* (pp. 1781–1785).

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, 120, 2285–2294.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Dynamics of speech production and perception across a language barrier

Ann Bradlow
Department of Linguistics, Northwestern University, Evanston, Illinois, USA
abradlow@northwestern.edu

The impact of non-dominant language status on the temporal domain of speech communication is evident at multiple levels, ranging from the phonetic level to the conversational level. In this presentation, I will discuss three strands of research that together aim to build a deeper understanding of the time-warping nature of a language barrier, particularly in relation to information transmission, that is, in the context of conversations with communicative intent.

First, at the phonetic level, one of the most salient differences between L2 and L1 speech production is the reduced tempo of L2 speech in terms of syllables produced per second. This reduced L2 speaking rate is well established across L1 and L2 talkers (e.g. L1 English versus L2 English) as well as within bilingual individuals (e.g. various L1s versus L2 English) (e.g. Guion et al., 2000; Baese-Berk & Morrill, 2015). Moreover, slower L2 than L1 speaking rates have been demonstrated in several L2s, including (but not limited to) L2 Spanish, L2 French, L2 Dutch, and L2 German (e.g. García Lecumberri et al., 2017; De Jong et al., 2013; Trouvain & Möbius XX). I will present data from a large corpus of both read and spontaneous speech recordings in both the L1 and L2 of a group of bilingual individuals (n=86) from various language backgrounds (n=10) showing that this dramatic difference between L2 and L1 speaking rates results from multiple interacting sources of variation. Using automatically extracted speaking rate measurements (syllables/second) (De Jong & Wempe, 2009), a comparison of L1 and L2 speaking rates within individual bilinguals revealed that, while speaking rate was always slower in L1 speech than in L2 speech, L1 speaking rate significantly predicted L2 speaking rate. That is, relatively fast or slow talkers in L1 were also relatively fast or slow in L2, respectively. These results indicate a persistent influence of a talker-specific articulatory setting, or "trait" characteristic, that combines with, rather than is overwhelmed by language dominance (i.e. a L1 versus L2 "state" characteristic) in the global temporal structure of bilingual speech production (Bradlow et al., 2017, see also De Jong et al., 2013; Derwing et al., 2009). In addition to state (L1 versus L2) and trait characteristics, a comparison across the various L1s of the bilinguals also revealed noteworthy language-specific speaking rate differences. Thus, the overall slow tempo of L2 speech reflects a confluence of L1 versus L2 state specificity, talker specificity, and language specificity.

Next, in order to gain further insight into the communicative impact of the distinctively slow tempo of L2 speech, the relationship between speaking rate (syllables/second) and information density (number of syllables for a given text) was examined in L1 and L2 recordings of a standard English reading passage (NWS, North Wind and the Sun passage) that was included in the bilingual speech corpus. This analysis followed the reasoning and methods of cross-language comparisons that have indicated an inverse relationship between speaking rate (syllables/second) and information density (number of speech units for a given meaning) such that information conveyed per second (information rate) remains relatively constant (Derwing et al., 2009; Pellegrino et al., 2011). Extending this approach to L2 speech allows us to see that L2 English productions of the NWS passage involved both slower rates (fewer intensity peaks (i.e. acoustic syllables) per second) and lower information density (production of more acoustic syllables) than L1 English. A follow-up comparison of the number of acoustic syllables (intensity peaks in the signal) versus orthographic syllables (dictionary-based counts of phonological syllables in the text) indicated substantial syllable reduction for L1 speech (number of acoustic syllables < number of orthographic syllables) but substantial syllable epenthesis (number of acoustic syllables > number of orthographic syllables) for L2 speech. Thus, compared to L1 speech, L2 speech involved information-sparse syllables (i.e. more syllables were produced to convey the same meaning/text) and slow speaking rates yielding an information transmission profile (bits of information conveyed per second) that, at an extreme, may fall outside the optimal range for human information processing of dynamic signals.

Finally, in order to assess the impact of a language barrier in a task with communicative intent rather than in decontextualized, laboratory-based, monologue recordings, I will show results of comparisons of task-completion time across pairs of L1 and L2 talkers (various combinations of L1 and L2 talkers) in a conversation-based, cooperative, picture-matching task, the diapix task (Van Engen et al., 2010; Baker & Hazan, 2011). These data showed that, while all pairs successfully completed the task, L2 talker pairs were substantially less efficient than L1 talker pairs in terms of both task-completion time and number of word repetitions (word type-to-token ratio). However, the L2 disadvantage for the diapix task was mitigated when the L2 conversation partners shared the same L1 as compared to L2 pairs in which the conversation partners came from different L1 backgrounds (Van Engen et al., 2010). We also found some evidence that the difference in communicative efficiency between L1 and L2 pairs diminished slightly across successive diapix task trials. Nevertheless, a comparison of the impact of a language barrier to that of other types of communication barriers, such as the presence of background babble or vocoded speech transmission (Hazan & Baker, 2011), has shown greater detriments to communicative efficiency for language barriers, most notably with respect to the time it takes to complete the diapix task.

Taken together, these three lines of research demonstrate an accumulation of time-related influences of the presence of a language barrier on speech communication. At the phonetic level, L2 speech production is distinguished from L1 speech by a slow tempo (fewer syllables per second) in combination with less syllable-level reduction, yielding a speech signal with a low information rate (bits of information conveyed per second) for a given text/meaning. While a direct link between information rate at the phonetic (i.e. syllable) level and communicative efficiency at the discourse level remains elusive and is undoubtedly compounded by lexical, syntactic, and other linguistic and cognitive components of L2 speech production and perception, it seems clear that speech communication across a language barrier provokes adjustment at multiple time scales. In this respect, a language barrier can be viewed as a multi-faceted time warp.

References

Baese-Berk, M. M., & Morrill, T. H. (2015). Speaking rate consistency in native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *138*(3), EL223-EL228.

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, *43*(3), 761-770.

Bradlow, A. R., Kim, M., & Blasingame, M. (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America*, *141*(2), 886-899.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, *41*(2), 385-390.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223-243.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533-557.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205-228.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139-2152.

Lecumberri, M. L. G., Cooke, M., & Wester, M. (2017). A bi-directional task-based corpus of learners' conversational speech. *International Journal of Learner Corpus Research*, *3*(2), 175-195.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 539-558.

Trouvain, J., & Möbius, B. (2014). Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. In *Proc. Speech Prosody* (Vol. 7, pp. 275-279).

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and speech*, *53*(4), 510-540.

Listening effort and hearing-aid benefit of older adults in everyday life

Inga Holube, Petra von Gablenz, Ulrik Kowalk, Jörg Bitzer
Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, Oldenburg, Germany
inga.holube@jade-hs.de

Common methods to assess hearing deficits and the benefit of hearing aids or other devices in clinical practice include questionnaires and speech tests under controlled conditions. For speech tests, words or sentences are typically presented in quiet or in (mostly stationary) noise coming from one or two loudspeaker directions. Laboratory hearing tests aim at creating reproducible results in controlled acoustic environments. However, listening demands in natural environments are imperfectly represented. In the course of hearing aid (HA) fitting, therefore, questionnaires are used to address the subjective perspective on hearing abilities. They focus on predefined listening situations and are completed retrospectively. Both methods, speech tests and questionnaires, rarely take into account the individual's specific needs. Hence, the impact of hearing impairment on the individual's everyday life and the benefit of hearing devices is not evident.

Weaknesses of this approach, e.g., memory bias and possible mismatch of the pre-defined and individually experienced listening situations, are overcome by Ecological Momentary Assessment (EMA), an ad-hoc query including prompt and repeated assessments in natural environments. Smartphones facilitate the implementation of questionnaires and rating schemes to be administered in individual environments of study participants or customers during or shortly after the experience. In addition, objective acoustical parameters extracted from head/body-worn microphones and/or settings from the hearing aid's signal processing unit can be stored alongside the questionnaire data. Advantages are participant-specific context-sensitive information on activities, experienced challenges, and preferences. Because the law in many countries does not allow audio recordings in order to preserve the privacy of all communication partners and bystanders, the information about environmental acoustics is limited to statistical data such as averages of e.g., levels and spectra. Other challenges for EMA are for example the unsupervised handling of the equipment by elderly people, the trade-off between the accuracy of description and the number of similar listening situations when performing comparisons (e.g., with and without hearing aids), the trade-off between the duration of recording intervals and the amounts of collected (and analyzed) data, the random or target-oriented reminder for subjective responses as well as the willingness and ability of the participants to respond while doing specific tasks.

We conducted an EMA study to examine how hearing aid uptake changes the perception of everyday hearing abilities in older adults. In collaboration with local hearing aid acousticians, 16 first-time and follow-up HA wearers were recruited. They used our smartphone-based EMA system olMEGA for 3-4 full days before HA fitting and also after HA acclimatization. The system includes one microphone on each side of the head and wireless data transmission to a smartphone. The audio signals recorded by the two microphones were analyzed in segments and deleted immediately thereafter. By storing only averaged characteristics of the environment's acoustical properties, the privacy of the user and bystanders was fully preserved. Further offline analysis aimed at own-voice detection and background estimation. The user query was carried out using an adaptive questionnaire that allows for specifying situations and sound sources as well as for assessing personal hearing-related dimensions like speech understanding, loudness, and listening effort. Study participants could perform an assessment at any time, but were also reminded about every 30 minutes.

EMA provided snapshots of individual listening experiences. Despite the participants spending most of their time at home, EMA profiles show considerable inter-individual variability both in the type of listening events assessed and hearing aid benefit. In this regard, EMA uncovers individual differences and hearing-related needs. However, this highly differentiated surveying also reveals its downside, particularly when aiming at a comparison of subjective assessments, e.g., before and after hearing aid uptake. Since everyday life may be similar but is never the same, the intra-individual match even for aggregated listening events was only moderate before and after intervention. As a consequence, the fragmentation of data and the variety of real-life assessments limit the possibilities to draw conclusions based on quantitative approaches. Nevertheless, the data was analyzed with respect to the experienced situations and the hearing aid benefit. From the collected subjective data, 933 queries out of a total of 1705 queries were related to speech listening events. Results showed a considerable individual variability regarding the type of reported events as well as the distribution and position of assessments. Overall, speech understanding improved by 1.1 scores and listening effort decreased by 1.3 scores on 7-point scales in post-intervention EMA compared to pre-intervention EMA.

This contribution shows results from a field study conducted with adults who were inexperienced in hearing testing beyond the regular ENT protocol. It points out pitfalls and opportunities that can arise when linking EMA to standard audiometric diagnostics. In summary, EMA gives new insights into the hearing-related quality of life on an individual basis. The method has advantages but bears challenges in terms of user's burden, hard- and software development, data analysis, and privacy issues.

Speech processing later in life: Processing complex sentences with declining hearing

Margreet Vogelzang[1] & Esther Ruigendijk[2]
[1] Department of Theoretical and Applied Linguistics, University of Cambridge, UK; mv498@cam.ac.uk
[2] Institute of Dutch Studies, University of Oldenburg and Cluster of Excellence 'Hearing4all', Germany

Older adults with mild-to-moderate age-related hearing loss typically exhibit issues with speech recognition and word understanding. Language ability beyond auditory (word) recognition, more specifically syntax, in hearing-impaired adults has not received much attention. It is generally known that for non-impaired populations, syntactically complex sentences are more difficult to process and understand than canonical sentences (see, e.g., Bader & Bayer, 2006; Tun et al., 2010; Wingfield et al., 2006). For example, Carroll et al. (2016) showed that German complex object-initial sentences were interpreted incorrectly more often than canonical subject-initial sentences and that complex sentences caused difficulties during on-line processing. In addition, difficulties with sentence processing typically occur when hearing takes place under adverse conditions, such as in sentence processing in noise, or with hearing impairment (e.g., Carroll & Ruigendijk, 2013; Stewart & Wingfield, 2009; Wingfield et al., 2006). Using neuroimaging methods, Peelle et al. (2011) found that already a mild-to-moderate decline in hearing ability results in less activation in the primary auditory cortex and several cortical brain regions of the speech network during auditory processing. Similar cortical regions as in Peelle et al. (2011) were found to be more active when listeners with normal hearing are presented with degraded speech (Bräuer et al., 2016).

Thus, there is evidence of syntactic complexity influencing sentence processing and comprehension, and evidence of hearing impairment influencing sentence processing and comprehension. Nevertheless, there is conflicting evidence on the relationship between syntactic complexity and hearing ability. Wingfield et al. (2006) found that, in comparison to normal-hearing peers, adults with mild-to-moderate hearing loss have more problems with the comprehension of complex object relative than subject relative sentences, especially at higher speech rates. Tun et al. (2010) found no effects of elderly adults' hearing on the comprehension of object relatives, but did find that adults with mild-to-moderate hearing loss responded slower to comprehension questions about complex sentences than normal-hearing peers, whereas this difference was absent in simpler canonical sentences. Moreover, complex sentence processing may differ between aided and unaided hearing impaired listeners under similar hearing conditions (Habicht et al., 2017). However, there are also studies that examined syntactic complexity and hearing ability that found no interactions between the two (e.g., Carroll et al., 2016; Wendt et al., 2015). So, despite these contributions, the relationship between hearing loss and higher-level language processes, such as syntax processing, is not well understood yet.

It is generally accepted that hearing impairment adds perceptual processing load. The speech processing difficulties that are typically found for hearing-impaired populations can thus be explained by perceptual difficulties caused by degraded input in combination with the increased demands of cognitive load or control required to process such degraded input (cf. Carroll & Ruigendijk, 2013; Mattys et al., 2009). In addition, the processing of complex sentences compared to canonical sentences has also been argued to add memory or processing load (e.g., Gordon et al., 2002; Grodzinsky, 2005). Therefore, processing cognitively taxing complex sentences from cognitively taxing degraded input may be increasingly taxing. The aim of our study is twofold. Firstly, we test the hypothesis that, indeed, hearing-impaired listeners show increased difficulties with processing complex syntactic structures and that these difficulties are due to increased demands on cognitive control with degraded auditory input that interfere with cognitive control processes required to process syntactically complex sentences (Hypothesis 1). Secondly, we examine with both behavioural and neural measures whether this perceptual and cognitive load can be reduced by hearing aid use and/or increased cognitive abilities (Hypothesis 2).

We performed an auditory sentence-processing paradigm followed by a picture-selection task with two pictures: one correct and one with thematic role-reversal. In German, subject-before-object order is the canonical word order, but structurally complex object-before-subject sentences can also be formed. By manipulating subject-object order and adjunct position we created 4 word orders (see Table 1), with SVAO being the canonical baseline condition. Sound levels were adjusted for each participant to 80% intelligibility. We manipulated cognitive load by introducing a secondary (dual) task, a fixation cross change detection task, in half of the trials. Neural processing (fMRI) and sentence comprehension were recorded and cognitive tests tapping into *working memory* (backwards Digit Span), *cognitive flexibility* (Trail Making), and *vocabulary size* were used. 20 mild-to-moderately hearing-impaired non-hearing aid users (HI; mean age 64.8), 19 hearing-impaired hearing aid users (HA; mean age 65.5; mean hearing aid experience 6.1 years) and 20 age-matched normal-hearing controls (NH; mean age 64.1) were tested.

***Table 1**. Examples of the four experimental conditions. Note that although the conditions use different word orders, their meaning remains the same.*

Subject-object order	**Adjunct position**	**Condition name**	**Example sentence**
sub-before-obj	3	SVAO	Der Igel berührt am Montag den Hasen
obj-before-sub	3	OVAS	Den Hasen berührt am Montag der Igel
sub-before-obj	1	AVSO	Am Montag berührt der Igel den Hasen
obj-before-sub	1	AVOS	Am Montag berührt den Hasen der Igel
			On Monday the_{NOM} *hedgehog touches* the_{ACC} *hare*

The influence of sentence condition, cognitive load and cognitive abilities on comprehension was examined with linear mixed-effects models. The results show more correct answers to canonical sentences compared to complex sentences ($p < 0.001$). The most difficult condition, AVOS, rendered only about 60% correct responses. In addition, an effect of hearing ability was found, with NH listeners outperforming both HI and HA listeners ($p < 0.01$), an effect that is driven by differences in performance on syntactically complex sentences. No differences between HI and HA participants were found. Surprisingly, the dual task did not influence offline comprehension. All of our cognitive measures (*working memory, cognitive flexibility, vocabulary size*) correlate with sentence comprehension (corrected p's < 0.05) in all three groups (HI, HA, and NH), indicating that participants with better cognitive abilities performed better on the picture selection task. This influence was most pronounced in the complex sentences with object-before-subject order (OVAS and AVOS). No effects of hearing aid experience were found.

The neuroimaging data were analysed as the difference in activity between the canonical baseline (SVAO) and the other sentence conditions. The results show effects of syntactic complexity in NHs in all non-canonical sentence conditions (Figure 1), but only in the most difficult AVOS condition in both groups of hearing-impaired listeners. Possibly, canonical sentence processing is already effortful for listeners with hearing loss and therefore no differences between the sentence conditions are observed. Under increased cognitive load (dual task), the NH group shows processing similar to the processing of the HI and HA groups in the single task; no differences between sentence conditions are found, except in the most difficult AVOS condition. Effects of hearing aid experience (in the HA group) can be seen in the most difficult AVOS condition (both in the single and the dual task; see Figure 2), with increased activity in areas including the right superior frontal gyrus and right precentral gyrus, possibly reflecting higher-level processing, correlating with longer hearing aid experience in the single task, and activity in the precuneus correlating with longer hearing aid experience in the dual task. Correlations with cognitive measures (entered as covariates) did not reach significance.

Figure 1. *Activity in the different conditions (single task) for normal-hearing participants, Red = OVAS, Green = AVSO, Blue = AVOS. (p < 0.05; FWE corrected on the cluster level)*

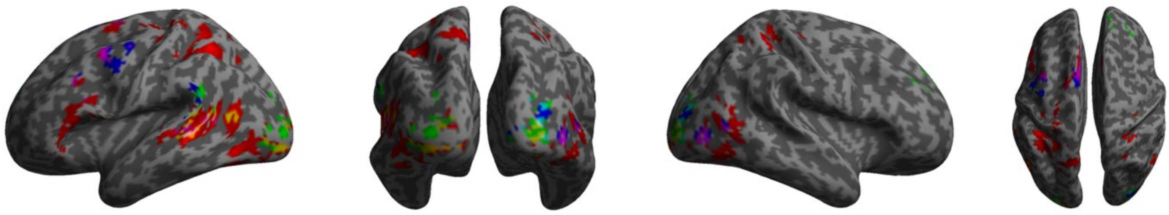
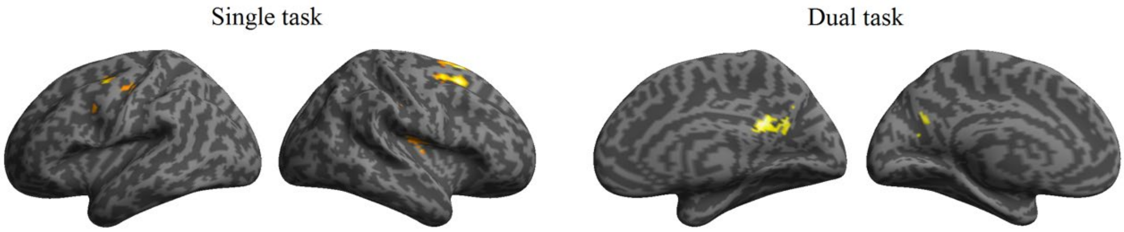


Figure 2. *Influence of hearing aid experience in AVOS. (p < 0.05; FWE corrected on the cluster level)*



In sum, these results show effects of hearing loss and of hearing aid experience on auditory complex sentence processing. In comprehension, hearing loss impairs complex sentence processing, but not canonical sentence processing. In neural processing, this is reflected in effects of complex sentence processing for normal-hearing, but not hearing-impaired adults. Under increased cognitive load, the normal-hearing group shows processing more similar to the processing of the hearing-impaired groups, indicating that increased demands on cognitive control with degraded auditory input may be influencing the processing of hearing-impaired adults (in line with Hypothesis 1). Hearing aid experience did not influence comprehension, but did influence neural processing of complex sentences. We interpret this finding as illustrating that the brain adapts to hearing aids and that this influences language processing, indicating that it is important to investigate the adaptive capacity of the brain with regard to the effect of hearing aids. Finally, it was shown that cognitive abilities strongly influence comprehension of complex sentences, but no differences between groups were found, indicating that cognitive abilities reduce the processing load of complex sentences, but do so in the same way for hearing-impaired and normal-hearing listeners (partially in line with Hypothesis 2).

References • Bader, M., & Bayer, J. (2006). *Case and Linking in Language Comprehension: Evidence from German*. Berlin, Germany: Springer. • Bräuer, S. et al. (2016). The role of ventral and dorsal pathway in sublexical speech perception: insights from fMRI. In *8th Annual Meeting for the Neurobiology of Language*. London, UK. • Carroll, R., & Ruigendijk, E. (2013). The Effects of Syntactic Complexity on Processing Sentences in Noise. *Journal of Psycholinguistic Research*, *42*(2), 139–159. • Carroll, R. et al. (2016). Processing Mechanisms in Hearing-Impaired Listeners: Evidence from reaction times and Sentence Interpretation. *Ear and Hearing*, *37*(6), e391–e401 • Gordon, P. C. et al. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430. • Grodzinsky, Y. (2005). Syntactic Dependencies as Memorized Sequences in the Brain. *Canadian Journal of Lingustics*, *50*, 241–266. • Habicht, J. et al. (2019). Exploring Differences in Speech Processing Among Older Hearing-Impaired Listeners With or Without Hearing Aid Experience: Eye-Tracking and fMRI Measurements. *Frontiers in Neuroscience*,*13*(May), 1–14. • Mattys, S. L. et al. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. • Peelle, J. E. et al. (2011). Hearing loss in older adults affects neural systems supporting speech comprehension. *Journal of Neuroscience*, *31*(35), 12638–12643. • Stewart, R., & Wingfield, A. (2009). Hearing Loss and Cognitive Effort in Older Adults' Report Accuracy for Verbal Materials. *Journal of the American Academy of Audiology*. • Tun, P.A. et al. (2010). Response latencies in auditory sentence comprehension: Effects of linguistic versus perceptual challenge. *Psychology and Aging*, *25*(3), 730–735. • Wendt, D. et al. (2015). How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing. *Trends in Hearing*, *19*, 1–18. • Wingfield, A. et al. (2006). Effects of Adult Aging and Hearing Loss on Comprehension of Rapid Speech Varying in Syntactic Complexity. *Journal of the American Academy of Audiology*, *17*(7), 487–497.

The effect of healthy ageing on auditory scene analysis – evidence from a change detection paradigm

Mathilde de Kerangal[1], Deborah Vickers[2,3], Maria Chait[2]
[1]Department of Bioengineering, Imperial College London, London, UK
[2]Ear Institute, University College London, London, UK
[3]Clinical Neurosciences Department, University of Cambridge, Cambridge, UK
m.de-kerangal@imperial.ac.uk

The ability to detect a change in a complex auditory scene, such as the appearance or the disappearance of a source, is essential in daily situations. Change detection has been studied quite substantially in the young population (Cervantes-Constantino et al., 2012; Sohoglu and Chait, 2016 a and b; Aman et al., 2018) but less is known about the effect of age on this ability.

We used a change detection paradigm, where listeners had to concurrently monitor six streams of tones — each characterised by a unique carrier frequency and temporal pattern — and listen out for the disappearance of one of them. Two age groups (N=40 each) of young (20–35) and older (60–86) adults with close to normal hearing were tested. Additional profile measures targeting hearing acuity, speech-in-noise perception, auditory and visual sustained attention were also collected. The study examined two aspects of scene analysis via the change detection paradigm.

First, sensitivity to temporal regularity was investigated by comparing performance on scenes that consisted of randomly patterned vs. regular streams. Whilst the older participants performed significantly worse than the younger listeners in the “random” condition (d′ and reaction time), there was no difference between groups in the “regular” condition. This suggests a relatively preserved sensitivity to regularity in the ageing brain.

The second set of experiments investigated the effect of age on auditory distraction. A brief distractor sound was added at the time of disappearance to half of the “regular” scenes presented to the participants; the other half did not include a distractor. The presence of the distractor degraded performance (d′ and reaction time) in both groups. However, an increase in criterion was only observed in the older group, demonstrating an inhibitory effect of the distractor specific to older age.

Overall, we found no correlation between change detection performance and major indicators used in the clinic, including age, audiometric measures or speech-in-noise perception. However, older listeners’ change detection ability consistently correlated with measures of auditory sustained attention and musical training. These results demonstrate that good measures of attentive abilities are critical for understanding how hearing in crowded environments is affected by healthy ageing.

References

Aman, L., Picken, S., Andreou, L., Chait, M., 2018. Sensitivity to statistical structure facilitates perceptual analysis of complex auditory scenes. bioRxiv 126763.

Cervantes Constantino, F., Pinggera, L., Paranamana, S., Kashino, M., Chait, M., 2012. Detection of Appearing and Disappearing Objects in Complex Acoustic Scenes. PLOS ONE 7, e46167.

Sohoglu, Ediz and Maria Chait (2016a). ‘Detecting and representing predictable structure during auditory scene analysis’. eLife 5.

- (2016b). ‘Neural dynamics of change detection in crowded acoustic scenes’. NeuroImage 126, pp. 164–172.

Intelligibility of speech improves after perceptual vowel training in L2 learners of English

Jayanthiny Kangatharan[1], Anastasia Giannakopoulou[2], Maria Uther[3]
[1]University of Winchester; [2] University of Bedfordshire, [3]University of Wolverhampton, United Kingdom
anastasia.giannakopoulou@beds.ac.uk

Previous research (Giannakopoulou et al., 2013) has shown that perceptual training helps L2 English adult learners to perceptually shift their attention to primary cues to correctly identify vowels in L2. Specifically, the finding that perceptual cue weighing can be re-learnt through high-variability perceptual training indicates that the perceptual training procedure appears to be a robust technique for improved outcomes in perceptual identification and learning of L2 speech sound categories. It has been shown that perception training of L1 Japanese speakers that resulted in the enhanced identification of English /r/-/l/ also led to the enhanced production of English /r/-/l/ after completing the perception training, such that production samples were more intelligible than before training (Bradlow et al., 1997). Previous research indicated that L2 speakers' pronunciation accuracy could affect the overall intelligibility of L2 speech in native English listeners (Purcell and Suter 1980; Varonis and Gass 1982). Thus, undergoing perceptual learning can help L2 speakers acquire a non-native perceptual contrast. It is therefore clear that there can be a transfer of L2 speakers' perceptual learning onto L2 speech production. This can suggest that L2 speakers' acquisition of non-native phonetic contrasts could affect their control over how that contrast is produced (e.g. Bradlow et al., 1997).

This study investigates whether the production of L2 speech by L2 adult learners was affected by high-variability training. One goal of this study was to investigate whether perceptual training improves production as assessed by native English listeners in an orthographic transcription task. A second goal was to investigate whether perceptual training improves production as assessed by native English listeners in a perceptual goodness rating task. Ratings of accuracy and clarity were tested from production samples of a previous study (Giannakopoulou, 2012), which aimed to find out whether L2 listeners' attention could be directed to cues that are critical for correct perceptual identification, for them to correctly weight the cues when identifying and discriminating between L2 phonetic segments. In that study L2 learners were trained with natural and modified duration stimuli using a list of minimal pairs depicting the tense-lax /i:/-/I/ vowel contrast. The pre-training test session compared Greek and English speakers on their use of spectral and duration cues while the post-training test session explored whether vowel perception and cue weighting were significantly affected by training.

Pre-training test results showed that while English native speakers used spectral and duration cues as primary and secondary cues respectively, Greek learners of L2 English used duration as primary cues as evidenced by their impaired performance for those tasks where identification and discrimination relied on spectral cues (Giannakopoulou et al., 2013). Post-training results revealed that in the condition where training used natural vowel stimuli it appears that Greek L2 speakers of English were likely to weight duration cues as primary and spectral cues as secondary whereas in the condition where training used modified duration stimuli performance was improved for the modified duration stimuli tasks and also for the natural duration stimuli in the post-training test.

In the present study, twenty native English listeners transcribed the production samples and rated them for accuracy and clarity. The design for both the orthographic transcription task and the goodness rating task was a within-subjects design. The independent variables include time (pre- versus post-training), item type (word or sentence type), and speaker (speaker 1-8). Among item types, words included, for example, 'sit' and 'seat', while sentences included, for example, 'Please take a seat' and 'Sit down please'. The dependent variables include intelligibility as measured by the orthographic transcription task and clarity as measured by the goodness rating. Judged as 'correct' in the transcription task were only transcribed words that were identified completely correctly. This was done so as not to be unclear about whether near hits resulted because of typing errors or because listeners actually did not recognise the presented word. A three-way within-subjects ANOVA (time, speaker, item) was used to analyse the transcription of the words and sentences that were produced before and after the perceptual training by native Greek speakers as learners of L2 English.

It was found that there was a higher accuracy in the transcription of samples at word and sentence level after training compared to baseline accuracy indicating that the intelligibility of L2 learners' speech samples were rated to be higher after the high-variability training than before. Moreover, a significant interaction between time and speaker indicated that speakers' average clarity score was higher after training than before. This provides strong evidence that high-variability training can enhance vowel intelligibility at both word and sentence level for L2 learners' perception and their production of L2 speech (Figures 1 and 2).

Figure 1: Transcription accuracy of Greek-speakers' speech samples by native speakers at word level. Error bars show +/- 1 standard error from the mean.

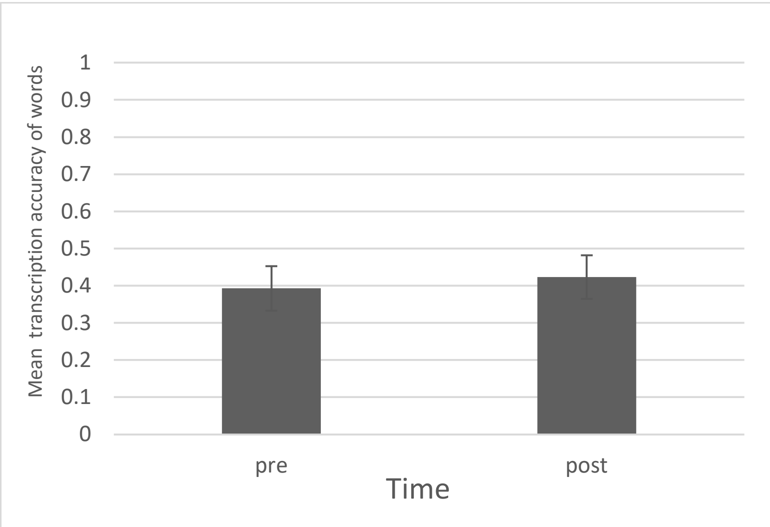
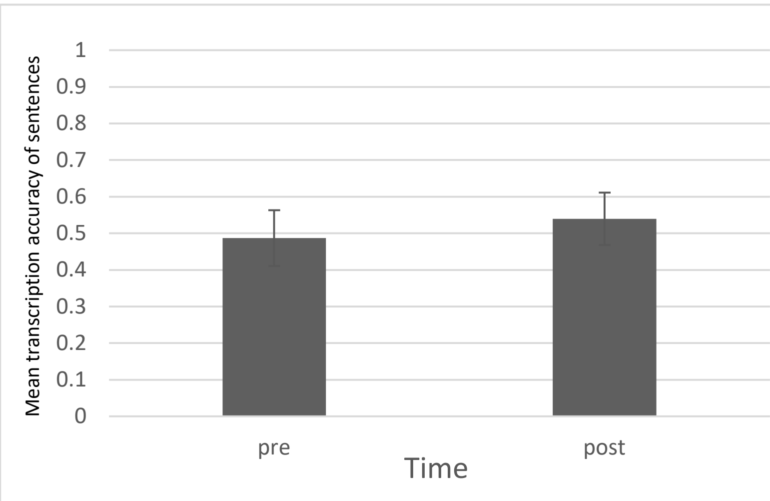


Figure 2: Transcription accuracy of Greek speakers' speech samples by native speakers at sentence level. Error bars show +/- 1 standard error from the mean.



Specifically, it could be suggested that high-variability phonetic training where particular cues are removed and attention is forced on relevant cues could possibly help L2 learners perceptually rearrange cues that were initially perceived as secondary into primary cues, and turn their perceptual benefit of enhancing their identification and discrimination of non-native speech contrasts into a tangible benefit when producing L2 speech. Implications of the findings are discussed in terms of theories concerning the link between speech perception and production, and the Speech Learning Model by Flege (1995).

References

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101 (4): 2299-2310. https://doi.org/10.1121/1.418276.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In S*peech perception and linguistic experience: Theoretical and methodological issues, edited by* W. Strange, 229-273. Timonium, MD: York Press.

Giannakopoulou, A. (2012). Plasticity in second language (L2) learning: perception of L2

phonemes by native Greek speakers of English. PhD dissertation, Brunel University. Retrieved from: http://bura.brunel.ac.uk/bitstream/2438/6592/1/FullTextThesis.pdf.

Giannakopoulou, A., Uther, M., and Ylinen, S. (2013). Enhanced plasticity in spoken

language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy* 29 (2) 201–218. https://doi.org/10.1177/0265659012467473.

Purcell, E. & Suter, R. (1980). Predictors of pronunciation accuracy: A re-examination. Language Learning, 30, 271-287.

Varonis, E., & Gass, S. (1982). The comprehensibility of non-native speech. Studies in Second Language Acquisition, 4, 114-136.

Age-specific effects of lexical-semantic networks on word production across the lifespan

Giulia Krethlow[1], Raphaël Fargier[2], Marina Laganaro[1]
[1] Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland
[2] Aix-Marseille University, Laboratoire Parole et Langage, France
Giulia.Krethlow@unige.ch

Introduction

Word production is an every-day activity that allows a person to express concepts by transforming them into lexical forms. This task can be influenced by different psycholinguistic factors as well as by the age of the speakers. Studies usually show lower performance in elderly and children when compared to young adults (D'Amico et al., 2001; Laganaro et al., 2015) in the picture naming task, in term of accuracy or production speed.

Among variables which are likely to underpin age-related changes in word production (Morrison et al., 2002; Taylor & Burke, 2002; Newman & German, 2005; Gordon & Cheimariou, 2013; Britt et al., 2016) changes in lexical-semantic processes may constitute a key point. There is however no unique consensus on which psycho-linguistic variable has the largest influence on word production and on age-related changes in picture naming (Thompson-Schill et al., 1998; Plaut & Booth, 2000; Lucas, 2000; Perea & Rosa, 2002; Ferrand & New, 2003; Hutchison, 2003). Semantic association has long been recognized as crucial for determining the conformation of semantic representations (Nelson et al., 2000; McRae et al., 2012). Nevertheless, the effects of lexical-semantic factors on word processing are usually based on studies enrolling young adult cohorts. To elucidate the role of age-specific semantic association networks on referential word production over the lifespan, in the current study we investigate whether the lexical-semantic organization predicts the speed of word production at different ages, from school-age children to older adults.

Methods

Collection of associative norms:
120 French native speakers aged from 10 to 80 years-old and divided into 6 age-groups (10-13; 16-18; 20-30; 40-50; 58-68; 69-80), participated in this data collection. Subjects had 10 seconds to give all associate words that came to mind for each of the 204 stimuli that the experimenter read them aloud (lexical-semantic fluency task). Two measures reflecting the lexical-semantic network of speakers have been computed from free associates given for each cue word and for each age-group: Available Richness of the mental lexicon (i.e. mean number of free associates given) and lexical-semantic Network Prototypicality.

Picture naming task:
120 other French native speakers aged from 10 to 80 years-old and divided into the same 6 age-groups (10-13; 16-18; 20-30; 40-50; 58-68; 69-80), participated in a picture naming task involving 120 pictures and their corresponding modal names for which classical psycholinguistic variables were available from two French databases (Alario & Ferrand, 1999; Bonin et al., 2003) in addition to the age-related associative norms collected for this study. Subjects were asked to overtly produce the word corresponding to the pictures. Reaction Times (RT) and Accuracy were calculated for each trial and entered in mixed models with the computed lexical-semantic variables and classical psycholinguistic variables as predictors.

Results

Available Richness and Prototypicality of the lexical-semantic association network change across ages. Available Richness displayed a U-shaped curve across the lifespan and was lower for children (10-13 years old) and elderly (69-80 years old) as compared to adults (40-50 years old) and younger adults (20-30 years old), whereas adolescents (16-18 years) and old adults (58-68 years old) showed intermediate results (main effect of Age, F (5,970) = 378.39, p < .001). Network prototypicality showed a relatively different pattern with highest prototypicality in old adults (main effect of Age, F (5,970) = 46.325, p < .001).

Then, mixed-effect regression models indicated that age-specific Available Richness (F(1, 10281) = 13.64, p<0.001) and age-specific Network Prototypicality (F(1, 10446) = 3.98, p<0.05) modulate production latencies. More crucially, results indicated that age-specific semantic association norms predict word production better than classical measures usually collected in young adults (McRae et al., 2005) (i.e. semantic associations or semantic features).

The effects revealed that a richer and more prototypical semantic network across subjects from a given age-group was associated with faster word production speed.

Discussion

The lexical-semantic network evolves in terms of Available Richness and Prototypicality, depending on age-groups. In fact, results show that the evolution in the lexical-semantic network, depending on age of the speaker, influences word production latencies in picture naming tasks.

The current results indicate that age-specific semantic organization is crucial to predict lexical-semantic behaviours across the lifespan. We may therefore underline the need of age specific norms rather than data collected only on a sample of young adults in order to better predict words processing at different ages across the lifespan. Finally, these results also provide cues to the understanding of the lexical-semantic properties of the mental lexicon and to lexical selection in referential tasks.

References

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531-552.

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 158-167.

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct Effects of Lexical and Semantic Competition during Picture Naming in Younger Adults, Older Adults, and People with Aphasia. *Frontiers in Psychology,* 7, 813.

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development, 2*(1), 71-105.

Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. *Mental lexicon: Some words to talk about words*, 25-43.

Gordon, J. K., & Cheimariou, S. (2013). Semantic interference in a randomized naming task: Effects of age, order, and category. *Cognitive neuropsychology,30(7-8),* 476-494.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813.

Laganaro, M., Tzieropoulos, H., Frauenfelder, U. H., Zesiger, P., 2015. Functional and time-course changes in single word production from childhood to adulthood. *NeuroImage, 111,* 204–214.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618-630.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547-559.

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations in Adolescents and Young Adults: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The Adolescent Brain: Learning, Reasoning, and Decision Making* (pp. 39-66). Washington, DC: APA.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology, 14(4),* 435-459.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & cognition*, *28*(6), 887-899.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech*, *48*(2), 123-156.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological research*, *66*(3), 180-194.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological review*, *107*(4), 786.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and aging*, *17*(4), 662.

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language*, *38*(4), 440-458.

Are bilingual children better, worse, or the same as monolingual children at spoken word recognition of foreign-accented speech?

Susannah V. Levi[1] & Nandita Karthikeyan[1,2]
[1]Department of Communicative Sciences and Disorders, New York University, USA
[2]Department of Psychology, University of Bath, UK
svlevi@nyu.edu

Introduction

The past two decades have seen an increased interest in how bilingualism affects cognitive and social processing. In particular, these studies have explored the existence of a bilingual advantage – better processing for individuals who are bilingual compared to those who are monolingual. In terms of executive function, several studies have found a bilingual advantage across the lifespan where bilinguals respond faster and are better at inhibiting irrelevant information or shifting to a new task (Bialystok, 1999; Carlson & Meltzoff, 2008; Costa, Hernández, & Sebastián-Gallés, 2008; Gold, Kim, Johnson, Kryscio, & Smith, 2013). Studies have also found a bilingual advantage for social processing, for example in tasks where children must alter their visual perspective to view a scene from the orientation of another party (Fan, Liberman, Keysar, & Kinzler, 2015; Greenberg, Bellana, & Bialystok, 2013). Research in the domain of speech perception has also found that bilingual children are better than monolingual children at processing information about the speaker (identifying the speaker or discriminating two different speakers) (Levi, 2018), a skill that could be interpreted as the social aspect of speech perception. It is important to note that the existence of a bilingual advantage, especially in the domain of executive function, has also been widely criticised (de Bruin, Treccani, Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2014). In the current study, we examine differences in spoken word recognition between bilingual and monolingual children. We use foreign-accented words as the target stimuli, because it may be the case that bilingual children have more experience with foreign-accented speech in general.

Methods

Participants

Thirty-nine children, ages 6;8–11;9, participated in the study. These participants were drawn from a larger study with additional tasks not discussed here. All children were English-dominant and were living in New York City at the time of testing. Children completed the Core Language subtests of the CELF-IV (Semel, Wiig, & Secord, 2003), which provides a measure of language ability, the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), which provides a measure of receptive vocabulary, and a hearing screening. One child was missing PPVT data, one was missing Core Language data, and one was missing both. Two different analyses were conducted based on different categorisations of bilingualism incorporating a broad definition versus a narrow definition. In the first categorisation, the same criteria were used as in Levi (2018), where children were monolingual if the parent report indicated that they only spoke and understood English and if there was no person living in the home or other caregiver who spoke another language. Children were classified in the bilingual category if the parent report indicated that they spoke or understood another language or if there was someone living in the home who spoke another language. Independent-samples t-tests confirmed no group differences for age ($p = .933$), Core Language ($p = .301$), or PPVT ($p = .903$). The second, narrow categorisation children as monolingual by the same criteria, but children were only categorised as bilingual if the parent report indicated that the child spoke or understood another language, which reduced the sample size in this group to eight. Independent-samples t-tests confirmed no group differences for age ($p = .665$), Core Language ($p = .479$), or PPVT ($p = .685$). Importantly, none of the children who were categorised as bilingual in either case had been exposed to German (the native language of the talkers, see next section). As such, any differences between the two groups cannot be attributed to experience with this particular language or foreign accent. Information about the children is provided in Table 1.

Table 1. *Group information.*

Categorisation	Group	Sample	Age in months	Language Ability	Receptive Vocabulary
Broad	Bilingual	17	111	107	108
	Monolingual	22	111	102	109
Narrow	Bilingual	8	108	106	106
	Monolingual	22	111	102	109

Materials and Procedure

Monosyllabic consonant-vowel-consonant English words were produced by six female German-English bilingual talkers. Words were evenly divided into two sets: high familiarity (e.g., 'hair') or low familiarity (e.g., 'loom'). In the spoken word recognition task, children heard 48 target words with a signal-to-noise ratio (SNR) of +5dB and were asked to repeat the word they heard aloud. This SNR was selected after pilot testing to ensure that children did not perform at ceiling or floor on the task. Words were presented binaurally over Sennheiser HD-280 circumaural headphones. Children's responses were coded at the time of testing, and also recorded and scored by a second coder offline. All disagreements were transcribed by a third coder. Prior to the experimental stimuli, children heard eight words that were spoken by a different speaker, but mixed at the same SNR to familiarise them with the task. For the purpose of this analysis, the target words were coded as whole word correct or incorrect.

Results

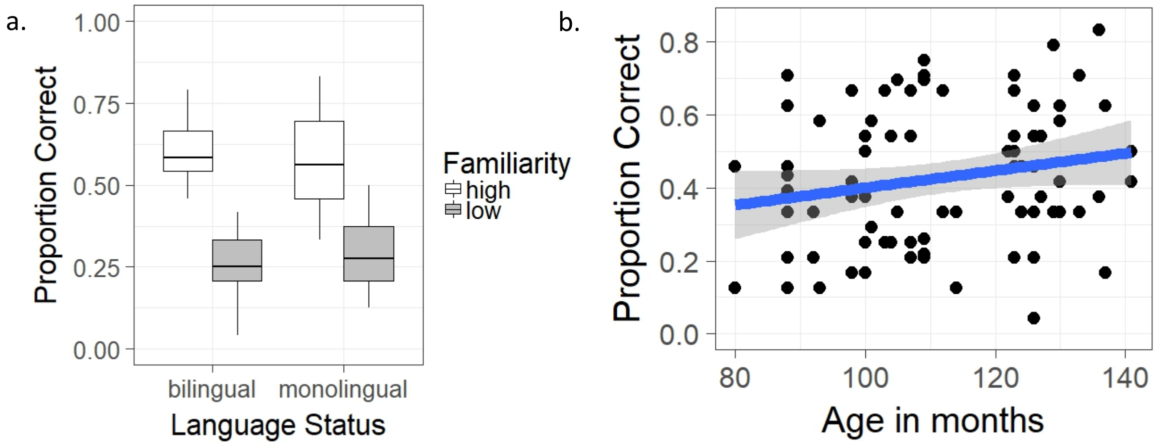
A logistic mixed-effects model was fit to the word recognition data and included fixed effects for group (bilingual, monolingual), word familiarity (high familiarity, low familiarity), age in months, and all two-way interactions. The two categorical predictors were sum-coded and the continuous predictor was scaled and centered. The model also included random slopes for talker by subject and random intercepts for talker and target word.

The first analysis (where the bilingual group included children with exposure to a second language in the home) revealed the expected effect of word familiarity (estimate = 0.858, standard error = 0.119, z-value = 7.19, $p < 0.001$), where children were more accurate at recognising high compared to low familiarity words. These results are visualised in Figure 1a below. The model also revealed an effect of age (estimate = 0.253, standard error = 0.079, z-value = 3.19, $p = 0.001$) with older children performing better than younger children, as shown in Figure 1b. The effect of group did not reach significance (estimate = -0.030, standard error = 0.076, z-value = -0.39, $p = 0.690$). No interactions reached significance (all $p > .118$). The second analysis (where the bilingual group only included children who spoke or understood a second language) revealed a similar pattern of results: High familiar words were recognised more accurately than low familiarity words (estimate = 0.957, standard error = 0.133, z-value = 7.15, $p < 0.001$) and older children performed better than younger children (estimate = 0.282, standard error = 0.112, z-value = 2.51, $p = 0.011$). The effect of group did not reach significance (estimate = -0.107, standard error = 0.109, z-value = -0.97, $p = 0.327$). No interactions reached significance (all $p > .118$).

Discussion

This study does not provide definitive evidence for a bilingual advantage (or disadvantage). One hypothesis was that bilingual children would be more familiar with listening to foreign-accented speech, presumably due to hearing other family or community members who are bilingual. However, no benefit of bilingualism was found. The current study was not designed to assess the actual amount of foreign-accented input, thus it is not possible to determine whether children in the bilingual group – regardless of how that group was defined – had more exposure to this type of speech than the monolingual groups. Indeed, one of the confounds may be that the children in the current study were living in a multilingual city and thus there may have been no major differences in exposure between the groups. Unfortunately, the small sample size in the current study makes it difficult to draw strong conclusions about this null result. It is important to note that the groups were well matched in terms of both receptive vocabulary and overall language ability.

Figure 1*. (a) Average proportion correct of whole words for bilingual and monolingual children split by high familiarity (white boxes) and low familiarity (grey boxes) for analysis 1. (b) Scatter plot of proportion correct by age in months for analysis 1.*



References

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636-644.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282-298. doi:10.1111/j.1467-7687.2008.00675.x

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59-86.

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science, 26*(1), 99-107. doi:10.1177/0956797614557866

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: NCS Pearson, Inc.

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The Exposure Advantage: Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological Science*. doi:10.1177/0956797615574699

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience, 33*(2), 387-396. doi:10.1523/jneurosci.3837-12.2013

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development, 28*(1), 41-50. doi:http://dx.doi.org/10.1016/j.cogdev.2012.10.002

Levi, S. V. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition, 21*(3), 523-536. doi:10.1017/S1366728917000153

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232-258. doi:http://dx.doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*(6), 615-639. doi:10.1080/20445911.2014.944914

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, fourth edition (CELF-4)*. Toronto, CA: The Psychological Corporation/A Harcourt Assessment Company.

The effect of dynamic acoustic cues on age classification

Luciana Albuquerque[1,2], António Teixeira[1,3], Catarina Oliveira[1,4], Daniela Figueiredo[2,4]
[1] Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal
[2] Center of Health technology and Services Research, University of Aveiro, Aveiro, Portugal
[3] Department of Electronics Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal, [4] School of Health Science, University of Aveiro, Aveiro, Portugal
lucianapereira@ua.pt

Introduction*:* Despite speech being inherently dynamic (Yuan, 2013), vowels have been traditionally characterized by static cues such as first two formants (F1 and F2) obtained at vowel midpoint. Nevertheless, several studies have reported that dynamic cues, in particular Vowel-Inherent Spectral Changes (VISC), contain essential information, not only for diphthongs but also for monophthong vowels (Almurashi et al., 2019). Also, improvements in classification of vowels by adding dynamic cues has been reported for some languages (e.g., Almurashi et al., 2019; Elvin et al., 2016; Yuan, 2013). Acoustic cues are not only used to decode the intended phoneme, speaker dependent information such as gender, age, and dialect is also extracted by humans (Yuan, 2013). All previous studies on European Portuguese (EP) vowels rely on single acoustic measurements of formant frequencies at the middle of the vowel.

Related work: Usually, studies that include dynamic cues of vowel extract formants at several time instants – typically from 3 to 30 – and, from the extracted trajectories derived parameters are obtained, being the Discrete Cosine Transform (DCT) coefficients the most used (Elvin et al., 2016; Williams & Escudero, 2014). Also, some studies use the formant Trajectory Length (TL) and Spectral Centroids (Jacewicz et al., 2011a, 2011b). Moreover, some studies use these dynamic cues in a classification task. For example, Themistocleous (2017) applies several classifiers to predict dialect, stress and vowel. The set of methods explored is vast, including Discriminant Analysis, Decision Trees, Neural Networks, and Support Vector Machine (SVM).

Objective: The age effects in speech production are an important issue, and the main goal of a research line that we have been pursuing for several years. For that, after studying the age effects in static cues (formants and f0) (Albuquerque et al., 2019, 2014; Oliveira et al., 2012), the determination of the potential usefulness of the dynamic acoustic proprieties of EP vowels as carriers of age classification is a natural objective. But before that, we need to investigate if there is a significant effect that justifies further research.

Method: The existence of a significant age effect in vowels' dynamics was indirectly studied by measuring the impact of using dynamic information in two classification tasks: classification into age groups; classification of vowel for different age groups. As a first approach, for age groups classification, a binary decision senior/non-senior was selected, with cut-off at 65 years. Finally, the vowel classification task was used to complement the first.

Speech Corpus: The sample consists of 112 healthy native Portuguese speakers (56 men and 56 women), from the central region of Portugal (without voice or speech problems), aged between 35 and 97. The corpus consists of 36 disyllabic words, with the EP vowels [i], [e], [↔], [a], [o], [↵] and [u] in stressed position and the vowels [ɨ] and [ɐ] in unstressed position, in stop and fricative consonant context. The stimuli were embedded in a carrier sentence. Each participant produced 12 repetitions of each vowel, in a total of 108 productions by speaker (112 participants x 36 words x 3 repetitions = 12096 recordings). The recorded data was first automatically segmented at phoneme level using WebMAUS and the vowel boundaries were manually check in Praat. For more details see Albuquerque (2019).

Acoustic Features: The Burg-LPC algorithm was used to compile values for F1, F2 and F3. A procedure was applied to optimize the formant ceiling for a certain vowel of a certain speaker (details in Albuquerque (2019) and Escudero (2009)). The “optimal ceiling” was used to extract the vowel formant frequencies at 35 equally spaced time points within the central 60% of each token (adapted from Williams & Escudero (2014)). Vowel duration was also measured. In order to characterize F1, F2, and F3 trajectories, each set of 35 formant values was transformed using the DCT. Each coefficient characterizes an aspect of a formant trajectory’s shape based on a cosine (Williams et al., 2014). To assess the extent of formant movement in a vowel, based on Jacewicz et al. (2011a, 2011b), the TL was also calculated. The obtained values, except duration, were normalised using a z-score to reduce the vowel and gender effects.

Classifiers: So far, experiments were conducted with the following classifiers: Linear Discriminant Analysis (LDA)(e.g., Themistocleous, 2017), Classification and Regression Trees (CART) (e.g., Schötz, 2003) and Neural Networks (NN). Experiments started by LDA and concentrated mainly in CARTs due to their advantage in terms of interpretability of results. Some initial experiments were performed with Deep Neural Networks (DNN). LDA and CART experiments were performed in Matlab (version R2018b) using 10-fold cross-validation. For DNN 80% of the samples were used for training and 20% for test.

Results: The main results of the classification experiments are summarized below. Additionally, more detailed results of ongoing experiments will be added.

LDA: The binary classification into age groups was attempted with different sets of features, starting with only static information for F1, F2 and F3 and gradually adding the other DCT coefficients (C1 to C3) as well as TL and duration. Very low sensitivity values (0.22) were obtained using only static cues. Inclusion of dynamic cues improves sensitivity, being the best results of 0.26 (without duration) and 0.45 (using duration). For all experiments, specificity was above 0.7. Generally, sensitivity is lower than the reported for dialect classification (Themistocleous, 2017), pointing to age group separation to be a more difficult task.

CART: The binary classification of age groups showed (see Table 1): (1) higher error rate when using only static cues; (2) error reduction by including additional DCT coefficients; (3) error reduction when adding TL; (4) great duration impact in error reduction. Predictor importance estimates were also analysed and the top 3 are included in Table 1. In the second task, vowel classification results, for both age groups, also showed higher error values when using only static cues (approx. 0.8 with only static versus approx. 0.40 with all DCT coefficients and no duration). With TL and duration added the error attains the minimum values, approx. 0.35. The best 3 predictors when using a mix of static and dynamic cues only include the dynamic ones, being common DCT coefficient C1 for F1, F2 and F3.

DNN: Only the binary age group classification was performed. A simple 2 hidden layers network was used. Hidden layers dimension was obtained by a growing process aiming at obtaining a high sensitivity for the training set. This process resulted in the selection of 2 hidden layers of 40 units each. After the definition of the network topology, one was trained with all static and dynamic cues as input. Training was stopped when a predefined error was reached (0.1), which occurred at iteration 362. A similar network was trained using only static cues as input, being the process stopped at the same number of iterations (error was at much higher values than in the first network). With the first network (including dynamics) sensitivity was 74.2 for training and 33.6 for test, values much higher than the 5.3 (training) and 5.5 (test) obtained with only static information.

Conclusion: This study shows that dynamic measurements of F1-F3 result in better classification performance of these two age groups, somehow supporting the hypothesis of relevant information existence regarding age in the acoustical dynamic cues. Additionally, the importance of duration identified in our previous related work (Albuquerque et al., 2019), was reconfirmed as an important predictor of age. Further investigation is needed regarding the age effects.

Table 1: *Results for the CART prediction of age group for each set of parameters. (Worst and best result in boldface). Different colours are used for each parameter in the predictor's columns*

	Parameters	DCT coefficients	Error	1st predictor	2nd predictor	3rd predictor
Static	F1, F2	C0	**0.483**	F1_C0	F2_C0	_
Static	F1, F2, F3	C0	0.478	F1_C0	F3_C0	F2_C0
Dynamic	F1, F2, F3	C0, C1	0.460	F1_C0	F2_C1	F1_C1
Dynamic	F1, F2, F3	C0, C1, C2	0.448	F1_C0	F2_C0	F2_C2
Dynamic	F1, F2, F3	C0, C1, C2, C3	0.440	F1_C0	F1_C1	F2_C0
Dynamic	F1, F2, F3; TL	C0, C1, C2, C3	**0.438**	F1_C0	F2_C1	F3_C1
	F1, F2, F3; TL; duration	C0, C1, C2, C3	0.406	duration	F1_C0	F2_C0

Acknowledgements

This research was financially supported by the project Vox Senes POCI-01-0145-FEDER-03082 (funded by FEDER, through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI), and by national funds (OE), through FCT/MCTES), by the grant SFRH/BD/115381/2016, and by IEETA (UID/CEC/00127/2019). We are very grateful to the institutions for having made possible the data collection, and also to all the adults who contributed as speakers.

References

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

Albuquerque, L., Oliveira, C., Teixeira, A., Sa-Couto, P., Freitas, J., & Dias, M. S. (2014). Impact of age in the production of European Portuguese vowels. *INTERSPEECH*, 940–944.

Almurashi, W., Al-Tamimi, J., & Khattab, G. (2019). Static and dynamic cues in vowel production in Hijazi Arabic. *19th ICPhS*, 3468–3472.

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *J. Acoust. Soc. Am.*, *140*(1), 576–581.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *J. Acoust. Soc. Am.*, *126*(3), 1379–1393.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011a). Cross-generational vowel change in American English. *Language Variation and Change*, *23*(1), 45–86.

Jacewicz, E., Fox, R. A., & Salmons, J. (2011b). Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics*, *39*(4), 683–693.

Oliveira, C., Cunha, M. M., Silva, S., Teixeira, A., & Sa-Couto, P. (2012). Acoustic analysis of European Portuguese oral vowels produced by children. *IberSPEECH*, *328*, 129–138.

Schötz, S. (2003). Automatic Estimation of Speaker Age using CART. *Course in Speech Recognition*, 1–8.

Themistocleous, C. (2017). Dialect classification using vowel acoustic parameters. *Speech Communication*, *92*, 13–22.

Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *J. Acoust. Soc. Am.*, *136*(5), 2751–2761.

Yuan, J. (2013). The Spectral Dynamics of Vowels in Mandarin Chinese. *INTERSPEECH*, 1193–1197.

Screening English and Arabic children for speech fluency and treating word-finding difficulty

Roa'a Alsulaiman, John Harris & Peter Howell
Division of Psychology and Language Sciences, University College London, United Kingdom
p.howell@ucl.ac.uk

Background

Andrews and Harris (1964) reported that the incidence of stuttering is 4.9% and Yairi and Ambrose (2005) confirmed that approximately 5% of pre-school age children exhibit episodes of stuttering. Such speech disfluencies need to be identified at an early age so that effective and appropriate intervention can take place before the problem becomes acute. This applies in countries like the UK and Saudi Arabia where both English and Arabic are spoken by some children. When children are not speaking their native language, their speech would include many instances of whole-word repetitions, which is an indication of word-finding difficulty (WFD). The latter is a type of communication difficulty that can occur in the speech of monolingual children, but it is particularly frequent in the speech of children who use English as an additional language (EAL) (Howell et al., 2017). Thus, it is important to have brief tests that can discriminate children with WFD from children who stutter (CWS) when they first enter school. CWS may then be referred for speech language therapy (SLT) to receive the right type of intervention.

Aims

The aims of this study included (1) separate children with speech dysfluencies from children with WFD using simple procedures that apply equally to English- and Arabic-speaking children; (2) treat WFD in monolingual English children and Arabic children with EAL. Non-word training materials were specifically designed to include patterns not used in the speaker's native language but that are required if that speaker is to use another targeted language (English or Arabic). The goal is to use them as interventions that train phonemic sequences that are exercised infrequently, or probably not at all, within the additional language used, either English or Arabic. For example, a native Arabic speaker learning English would be expected to struggle most on those phonotactic patterns not used in Arabic but required for English. It is hypothesized that phonological priming which involves repeated exposure to the unfamiliar sequences would increase familiarity with such structures and improve overall fluency. To support that, NWR performance should improve initially, which should generalize to lexical performance (less WFD, improve speech rate and fluency) and improve language skills.

Methods

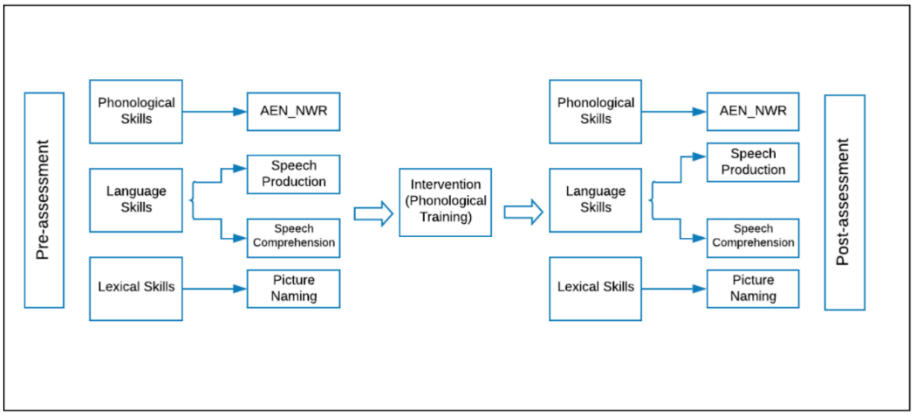
Participants

All children from reception classes and year 1 at King Fahad Academy in London whose parents consented to participation were tested. Twenty-six 4-6-year-old Arab-speaking children were assessed pre- and post-intervention. None of the children had hearing impairments. Language history information was obtained from the children and verified by teachers.

Procedure

Children were tested individually in a quiet room in 15-minute sessions. All sessions were audio recorded using a Sennheiser SC 660 USB ML wired headset and Audacity software for subsequent analysis. In the pre-test, a spontaneous speech sample of 200 syllables (minimum) was recorded and used to obtain measures of speech dysfluency according to Riley's (1994) Stuttering Severity Instrument (SSI-3) and word-finding difficulties. Characteristics of word-finding difficulty included unnecessary repetition of words and empty words that add no content or specificity to the message. Children's phonological skills were assessed based on their accuracy in repeating non-words of different syllable lengths from the Arabic-English non-word repetition (AEN_NWR) task. The AEN takes the phonotactic constraints of Arabic and English and generates materials that are phonologically well-formed and are appropriate to children who speak either of the two languages. Presenting children with those non-word materials can aid in identifying production difficulties and facilitate separating children who are fluent from those who have WFD and from children who are dysfluent (Howell et al., 2017). Additional language tests included a narrative comprehension task to measure a child’s comprehension abilities after being narrated a story based on a picture they were presented with and had to answer questions about. A picture-naming task was also conducted where children named pictures of English words with selected phonological structures that corresponded to those in the intervention materials. Pictures were selected from CBeebies; a sub corpus of the SUBTLEX (Van Heuven at al., 2014) British English corpora. CBeebies was built using children spoken language and has a word frequency scale ranging from 1 to 6 to indicate the number of times every word has appeared in the corpus (1=low frequency and 6=high frequency). Only words with a Zip scale higher than 4.60 were selected to minimize the number of errors due to not knowing the picture. Reaction time (RT) was calculated from the onset of the picture to the participant's voice onset. During the intervention, children repeated the English set of non-words (which served as primes) that were designed to expose them to English phonotactic structures that are challenging for them because they represent patterns of English not found in Arabic. The stimuli were presented starting with the least phonologically complex materials, like vowels and consonant phonemes that occur in monosyllables and progressed through complex clusters word initially and word finally. Children were exposed to all intervention materials and no feedback was given. Children were told prior to the non-word tests (screening and intervention) that they would hear made-up words that they should repeat. Measures were taken pre and post the intervention (see Figure 1) to establish any improvements as a result of the intervention. All stimuli were pre-recorded by a native British male speaker who is phonetically trained in English and Arabic.

Figure 1. *Study design outlining the different tasks during pre and post tests and the intervention*



Outcome and results

Three repeated-measures ANOVAs were conducted with one within group factor (two levels, before and after the intervention) using RT, picture-naming accuracy or NWR scores as the main dependent variables. As expected, a main effect of training was found, indicating that RT significantly decreased across pre and post assessment sessions ($F(1,25) = 9.719$, $p<.000$). In the second analysis, the mean picture-naming accuracy increased significantly after the treatment ($F\ (1, 25) = 8.77$, $p=.007$). The third analysis also determined that NWR scores increases significantly across pre and post assessment sessions ($F\ (1, 24) = 14.298$, $p<0.000$). A slight reduction was found in the rate of whole-word repetition and the percentage of speech syllables, but these were not significant.

Conclusion and Implications

The results of this study suggest that phonological priming has promise for treating WFD and associated fluency problems for Arabic children with EAL. The intervention procedure is fast and easy to administer by schoolteachers. The improvements relative to baseline occurred immediately post-intervention but follow-up tests are needed to examine whether there are sustained improvements. Future work should consider the recruitment of a control group to show whether any training effects are due to teaching input rather than the intervention. In ongoing work, we are testing a larger number of children with diverse language backgrounds including monolingual English and English children with Arabic as additional language. Differences between the groups are to be examined and reported at the conference.

Acknowledgment

King Saud University supports Roa'a AlSulamainan's PhD at University College London. Their support is gratefully acknowledged.

References

Andrews, G., & Harris, M. (1964). The syndrome of stuttering.

Howell, P., Tang, K., Tuomainen, O., Chan, S. K., Beltran, K., Mirawdeli, A., & Harris, J. (2017). Identification of fluency and word-finding difficulty in samples of children with diverse language backgrounds. *International Journal of Language & Communication Disorders*, 52(5), 595-611.

Riley, G. (1994). Stuttering severity instrument for young children (SSI-3). *Stuttering Severity Instrument for Young Children (SSI-3)*.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176-1190.

Yairi, E., Ambrose, N. G., Paden, E. P., & Watkins, R. V. (2005). *Early childhood stuttering for clinicians by clinicians*. Austin, TX: Pro-ed.

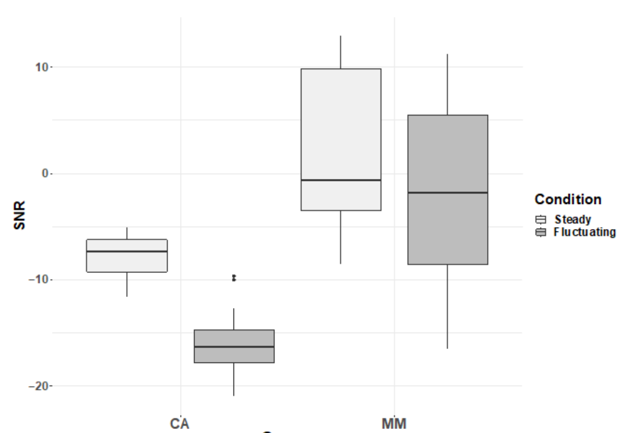
Neural Processing and Perception of Speech in Children with Mild to Moderate Sensorineural Hearing Loss

Axelle Calcus[1], Stuart Rosen[2], Lorna Halliday[3]
[1] Laboratoire des Systèmes Perceptifs, Ecole Normale Supérieure, Paris, France [2] Department of Speech, Hearing and Phonetic Sciences, University College London, UK [3] MRC Cognition and Brain Sciences Unit, University of Cambridge, UK
axelle.calcus@ens.fr

Mild (21-40 dB HL) or moderate (41-70 dB HL) sensorineural hearing loss (MMHL) can lead to persistent changes to the cortical processing of speech sounds. This was evidenced in a study conducted on 46, 8- to 16-year old children with MMHL and 44 normally-hearing (NH) age-matched controls (Calcus, Tuomainen, Campos, Rosen, & Halliday, 2019). While present in younger children with MMHL, there was no significant mismatch negativity (MMN) in older children with MMHL. However, to date no studies have examined speech processing at the subcortical level in children with MMHL, yet this is known to be linked to speech perception in noise (SIN) in NH children. Moreover, the effects of amplification on the neural encoding of speech remain poorly understood, with previous data suggesting a benefit at the subcortical but not the cortical level (Anderson et al., 2013; Billings, Tremblay, Souza, & Binns, 2007).

Here, we aimed to investigate (1) the subcortical and cortical processing of speech sounds in children with MMHL, (2) the relation with SIN, and (3) the effects of amplification on the neural processing of speech, for children with MMHL. Behavioural thresholds were measured at 70 dB SPL for consonant identification in both steady and fluctuating noise. Subcortical and cortical EEG activity evoked by speech stimuli were simultaneously recorded in 18, 8- to 16-year old children with MMHL and 15 age-matched NH controls. The frequency-following-response (FFR) and MMN were used as indices of speech processing at the subcortical and cortical levels, respectively. Speech stimuli were presented in an oddball paradigm. The FFR was computed in response to the standard stimulus (/ba/), while the MMN was computed as the differential wave evoked by the deviant (/da/) compared to the standard. For the MMHL group, stimuli were presented both unamplified (70 dB SPL), and with a frequency-specific gain (without compression) based on their individual audiograms.

Figure 1: *Boxplot of speech reception thresholds (SRT, in dB SNR) for age-matched NH controls (CA) and children with MMHL (MM), presented with steady (light grey) and fluctuating (dark grey) background noise.*



Behavioural thresholds were poorer for children with MMHL than NH controls, whatever the background noise (Fig 1). Children with MMHL did not show masking release in the fluctuating, as compared to the steady background noise, yet NH controls did. At the cortical level, MMN was significantly smaller, but not later, in children with MMHL than NH controls. There was no significant benefit of amplification on the MMN amplitude (Fig 2). At the subcortical level, responses were significantly smaller in children with MMHL than controls. Importantly, the subcortical responses were significantly larger in the amplified than the unamplified condition for children with MMHL, reaching an amplitude that was comparable to that of NH controls (Fig 3). However, the correlations between speech perception in noise and subcortical measures of speech encoding was not significant for either the MMHL or the NH group.

Figure 2: *Grand average of the cortical responses recorded in NH controls (left), in MM children when unamplified (middle) and when amplified (right). The light trace is obtained in response to the standard sound; the dark trace is obtained in response to the deviant sound.*

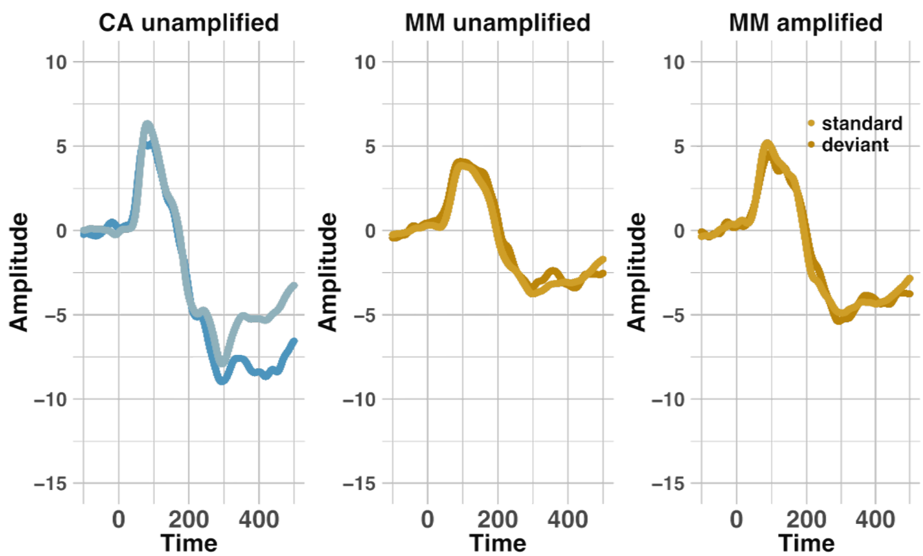
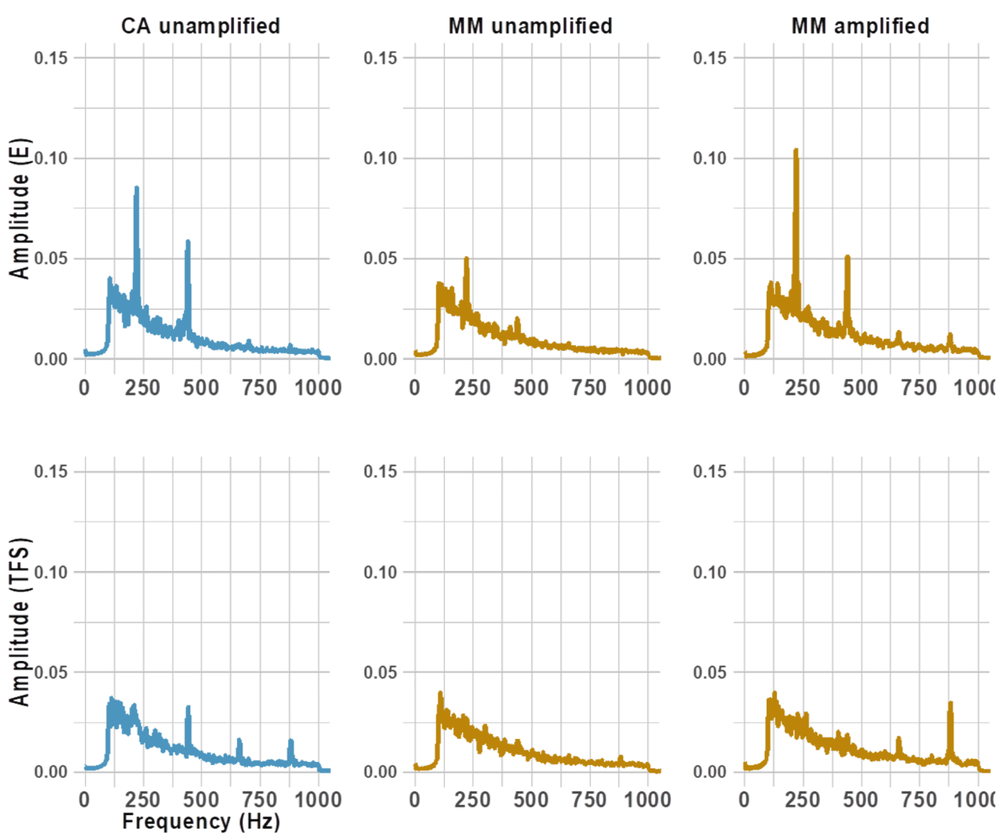


Figure 3: *Grand average spectra of the subcortical responses recorded in NH controls (left), in MM children when unamplified (middle) and when amplified (right).*



The neural processing of unamplified speech may be impaired at both subcortical and cortical levels in children with MMHL. Moreover, amplification may benefit auditory processing at subcortical but not cortical levels in this group. We offer two alternative explanations for our findings: increasing multi-sensory integration at successive levels of the auditory system (Calvert, Hansen, Iversen, & Brammer, 2001), and/or later maturation of the auditory cortex compared to the inferior colliculus (Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008).

Acknowledgments

This research was supported by an Economic and Social Research Council (ESRC) First Grants Award (RES-061-25-0440) to LH and a European Union ITN grant (FP7-607139).

References

Anderson, S., Anderson, S., Parbery-Clark, A., Parbery-Clark, A., White-Schwoch, T., White-Schwoch, T., et al. (2013). Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America*, *133*(5), 3030–3038.

Billings, C. J., Tremblay, K. L., Souza, P. E., & Binns, M. A. (2007). Effects of Hearing Aid Amplification and Stimulus Intensity on Cortical Auditory Evoked Potentials. *Audiology and Neurotology*, *12*(4), 234–246.

Calcus, A., Tuomainen, O., Campos, A., Rosen, S., & Halliday, L. F. (2019). Functional brain alterations following mild-to-moderate sensorineural hearing loss in children. *eLife*, 1–37.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of Audio-Visual Integration Sites in Humans by Application of Electrophysiological Criteria to the BOLD Effect. *NeuroImage*, *14*(2), 427–438.

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research*, *236*(1-2), 61–79.

Cortical responses to speech and linguistic contrasts in children with typical and atypical language development

Ana Campos Espinoza [1], Outi Tuomainen[1], Stuart Rosen[1] & Lorna F. Halliday[1,2]
[1] Department of Speech, Hearing and Phonetic Sciences, University College London, UK
[2]MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

Previous research has shown that brain speech processing is influenced by the perceived language content and individual knowledge (Ding & Simon, 2014), with evidence of enhanced cortical discrimination for higher-level linguistic representations (Gansonre, Højlund, Leminen, Bailey & Shtyrov, 2018) and greater language abilities (Bent, Bradlow & Wright, 2006). During early childhood, important age-related changes in language skills are associated with language experience and brain structural/functional development (Skeide & Friederici, 2016). However, top-down linguistic influences on cortical speech perception responses may be affected in language-impaired individuals with normal auditory thresholds. Previous research has shown evidence of atypical auditory Event-Related Potential (ERP) responses in normal-hearing children affected by Developmental Language Disorder (DLD) when compared to age-matched controls with typical language development (TLD). For example, several studies using acoustic and speech sound contrasts have shown smaller or absent Mismatch Negativity (MMN) or Late Discriminative Negativity (LDN) cortical responses and longer latencies in DLD-affected children than in age-matched, TLD controls (Kujala, & Leminen, 2017; Näätänen, Sussman, Salisbury & Shafer, 2014). However, this has been contradicted by other findings of similar auditory discrimination between TLD-DLD groups (Kujala, & Leminen, 2017) and by the fact that most studies reporting atypical cortical responses in DLD children are consistent at a group but not at an individual level (Bishop, 2007). In addition, previous studies have usually focused on low-level auditory discrimination by using acoustic (e.g. non-speech sounds with frequency or duration changes) and phonemic contrasts (e.g. /ba/-/da/ syllables or vowel changes), without assessing the effects of higher-order lexical or semantic distinctions and potential top-down linguistic on speech perception.

This study investigates cortical discriminatory responses to speech contrasts involving multiple linguistic levels, in two groups of Spanish-speaking children aged 4.9 to 5.7-year-old. Both groups differed in their language developmental status: a group diagnosed with DLD (n=16, 6 female, mean age 5.2 years, range 4.9-5.7 years) and a group of TLD controls (n=11, 7 female, mean age 5.2 years, range 4.10-5.6 years), all recruited from the same preschool in Chile (special education and mainstream divisions). Electroencephalography (EEG, 32-channel) was recorded using a task-free Multifeature oddball paradigm (Niemitalo-Haapola, 2013) during auditory perception of CVC syllables sequences with initial phonemic changes. By changing the initial phoneme of the Spanish monosyllabic non-word /FUS/ (standard stimulus), we produced four deviant stimuli, manipulating the linguistic content at three processing levels: phonological (native vs non-native phonemes), lexical (word vs non-word) and semantic (content vs function word). Phonological contrasts involved changing from the standard stimulus (ST) to a non-native phoneme (/SH/) or to a Spanish native one (/H/), presented as the non-words /SHUS/ (Deviant 1) and /HUS/ (Deviant 2), respectively. Lexical/non-lexical comparisons involved native phonemic changes from the ST to monosyllables with and without lexical status, the Spanish words /TUS/ (Deviant 3) and /LUZ/ (Deviant 4) versus the non-word /HUS/. Finally, the semantic level involved native phonemic changes from the ST, towards a Spanish function word (/TUS/) versus a content word (/LUZ/). A total of 288 standards (50%) and 288 deviants (72 per type, 12.5% each) stimuli were alternately presented, in four blocks of 3 min 20 seconds duration. Participants were behaviourally screened to corroborate their normal-range hearing and non-verbal skills. In addition, their phonological awareness and speech perception skills were tested to determine correlation with cortical responses.

We hypothesised that cortical discrimination of speech contrasts will be associated with language typical/atypical developmental status (TDL/DLD), with better language skills facilitating the perception of acoustic differences involving higher-level linguistic representations. Using the auditory Mismatch Response (MMR) as cortical discrimination index, we expect significant between-group differences in speech responses and correlations between brain activity and language measures. Specifically, in the TLD group, we predict significantly larger MMR mean amplitude and shorter latency for change detection than for the DLD group. We also predict a significant "Group x Condition" interaction, this is, a larger effect in the TLD group for discrimination of i) native than non-native contrasts, ii) words than nonwords iii) content than function words. On the contrary, we expected these effects be reduced or absent in the DLD group. Preliminary EEG and behavioural results will be presented and discussed.

References

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 97-103. https://doi.org/10.1037/0096-1523.32.1.97

Bishop, D. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin*, 133(4), 651–672. https://doi.org/10.1037/0033-2909.133.4.651

Ding, N., & Simon, J. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Front Hum Neurosci*, 8, 311. https://doi.org/10.3389/fnhum.2014.00311

Gansonre, C., Højlund, A., Leminen, A., Bailey, C., & Shtyrov, Y. (2018). Task-free auditory EEG paradigm for probing multiple levels of speech processing in the brain. *Psychophysiol,* 55(11). doi: 10.1111/psyp.13216.

Kujala, T. & Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment: A review on mismatch negativity findings. *Developmental Cognitive Neuroscience*, 28, 65-75. doi: 10.1016/j.dcn.2017.10.005.

Näätänen, R., S. Sussman, E., Salisbury, D. & Shafer, V. (2014). Mismatch Negativity (MMN) as an Index of Cognitive Dysfunction. *Brain Topogr* 27, 451–466. https://doi.org/10.1007/s10548-014-0374-6

Niemitalo-Haapola, E., Lapinlampi, S., Kujala, T., Alku, P., Kujala, T., Suominen, K., & Jansson-Verkasalo, E. (2013). Linguistic multi-feature paradigm as an eligible measure of central auditory processing and novelty detection in 2-year-old children. *Cognitive Neuroscience*, 4(2), 99-106. doi: 10.1080/17588928.2013.781146

Skeide, M. & Friederici, A. (2016). The ontogeny of the cortical language network. *Nat Rev Neuroscience*, 17(5), 323-332. doi:10.1038/nrn.2016.23.

The production of French consonant sequences in typically-developing children and in people with aphasia

Faith Chiu [1] and Typhanie Prince [2]

[1] Department of Language and Linguistics, University of Essex

[2] University Training Center for Speech-Language Pathology, University of Nantes

Introduction

All models of language production postulates that speakers generate phonological forms of words out of sublexical components as opposed to retrieving them as units from the mental lexicon (Meyer & Belke, 2008). Phonological forms differentiate segments from metrical structure, and models vary in which the way segmental material is linked to metrical structure (syllables) and whether or not activation of segmental material is parallel in the syllable (e.g., Dell, 1986 compared to Levelt et al., 1999). Our study considers if speech errors and modifications made by typically-developing children during language acquisition are mirrored by those made by people with aphasia (PwA). We focus in particular on consonant sequences (also consonant clusters) which have been theoretically shown to involve interplay between segmental material and syllable structure.

Methods

Participants

Twenty monolingual French typcally-developing children (M_{age} = 2;34, 8 female) and twenty monolingual French people with vascular aphasia participated in this study (M_{age} = 64;10, 11 female). None of the children's parents reported their children to have any speech and hearing deficits, nor any neurological or motor control disorders. All but one PwA suffered from a left hemisphere stroke; the exception had a right cerebrovascular accident. Parts of standardised batteries (HDAE, Mazaux & Orgogozo, 1982 and MT-86, Nespoulous & al. 1986-1992) were administered by a speech therapist to PwA to obtain their cognitive and linguistic profiles; only those with phonological deficits were selected for the study. All but two PwA participated in the experiment 2 to 25 days post-stroke; the other two completed the task during speech therapy after at least 25 days. The breakdown of aphasia types by lesion were as follows: those resembling Broca's aphasia with an anterior damage (n = 7, where 4 of which also suffered from apraxia of speech), those with Wernicke's aphasia (n = 6), conduction aphasia (n = 4), and sensorial transcortical aphasia (n = 3). All participants were based in Nantes: children were tested in their local kindergarten; PwA participated in the task in the hospital Neurovascular unit. Another group of native French speakers were also tested as controls (n = 20).

Material and Procedure

PwA and children completed a picture-naming task containing a total of forty target tokens. These tokens were bisyllabic or trisyllabic and contained consonant sequences of two adjacent consonants at the beginning of the word (e.g. ***st**ylo 'pen'*), or the middle (*ca**sq**uette* 'cap') or the end (*ma**sq**ue 'mask'*, *ve**ste** 'jacket'*)**.** Participants were prompted by the experimenter when they were unable to retrieve a word from memory, after which they repeated the token. In this conference paper, only speech error data obtained by means of unprompted production were analysed (*i.e.* without any repetition prompt). All utterances were transcribed by a trained French phonetician on PHON (Rose & al., 2006), and were coded for transformations made to consonant sequences in the target token. These transformations could, for example, take the form of omitting the first or second of two consonants (deletion), and/or inserting a vowel to break up the two adjacent consonants in the sequence (epenthesis), and/or replacing one or both consonants with another consonant (substitution).

Preliminary Results

Data analysis is currently in progress.

Examples of modifications in speech error data

Deletion	***st**ylo*	/stilo/	[**t**ilo]
Epenthesis	*se**rp**ent*	/sɛʁpɑ̃/	[sɛʁ**ə**pɑ]
Substitution	*to**rt**ue*	/tɔʁty/	[**k**ɔʁ**k**y]

Main modification types

Across all word positions, both children and PwA produced speech errors on consonant sequences. In both groups, the predominant modification to phonological form was that of consonant deletion, resulting in the transformation of consonant sequences to simple onsets. Compared to children, people with aphasia performed a larger amount of consonant substitutions. In this early stage, we interpret that this difference is driven by PwA's possession of adult, fully-developed phonological processing. In substitutions, the choice to retain syllable positions over maintain integrity of segmental material could be an attempt to preserve metrical structure still existent in post-stroke lexical entries. Some further analysis of variation by aphasia-type also revealed that epenthesis was most common as a modification strategy only in people with exclusive Broca's aphasia (i.e. without co-occuring apraxia of speech). These PwA, unlike the other four groups (Broca's with apraxia, Wernicke's, conduction, and sensorial transcortical), tend to break up adjacent consonants within a word by inserting a vowel.

Variation of modification by consonant sequence position within a word

There is also variation in word position by group for recorded modifications. PwA transformed mainly word-initial consonant sequences. PwA with and without apraxia of speech produced initial sequences differently: deletion of the first consonant dominated speech errors by those with apraxia, while other transformation options were also observed for PwA without apraxia. It was only children who mostly struggled with consonant sequences in the middle of word; their speech errors for medial sequences involved transformations via deletion of either the first or both consonants in the sequence. All groups made transformations to word-final sequences via a mixture of transformation methods, predominantly deletion of either the first or second consonant and/or substitution.

Discussion

We interpret that transformations by PwA of consonant sequences at the beginning of the word to indicate complexity of such initial sequences in both motor planning and in phonological structure. The fact that PwA without apraxia of speech also struggled with such initial sequences and transformed them differentially demonstrates that, even when motor planning difficulties were accounted for, such initial sequences also present difficulty at the level of phonological representation. Children's difficulty with medial sequences could imply something about the nature of these sequences which makes them difficult to acquire. Yet, once these medial sequences are mastered, they remain relatively stable in phonological representation since PwA made few speech errors with consonant sequences in word-medial positions.

Our study demonstrated differential treatment of phonologically complex consonant sequences in PwA and during language acquisition. The fact that consonant sequences are modified in both acquisition (in children) and loss (in PwA with and without apraxia of speech) indicates an underlying complexity resultant of metrical structure (*i.e.* the syllable). Differential modifications by both groups should help shed light on models of language production and the various levels within phonological encoding also for the healthy individual.

References

Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93: 283–321.

Levelt, W.JL, Roelofs, A. & Meyer, A. (1999). A theory of lexical access in Speech production. *Behavioral and Brain Sciences*, 22: 1-75.

Meyer, A. & Belke, E. (2008). Word form retrieval in language production, in M. G. Gaskell (ed.). *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 471-487.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K. O'Brien, P. and Wareham, T. (2006) Introducing Phon: A software solution for the study of phonological acquisition, in Bamman, D, Magnitskaia, T, and Zaller, C, (eds.) *Proceedings of the 30th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 489–500.

Articulatory and acoustic changes in pre-adolescent and adolescent childrens' production of /s/ and /ʃ/
A case study in Hungarian

Tekla Etelka Gráczi[1,4], Alexandra Markó[2,4], Tamás Gábor Csapó[3,4], Andrea Deme[2,4], Márton Bartók[2,4]
[1] MTA Research Institute for Linguistics, Budapest, Hungary, [2] Department of Applied Linguistics and Phonetics, Eötvös Loránd University, Budapest, Hungary, [3] Budapest University of Technology and Economics, [4] MTA–ELTE „Lendület" Lingual Articulation Research Group, Budapest, Hungary
graczi.tekla.etelka@nytud.hu

Introduction

The fricative manner of articulation is acquired among the latest ones during language acquisition (Dinnsen, 1992; Kent, 1992) and its production still undergoes changes in early adolescence (Romeo, Hazan & Pettinato, 2013). The fricatives /s/ and /ʃ/ produced by 3- to 6-year-olds were found to be classifiable on the basis of the 1st and 3rd spectral moments (Nissen & Fox, 2005). Pre-adolescents' and adults' /s/ and /ʃ/ production was found to be similar both in their articulatory and acoustic characteristics (Zharkova, 2016); however, age-related differences occurred in the centre of gravity (CoG) values at consonant-vowel boundaries, and in the variability of CoG for /ʃ/. At most of the measurement points throughout the consonant duration, both speaker groups consistently differentiated between /s/ and /ʃ/ in both the CoG and in the tongue shape (LOC_{ai}) measures. Tongue contours were found to be affected by V-to-C coarticulation in /s/, but not in /ʃ/ in children, while in adults, both segments showed the effect of coarticulation (Zharkova, Hewlett, & Hardcastle, 2011, 2012). The present study focuses on a longitudinal-developmental aspect of changes in the articulatory and acoustic characteristics of the contrast of /s/ and /ʃ/ in pre-adolescents.

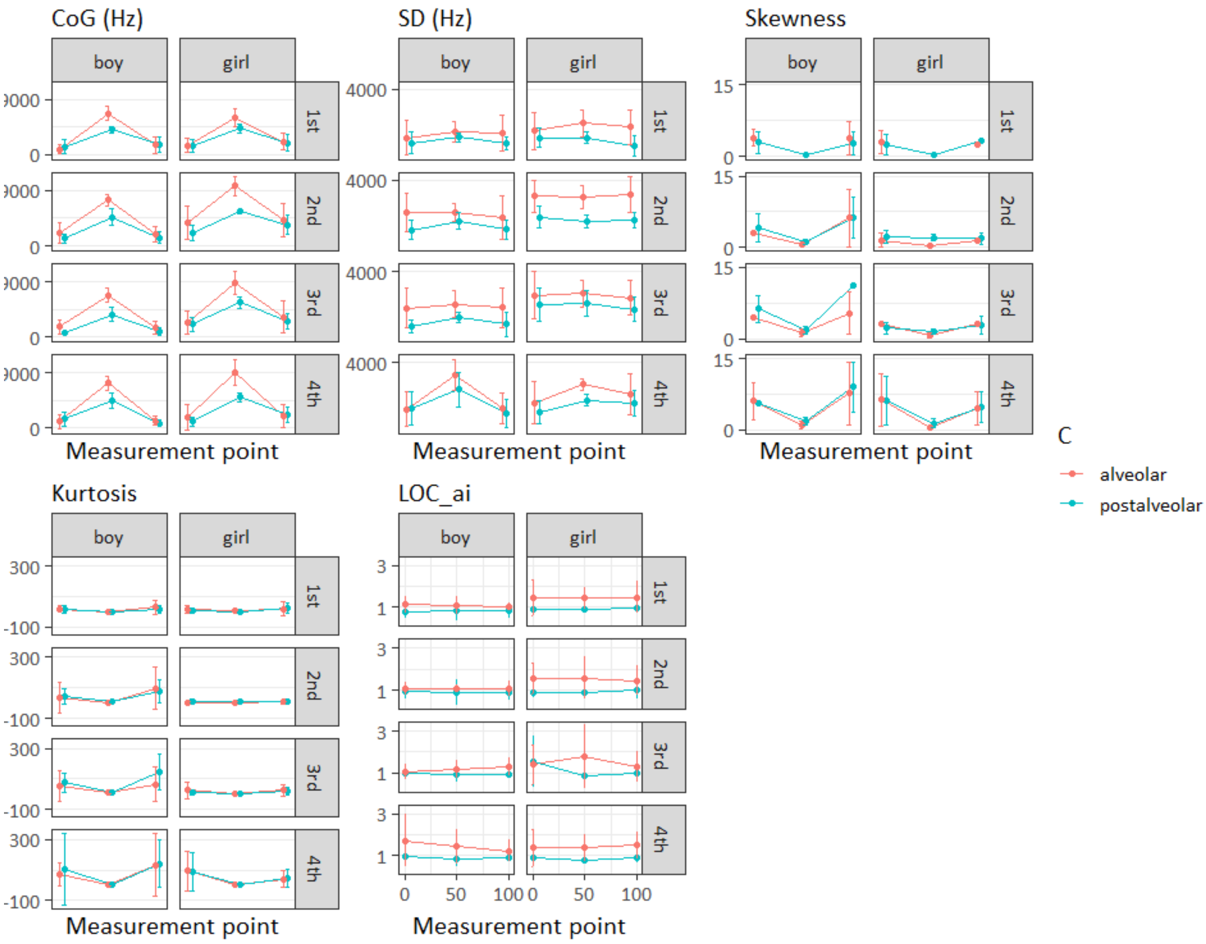
Method

We recorded speech samples of a boy and a girl (siblings, both were native, monolingual speakers of Hungarian) at the following ages: 7;5, 7;11, 8;5, 8;10 years in the case of the girl, and 11;0, 11;6, 12;0, 12;5 years in the case of the boy. Their task was to read aloud VCVCV nonsense words (2 times per token) in which all three Vs and two Cs were identical. We analyzed midsagittal tongue contours recorded by ultrasound (Articulate Instruments Ltd., 2012) and acoustic data of the consonants /s/ and /ʃ/ in /ɒ aː ɛ i u/ contexts at the time point of the closes US-image to the 0%, 50% and 100% of the total fricative duration. We also measured acoustic parameters for the total consonant duration. We measured CoG, SD, skewness (SK) and kurtosis (KU) in Praat (Boersma & Weenink, 2019). The tongue shapes were parameterized using LOC_{ai} (Zharkova, 2013). In order to avoid misunderstandings between the SD of the spectral shape and the SD of the measures themselves, the first one will be named sSD, the latter one will be referred to as sd henceforth. Repeated measures ANOVA tests were applied on the data separately for the two children in R (R Core Team, 2018). The dependent variables were CoG, sSD, skewness and kurtosis, $LOC_{a\text{-}i}$. The independent variables were the consonant (C), the recording age (A) and the measurement point (P) allowing the interactions of the factors.

Results

The difference of the CoG between /s/ and /ʃ/ is the largest at the midpoint in all four recording sessions in the case of both children. In the case of the boy, the interactions C*A and P*A had a significant effect on CoG, while in the case of the girl, all three pairwise interactions (C*A, P*A, C*P) The difference of the sSD between the two fricatives shows a development into this tendency. In the case of the boy, all three pairwise interactions (C*A, P*A, C*P) had a significant effect on sSD, while in the case of the girl only the C*A, P*A did. None of the further three factors showed the largest difference between the two fricatives at the midpoint. The sd of the skewness and kurtosis are larger in the 0% and 100% points in the fricatives. In the case of the girl, only the P*A interaction had a significant effect on both skewness and kurtosis. In the case of the boy, P and A had a significant effect on the skewness, and C, P and A on kurtosis. This means that these skewness was not distinctive in the case of any of the children between the two fricatives, and neither kurtosis in the case of the girl. This means that the distinction of the two fricatives is the largest at the midpoint in the measures of CoG and sSD, while skewness and kurtosis either reflect the contextual effects on the starting and end point of the pronunciation.

Figure 1: *The mean and sd of CoG (Hz), SD (Hz), skewness, kurtosis and LOC$_{ai}$ at the three measurement points in the four recording sessions*



The results for the tongue shape, i.e. LOC$_{ai}$ showed a diverse picture. The difference of LOC$_{ai}$ between the two fricatives was found at all measurement points in all four recording sessions for both children, however, their difference was highly variable, and did not show the tendency to be larger comparing the midpoint of the fricatives than at the starting or end points. The factors C and A had a significant effect on this measure in the case of the boy, while only C in the case of the girl. These might mean that either the vowel context plays a large role throughout the total duration of the fricatives on the tongue shapes, or that the variability of the tongue shape itself is large, but the distinction of the two fricatives are apparent in all four recording sessions throughout the fricative duration in both children's pronunciation.

Discussion and conclusions

In the present study we analyzed articulatory and acoustic characteristics of /s/ and /ʃ/ in a pre-adolescent and an adolescent child in a longitudinal aspect. The difference of /s/ and /ʃ/ was found to be present in both children's pronunciation: all measured acoustic (CoG, SD, skewness and kurtosis) and the articulatory (LOC$_{ai}$) factors showed significant difference at least at the midpoint of the consonants. The exact values changed across the four recording sessions; however, this variation did not show a quasilinear change as a function of age, which means that the exact distinction – although being present – is still undergoing changes along their age. The variability of skewness and kurtosis at the starting and end point of the fricatives were shown to be increasing along the age. This might have arisen due to the clearer coarticulatory effect of the context that included 5 different vowel qualities. This contextual effect was not apparent in the CoG or in the LOC_{ai}, and was not clear in the SD.

The time-frame of this longitudinal study was approximately 2 years (yet) in pre-adolescence, in which the children still undergo large and quick physiological changes, and their motor skills also improve. Although in the present study we included a low number of subjects, we were able to observe that (i) the results were consistent with previous findings, i.e. the distinction was apparent between the consonants, however, the fine tuning of the distinction and the contextual effect are still in a developmental phase, (ii) there is some shorter term changes that might be of importance in the study of the development of motor skills in pre-adolescence.

References

Dinnsen, D. A. (1992). Variation in developing and fully developed phonetic inventories. In C. Ferguson, L.M., & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications.* Timonium, MD: York Press. pp. 191–210.

Kent, R.D. (1992). The biology of phonological development. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications.* Timonium, MD: York Press. pp. 65–90.

Nissen, S.L., & Fox, R.A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *J. Acoust. Soc. Am.* 118, pp. 2570–2578.

Romeo, R., Hazan, V., & Pettinato, M. (2013). Developmental and gender-related trends of intra-talker variability in consonant production, *J. Acoust. Soc. Am.* 134, pp. 3781–3792.

Zharkova, N. (2016). Ultrasound and acoustic analysis of sibilant fricatives in preadolescents and adults. *J. Acoust. Soc. Am.* 139/5, pp. 2342-2351

Zharkova, N., Hewlett, N. & Hardcastle, W.J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control* 15, 118–140.

Zharkova, N., Hewlett, N., & Hardcastle, W.J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *JIPA* 42, pp. 193–-208.

Articulate Instruments Ltd. (2012). *Articulate Assistant Advanced Ultrasound Module User Guide: Version 2.14* (Articulate Instruments Ltd., Edinburgh, UK).

Boersma, P. & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program]. Version 6.1.08, retrieved 5 December 2019 from http://www.praat.org/.

Zharkova, N. (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal,* 50, pp. 76-81.

R Core Team (2018) *R: A Language and Environment of Statistical Computing.* R Foundation for Computing, Vienna.

Speech in noise perception in childhood: Role of sensory auditory processing and processing efficiency

Irene Lorenzini[1,2], Christian Lorenzi[3], Laurianne Cabrera[1,2]
[1]Integrative Neuroscience and Cognitive Centre, Université de Paris
[2] Integrative Neuroscience and Cognitive Center, CNRS
[3] Ecole Normale Supérieure, Université Paris Sciences et Lettres

Temporal cues (e.g., amplitude modulation, AM) play a crucial role in speech intelligibility for adults (Rosen, 1992). How the ability to track these temporal cues develops and interacts with the development of speech perception is however unclear. Previous work shows that AM sensitivity improves until late childhood (Hall & Grose, 1994). This may relate to the development of sensory factors (tuning of AM filters, susceptibility to AM masking) or to changes in processing efficiency (reduction in internal noise, optimization of decision strategies). A recent experimental and modelling study (Cabrera et al., 2019) showed that AM detection thresholds improved between 5 and 11 years of age and that this developmental trend was better accounted for by a reduction in internal noise in the temporal-envelope domain by a factor 10 rather than by changes in AM selectivity or by changes in decision strategy. Moreover, children's consonant identification thresholds in noise were related to some extent to AM sensitivity.

The present study aimed at evaluating more whether a reduction in internal noise in the temporal-envelope domain may support better use of temporal information in speech during childhood. To this goal, 82 children aged from 6 to 8 years completed three behavioural tasks assessing: (i) AM sensitivity and AM masking (i.e., modulation filtering); (ii) double-pass consistency for AM detection (i.e., internal noise) and (iii) speech-in-noise identification thresholds.

Task i: Development of AM sensitivity and AM masking

The development of AM filtering was evaluated through a 2-alternative-forced choice task (AFC), using an adaptive procedure to estimate AM detection thresholds for an 8-Hz sinusoidal AM. AM sensitivity was tested using either: (a) a 500-Hz sine tone (No Masking condition), and (b) a 4-Hz wide narrowband noise centred at 500 Hz (Masking condition).

Task ii: Double-pass consistency for AM detection

Internal noise for AM detection was estimated from a 'double-pass technique' using a constant-stimuli procedure (Green, 1964) testing the consistency of children's responses in an AM-detection task. Here, AM detection was measured at threshold in the Masking condition for 200 trials repeated twice (2 'passes') using the same 2-AFC task as in the first task. The Percentage of Correct AM detection in each pass (PC) and the Percentage of Agreement between the responses across the 2 passes (PA) were used to estimate within-listener consistency, a proxy of internal noise.

Task iii: Speech-in-noise identification

The stimuli were /a/-Consonant-/a/ syllables where the consonant was varied between French fricatives ([f], [v], [s], [z], [ʃ], [ʒ]) and stops ([b], [p], [d], [t], [g], [k]) in 6 different phonetic conditions. In each condition, these consonants were minimally contrasted on voicing, place or manner of articulation within a pair. Children completed an XAB adaptive task. They were asked to identify whether the A or B sound corresponded to the target X sound. The minimal phonetic contrast was between the sounds A and B. These two sounds were presented within a speech-shaped noise whose level varied according to the participant's response on a trial-to-trial basis. Here, we measured consonant identification thresholds in noise as a function of the phonetic condition.

All the tasks were completed on a touch-screen tablet through a child-friendly interface (cf. Figures 1 and 2). Additionally, all children completed two standardized tests assessing receptive vocabulary and non-verbal reasoning.

Results showed that AM detection thresholds obtained with both carriers did not significantly improve from 6 to 8 years (*ps* > .37) and that all children were similarly affected by AM masking ($p < .001$). As to double-pass consistency, both PC and PA increased with age (*ps* = .03). Thus, AM filtering was not affected by age, but our proxy of internal noise was. Regarding speech-in-noise, thresholds significantly improved with age ($p = .02$) and were affected by phonetic feature ($p < .001$). Backward regression analyses (for age-normalized scores) showed that AM masking scores significantly predicted the identification thresholds in noise for Manner (R^2 = 6.8%); PA predicted to a small extent identification thresholds for Voicing (R^2 = 4.7%), and vocabulary associated with non-verbal reasoning predicted data for Place (R^2 = 5.6%).

Overall, this research displays that modulation filtering, internal noise and linguistic level determine consonant-in-noise identification abilities in childhood. In particular, the role of temporal processing seems to vary according to the phonetic contrast to identify in noise.

Figure 1*: 2-alternative-forced choice task (Task i: AM detection and Task ii: double-pass consistency). Two characters were presented, one speaking after the other. The children were asked to touch the character who emitted the sound containing the target AM (i.e. the character "who made a buzzing sound").*

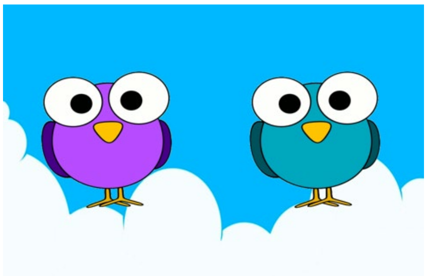
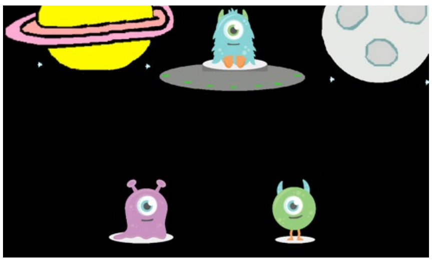


Figure 2: *XAB identification task (Task iii: speech-in-noise). Three characters were presented: the first (male voice) uttered a clearly spoken target syllable (no superimposed noise), the second and the third either repeated the target syllable or uttered a distractor, both masked by a speech-shaped noise. The children were asked to touch the character who repeated the same target syllable as the first character.*



References

Cabrera, L., Varnet, L., Buss, E., Rosen, S., & Lorenzi, C. (2019). Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity. *JASA,* 146(4), 2415-2429.

Green, D. M. (1964). Consistency of auditory detection judgments. Psychol Rev, 71(5), 392.

Hall III, J. W., & Grose, J. H. (1994). Development of temporal resolution in children as measured by the temporal modulation transfer function. *JASA*, 96(1), 150-154.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos T R Soc B*, *336*(1278), 367-373.

Aging effects on prosodic marking in German: An acoustic analysis

Jane Mertens[1], Doris Mücke[1] & Anne Hermes[2]
[1]IfL Phonetics, University of Cologne, Germany
[2]Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, France
{jane.mertens, doris.muecke }@uni-koeln.de; anne.hermes@sorbonne-nouvelle.fr

Aging entails various physiological changes, including loss of muscle cells, slowing down of movements and ossification of cartilage structures. With respect to speech motor control, these changes are likely to affect speech. Many studies revealed age-induced effects investigating speaking rate, fundamental frequency and vowel formants (Amerman & Parnell, 1992, Ramig, 1983, Nishio & Niimi, 2008, Eichhorn et al., 2018). However, not much is known about the effect of aging on the expression of prosodic prominence involving the highlighting of important information by fine-tuned modifications in the physical system (Cho 2011, Mücke & Grice, 2014). Thus, as aging has an effect on movement velocity, muscular strength and flexibility, strategies for prosodic marking are prone to be affected as a result of physiological constraints. In this study, we explore the effect of aging on the realization of accented and unaccented syllables in German by exploring prosodic marking in the acoustic dimension. Our results indicate that although both age groups mark prosodic prominence by means of lengthening the stressed vowel, older speakers exhibit even more lengthening, less variable formant values and a smaller vowel space area.

We carried out acoustic recordings from 20 older (10f, 10m; aged 65-80 years) and 20 younger (10 f, 10 m; aged 20-30 years) native speakers of German. The speech material consists of target words representing 10 different girl names (e.g. <*Mila*>) containing the vowels /a:, e:, i:, o:, u:/. Participants were presented with animated scenes and were asked to give the 'appropriate' answer to an auditorily and visually-presented question (see Table 1).

Table 1: *Example for questions and corresponding answers. Words with nuclear accents in bold.*

Condition		***Question***	***Answer***
Unaccented	*Background*	Hat die Oma der Mila gewunken? *'Had the grandma waved to Mila?'*	Der **Opa** hat der Mila gewunken. *'The **grandpa** had waved to Mila.'*
Accented	*Broad*	Was ist passiert? *'What happened?'*	Der Opa hat der **Mila** gewunken. *'The grandpa had waved to **Mila**.'*
	Contrastive	Hat der Opa der Kati gewunken? *'Had the grandpa waved to Kati?'*	Der Opa hat der **Mila** gewunken. *'The grandpa had waved to **Mila**.'*

Questions were designed to trigger the target word as either being in accented condition, i.e. *contrastive* focus or *broad* focus or in unaccented condition i.e. *background*. We discarded target words in background condition which were accented. We labelled segmental boundaries for the vowel (V1) of the stressed syllable, and we extracted the formants F1 and F2 at the midpoint of the V1. For determining the acoustic vowel space, the vowel articulation index (*VAI*) as well as the vowel space area were calculated using the formant values of /i:, a:, u:/ (Roy et al., 2009, Sapir et al. 2011). Here, we present preliminary data of 3 older (2 f, 1 m) and 3 younger speakers (2 f, 1 m). In what follows, we restrict ourselves to descriptive results, due to the small amount of data examined.

Figure 1: *Vowel duration (in ms) dependent on focus condition in younger and older speakers.*

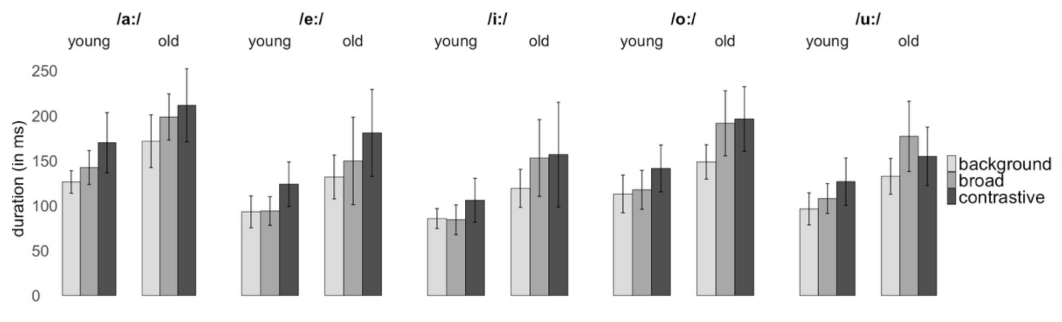
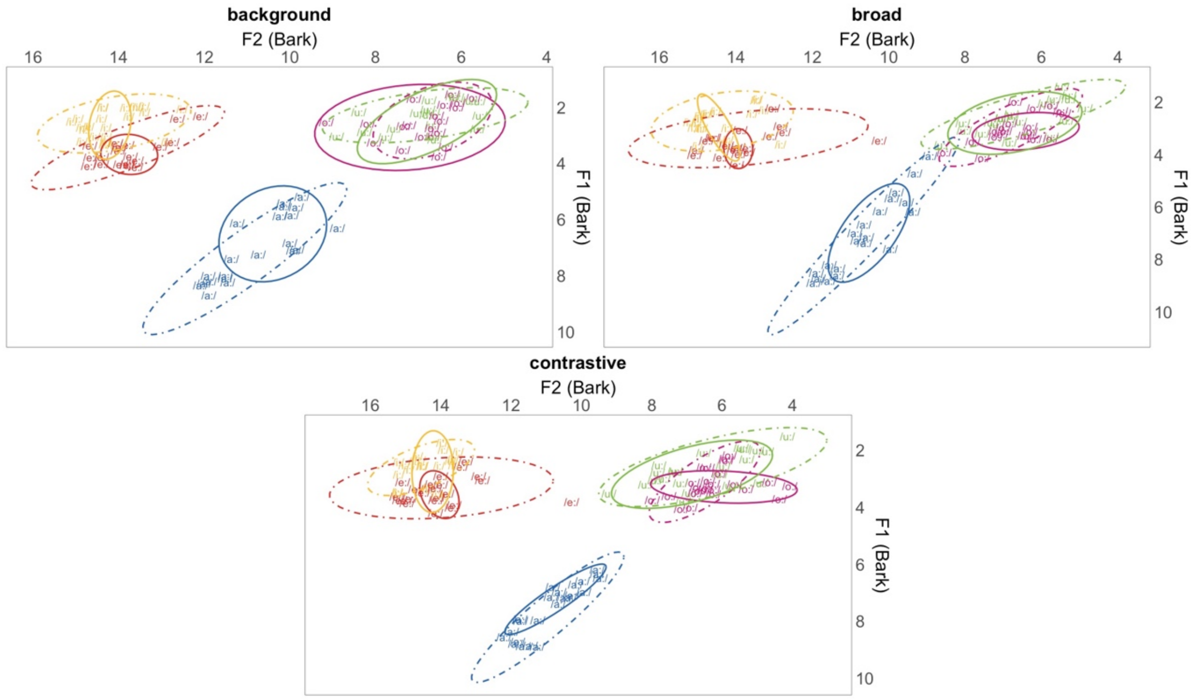


Figure 1 displays the results for the duration of the stressed vowel, separately for vowel quality, age group and focus condition. Looking at the maximally diverging focus structures, the vowel duration increased comparing background with contrastive focus for all vowels. When comparing within accentuation, vowel durations increased from broad to contrastive focus for the vowels /a:/ and /u:/ in the younger speakers and for the vowels /a:/ and /e:/ in the older speakers. Looking at the less divergent focus structure such as background and broad focus, vowel duration always increased from background to broad focus in the older speaker group (Hermes et al., 2008), whereas for the younger speakers no such systematic increase can be observed (Mücke & Grice, 2014). Additionally, age had an effect on vowel duration: Older speakers showed longer durations in all vowel qualities compared to younger speakers in the respective focus conditions. Turning to the formant measures, older speakers produced vowel /a:/ more centralised than younger speakers, especially in background and contrastive focus (see Figure 2).

Figure 2: *F1 and F2 (in Bark) for younger (dashed lines) and older speakers (solid lines) in background (upper left), broad focus (upper right) and contrastive focus (bottom).*

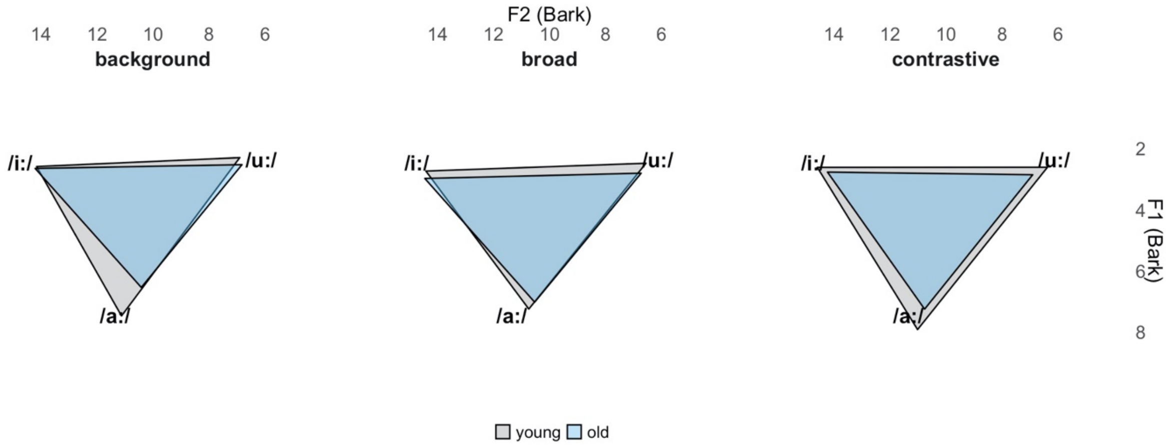


Furthermore, older speakers did show less variability than younger speakers in the formant values in all vowel types. Looking at the vowel articulation index (VAI; Roy et al., 2009, Sapir et al. 2011), cf. Table 2), younger speakers did show a stepwise increase from background over broad to contrastive focus. However, in older speakers the VAI increased from background to broad focus, but decreased from broad to contrastive focus. In Figure 3, the vowel space area is displayed. Age differences could be observed in all three conditions, with a reduced vowel space in older speakers, especially for the open vowel /a:/. The aging effect for VAI and vowel space area was strongest observable in the most prominent positions, i.e. contrastive focus.

Table 2: *Vowel Articulation Index [8,9] across focus condition for younger and older speakers.*

	Young	Old
Background	1.17	1.19
Broad	1.24	1.21
Contrastive	1.28	1.17

Figure 3: *Vowel space area for younger (grey) and older (blue) speakers in background (left), broad focus (middle) and contrastive focus (right).*



The preliminary results reveal that prosodic marking is reached by temporal and spatial modifications of the stressed vowel. In prominent positions, vowels are produced longer and more peripheral. This is in line with the literature on prosodic strengthening in German (Mücke & Grice, 2014, Hermes et al., 2008). Furthermore, there are age effects observable in the strategies for marking prosodic prominence: Older speakers tend to produce longer, but more centralized vowels than younger speakers. Both, the increase in vowel duration and the decrease in vowel space might be due to physiological constraints of the vocal tract and age-induced slowing down of movements (Ketcham & Stelmach, 2004, Hermes et al. 2018). However, the increase in duration could also be interpreted as a compensation strategy for the reduced vowel space. More strikingly, younger speakers show higher variability and therefore a more flexible vowel system within and across focus conditions than older speakers.

References

Amerman, J. D., & Parnell, M. M. (1992). Speech timing strategies in elderly adults. *Journal of Phonetics*, *20*(1), 65-76.

Cho, T. (2011). Laboratory phonology. *The continuum companion to phonology*, 343-368.

Eichhorn, J. T., Kent, R. D., Austin, D., & Vorperian, H. K. (2018). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice*, *32*(5), 644-e1.

Hermes, A., Mertens, J., & Mücke, D. (2018, September). Age-related Effects on Sensorimotor Control of Speech Production. In *Interspeech* (pp. 1526-1530).

Hermes, A., Becker, J., Mücke, D., Baumann, S., & Grice, M. (2008). Articulatory gestures and focus marking in German. *Proc. Speech Prosody 2008*, 457-460.

Ketcham, C. J., & Stelmach, G. E. (2004). Movement control in the older adult. In *Technology for adaptive aging*. National Academies Press (US).

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation–Is it mediated by accentuation? *Journal of Phonetics*, *44*, 47-61.

Nishio, M., & Niimi, S. (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia phoniatrica et logopaedica*, *60*(3), 120-127.

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of communication disorders*, *16*(3), 217-226.

Roy, N., Nissen, S. L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of communication disorders*, *42*(2), 124-135.

Sapir, S., Ramig, L. O., Spielman, J., & Fox, C. (2011). Acoustic metrics of vowel articulation in Parkinson's disease: Vowel space area (VSA) vs. vowel articulation index (VAI).

The development of narrative skills of Hungarian preschoolers

Sarolta Murányi
ELTE Eötvös Loránd University, Budapest, Hungary
muranyi.sarolta@gmail.com

Introduction

One of the important fields in the research of children's language is the examination of oral narratives. Children's narratives are analysed from numerous perspectives in both Hungarian and international literature.

The development of narrative skills is closely linked to linguistic and cognitive development (Applebee, 1978). Over the past few years, several studies have focused on children's oral narratives and stories. Based on the results of psychological, pedagogical and linguistic research, it can be concluded that age-related peculiarities in children's storytelling skills can be observed and described (Bruner et al., 1993; Schneider et al., 2005). Other than Applebee's (1978) model, there are now several new theories describing the narrative structure and its various development stages during the acquisition of language (Stein–Policastro, 1984; Meritt, 1987; ENNI, 2006). In the performance of children in this matter, family background, the teaching institution and targeted development all play an essential role (Fekonja–Peklaj et al., 2010).

Although many international studies have been investigating the development of narrative skills, their results cannot be generalized to Hungarian children for cultural, linguistic and educational reasons. Hungarian literature and educational practice lack a general investigation of narrative skills that would also examine the linguistic and cognitive factors influencing them, and perform statistical analyses on the results of tests measuring the development of these areas.

The Hungarian linguistic studies on the narrative skills of preschool-aged children so far have mainly examined the activated vocabulary, the length of the narrative and the grammatical structures. However, a few studies have also commented on the macrostructure of the narratives created. The studies mapping the attributes of preschool narratives have confirmed that even school-aged children find this kind of productive narrative-building challenging. In this age, children may have trouble with identifying the characters, the space and the time of the events, and the structure of the narrative (Csákberényiné–Hajdú 2011).

Most 3-4 year olds cannot form a narrative based on pictures on their own and without asking questions. In this age, the narrative of children typically consists of listing the characters and items shown in the pictures, and naming a few connections. 6-7 years olds however can form a narrative on their own that contains more causal relations (Murányi 2017).

This longitudinal study focuses on analysing picture-narratives formed by preschool children. The study follows the changes in narrative skills for a three-year period. It also examines the ways in which narratives are formed in various ages, how fluently and independently they are formed, and observes their coherence, structure and vocabulary.

Methodology

The study group included 12 Hungarian native preschoolers with typical language development. At the beginning of the study, the participants had just started preschool, and by the recording of the last audio tape, they were in their final year, so altogether three years passed between the first and the last recording. The children had to tell stories using the same set of pictures (consisting of 6 images) on every occasion with the same instructions given to them.

The detailed analysis of the narratives was based on considering their length, chronology, structure and the presence of descriptive parts and causal relationships. In addition to the above, I have also examined how independently the narrative was formed, and how much help did the child need during the telling of the story.

Results

The findings confirm that there is significant improvement in this area during the preschool years – this improvement is confirmed in all aspects considered during the analysis of the narratives. The greatest change is observed between the first and second year narratives. The vocabulary has doubled, and the indicators signifying coherence have typically all increased. Besides the age-related analysis, it should be emphasized that huge individual differences can be observed in the studied variables.

Conclusions

The observations made during this experiment contribute to understanding children's narrative competence better. For a more detailed analysis of the results of this study, the experiment should be repeated with a bigger sample, and the introduction of newer indicators.

References

Adamikné Jászó A. (2001) Anyanyelvi nevelés az ábécétől az érettségiig. Budapest: Trezor Kiadó.

Applebee, A. (1978). The child's concept of story. Chicago: University of Chicago Press.

Bruner, J., Feldman, C., Kalmar, D., & Renderer, B.(1993) Plot, plight, and dramatism: Interpretation at three ages. Human Development, 36. 6. 309–323.

Csákberényiné Tóth K. & Hajdú T. (2011) Az iskolába lépő gyerekek szóbeli szövegalkotásának jellemzői. Anyanyelv-pedagógia 3.

Fekonja-Peklaj, Urška M., & Ljubica, S. (2010) Children's storytelling: The effect of preschool and family environment. European Early Childhood Education Research Journal 18/1.55–73.

Merritt, D.D., & Liles, B.Z. (1987) Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. Journal of Speech and Hearing Research, 30, 539-552.

Murányi S. (2018): A szavaktól a történetekig. Anyanyelv-pedagógia, 11. 3. sz.

Schneider, P., Dubé, R. V., & Hayward, D. (2005) The Edmonton Narrative Norms Instrument. Retrieved from University of Alberta Faculty of Rehabilitation Medicine. Elérhető: University of Alberta Faculty of Rehabilitation Medicine website:www.rehabresearch.ualberta.ca/enni.

Schneider, P., Hayward, D., & Dubé, R. V. (2006) Storytelling from pictures using the Edmonton Narrative Norms Instrument. Journal of Speech-Language Pathology and Audiology, 30,224-238.

Stein, N. L., Glenn, C. (1979). An analysis of story comprehension in elementary school children. New directions in discourse processing 53–120 Norwood, NJ: Ablex.

Stein, N. L., & Policastro, M. (1984) The concept of a story: A comparison betweenchildren's and teachers' viewpoints. Learning and the comprehension of text pp. 113-155. Hillsdale, NJ: Lawrence Erlbaum Associates.

The influence of conceptual and visual factors on sentence production in younger and older adults

Sabine Reuters
Department of Rehabilitation and Special Education, University of Cologne, Germany
sabine.reuters@uni-koeln.de

The current study investigated sentence production in young and elderly German subjects with regard to scene description. In particular, it set out to test in how far certain attention-orienting conceptual and visual manipulations such as patient animacy and patient position lead to systematic variations between active sentences (*Der Engel schiebt den Piraten*) and passive sentences (*Der Pirat wird vom Engel geschoben*). Additionally, it examined how these factors influence speech onset times (SOT). Most studies which found that language production is affected by the animacy status and position of actants in event scenes have been conducted with younger adults (e.g. Esaulova et al., 2019) whereas literature on experimental studies with elderly adults is sparse. Since the human aging process is associated with a gain in linguistic experience, but also with diverse cognitive changes regarding attention mechanisms (e.g. Benwell et al., 2014), experiments with elderly participants might lead to different results. My goals were: (i) to test the overall preference of syntactic structures and SOT in young and elderly German subjects during a picture description task and (ii) to investigate how patient animacy and patient position in this task determine the selection of syntactic structures and SOT in young and older German participants.

The experiment was based on a 2x2x2 mixed factorial design with the within-factors patient animacy and patient position and the between-factor age group as independent variables and frequency of passive sentences and SOT for actives as dependent variables. In total, 60 participants were tested. These subjects were divided in two age groups. The first group consisted of German younger adults (N=30, age range=18-27 years, mean age=22.57 years, *SD*=2.33) whereas the second group comprised elderly German participants (N=30, age range=63-80 years, mean age=72.37 years, *SD*=5.28). The subjects were asked to describe simple black-and-white drawings depicting diverse transitive interactions (e.g. pushing) in a single sentence. The stimuli showed scenes with animate or inanimate patients. Additionally, the patient was either positioned left or right. All nouns were controlled for word length, word form frequency, morphological complexity and grammatical gender. Verbs were controlled for their occurrence in passive voice. In terms of visual aspects, I also controlled referent size, referent color and the distance between referents. Furthermore, both age groups were matched for handedness, gender and educational level.

The statistical analysis was carried out by means of separate mixed three-way ANOVAs for the frequency of passivizations and SOT. Regarding the frequency of passives, the statistical analysis revealed a significant main effect of age group on the production of passive utterances ($F(1, 58)=9.76$, $p=.003$, $f=0.41$). The older participants produced a significantly higher number of passivizations than younger adults (see figure 1). In addition, a significant main effect of animacy on the frequency of passive utterances could be observed ($F(1, 58)=23.39$, $p<.001$, $f=0.63$). Overall, participants produced more patient-initial passive utterances in the conditions with an animate patient than in the conditions with an inanimate patient. Interestingly, I also found a significant interaction between age group and animacy ($F(1, 58)=4.77$, $p=.033$, $f=0.29$). Elderly participants produced significantly more passive utterances in the conditions with an animate patient than in the conditions with an inanimate patient whereas younger adults only tended to utter more passivizations in the conditions with an animate patient than in the conditions with an inanimate patient (see figure 3). According to Cohen (1988), both the main effect of age group and the main effect of animacy on the production of passivizations can be regarded as strong effect, while the combined effect of age group and animacy on the occurrence of passivizations can be classified as moderate effect. Regarding SOT, the ANOVA led to a significant main effect of age group ($F(1, 58)=24.86$, $p<.001$, $f=0.65$). Elderly participants needed more time for the initiation of active utterances than younger participants (see figure 2). In addition, a significant main effect of position on SOT for actives could be observed (*F*(1, 58)=4.61, *p*=.036, *f*=0.28). Thus, overall participants needed more time for the initiation of actives when the target stimulus depicted a left-positioned patient than when it depicted a right-positioned patient. Concerning SOT for active utterances, there was no interaction between age group and position (*F*(1, 58)=0.02, *p*=.899) (see figure 4). According to Cohen (1988), the significant main effect of age group on SOT can be categorized as strong effect and the significant effect of position on reaction times as moderate effect.

Figure 1

Frequency of passives in young and old adults

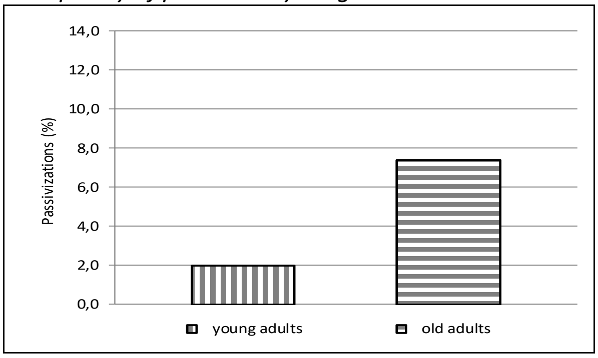


Figure 2

SOT for actives in young and old adults

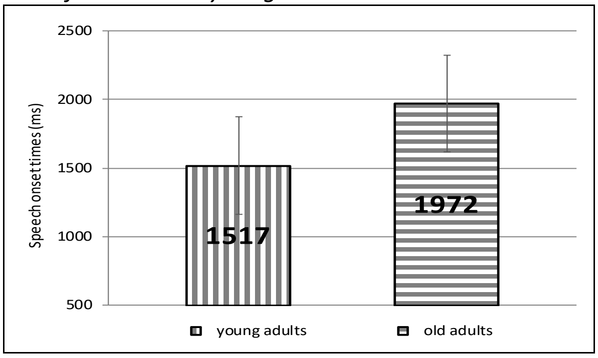


Figure 3

Animacy effects on frequency of passives in young and old adults

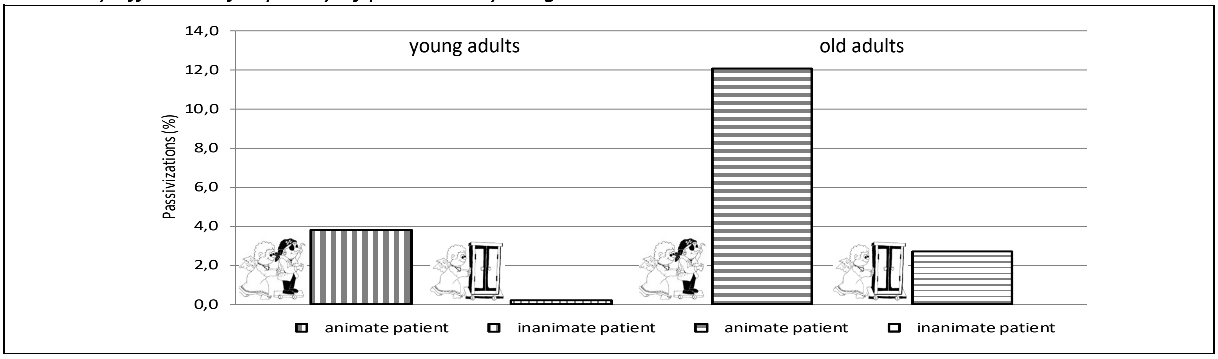
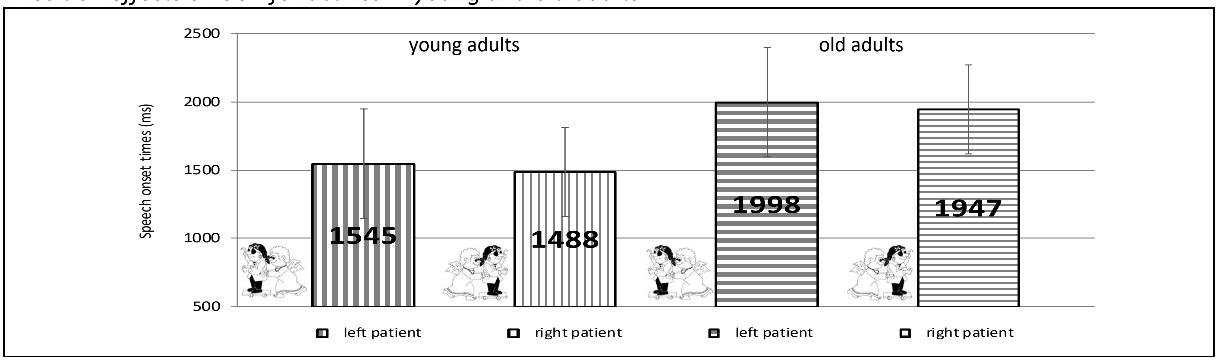


Figure 4

Position effects on SOT for actives in young and old adults



In my study, similarities and differences regarding young and elderly participants' sentence production were found. First of all, in line with the overall slowing hypothesis by Salthouse (1994), older adults needed much more time for the initiation of actives than younger participants. Secondly, elderly participants in total produced a significantly higher number of passives in comparison to younger participants. Regarding the influence of conceptual and visual factors on sentence production, in line with Esaulova et al. (2019) it could be observed that both young and older participants produced more passivizations regarding stimuli with animate patients in comparison to stimuli with inanimate patients, yet this effect was larger in the elderly. In contrast, the spatial positioning of actants affected SOT for actives since in line with Esaulova et al. (2019) both the young and the elderly showed longer SOT for stimuli with left-positioned patients in comparison to stimuli with right-positioned patients.

The fact that elderly participants needed more time for the initiation of actives than younger adults supports the overall slowing hypothesis postulated by Salthouse (1994) which claims that all cognitive computations are slowed down with aging. The combination of a higher number of passives and longer SOT in older adults can be considered as evidence for the assumption that due to delayed sentence production older participants apply a sentence planning strategy called word-by-word linear incrementality. Since older adults start speaking as soon as the first lemma is available without advanced grammatical planning these speakers are forced to produce an infrequent non-canonical structure in case the patient is retrieved first. In contrast, younger participants apply a sentence planning mechanism called structural (linear) incrementality meaning that sentence planning is controlled by a larger linguistic unit which enables the speaker to look ahead to the following structure leading to a lower number of passives (Thompson et al., 2015).

As to the influence of conceptual and visual manipulations, the choice of a grammatical structure was only determined by the animacy of thematic roles. In both age groups, more passives were produced for stimuli with animate patients in comparison to stimuli with inanimate patients, but this effect was even larger in the elderly. Thus, the study delivers counterevidence to the assumption that elderly participants use differences in animacy less efficiently in choosing a referent as sentence-initial subject (Altmann & Kemper, 2006). Apparently, elderly people use their semantic and syntactic knowledge that it is more likely to start a sentence with an animate than with an inanimate patient to a greater extent than younger adults as a strategy to compensate for delayed speech production.

In contrast, overall the initiation of actives was mostly affected by the spatial positioning of patients in event scenes. Stimuli with left-positioned patients led to slower initiations of actives than stimuli with right-positioned patients indicating higher processing costs for pictures with left-positioned patients. Since hemispheric lateralization, i.e. the dominance of the right hemisphere in visual-spatial tasks, causes young adults to favor the left visual field (e.g. Benwell et al., 2014) and thus to process left-positioned referents first, pictures in which the more prominent left position is occupied by the less prominent patient lead to higher cognitive costs in planning the utterance. This effect is strengthened by reading habits because reading orientation in German goes from left to right. Importantly, this effect also occurred in elderly participants which shows that in this case the influence of reading habits apparently overrides the impact of the reduced dominance of the right hemisphere in visual-spatial tasks with aging (e.g. Benwell et al., 2014). In summary, not all results found in studies with younger adults (e.g. Esaulova et al., 2019) can be generalized to people of all ages. The new insights can serve as tool to develop material used in speech and language therapy and for marketing strategies.

References

Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes*, *21* (1–3), 322–354. https://doi.org.10.1080/016909605400006

Benwell, C. S. Y., Thut, G., Grant, A., & Harvey, M. (2014). A rightward shift in the visuospatial attention vector with healthy aging. *Frontiers in Aging Neuroscience, 6*, 1–11. https://doi.org/10.3389/fnagi.2014.00113

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Esaulova, Y., Penke, M., & Dolscheid, S. (2019). Describing events: changes in eye movements and speech due to visual and conceptual properties of scenes. *Frontiers in Psychology*, *10*, 1–15. https://doi.org/10.3389/fpsyg.2019.00835

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8* (4), 535–543.

Thompson, C. K., Faroqi-Shah, Y., & Lee, Y. (2015). Models of sentence production. In A. E. Hillis (Ed.), *The handbook of adult language disorders* (pp. 328–354). Hove, Great Britain: Psychology Press.

The impact of levodopa on acoustic parameters of prominence marking and tongue body movements in patients with Parkinson's disease

Tabea Thies[1,2], Doris Mücke[1], Richard Dano[2], Michael T. Barbe[2]
[1]University of Cologne, Faculty of Arts and Humanities, IfL Phonetics
[2]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Neurology; tabea.thies@uni-koeln.de

Parkinson's Disease (PD) is a neurodegenerative disorder which is characterized by a progressive loss of dopaminergic neurons in the substantia nigra. As dopamine controls non-motor and motor circuits, PD affects non-motor and motor functions. Gross motor symptoms are manifested in smaller, slower and less extended movements and/or a resting tremor. In addition, many patients develop a speech disorder (hypokinetic dysarthria). The speech is characterized as monotone, quiet and slurred (Duffy, 2013). Despite the lack of controlling speech, Thies et al. (2020) investigated that mild dysarthric patients with PD can produce prosodic prominence by modulating syllable duration, intensity and fundamental frequency (F0). Although patients also changed their vowel articulation, the overall vowel space was smaller in patients compared to healthy controls. Patients with PD are treated with dopaminergic medication such as levodopa, which raises the dopamine concentration in the brain to improve motor functions. Whereas it is proven that levodopa is an effective treatment for gross motor performance (Katzenschlager & Lees, 2002), it remains unclear to what extent it influences speech production. Therefore, this study investigates the influence of levodopa on speech performance of patients with PD based on acoustic and articulatory measurements.

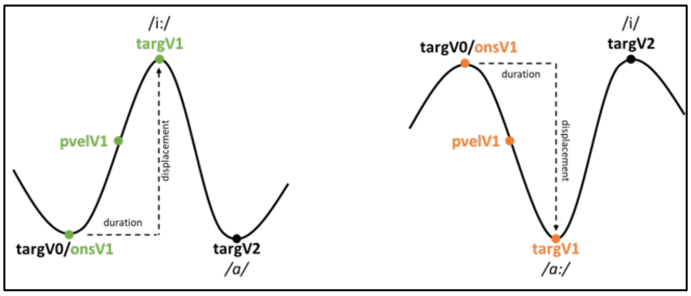
Fifteen native German speaking patients with idiopathic PD were recorded with an Electromagnetic Articulograph (AG 501, Carstens system) within one recording session in two conditions in the following order: (1) without drug intake (med-OFF, 12 hours cessation of PD medication) and (2) with drug intake (med-ON, 200 mg levodopa). Patients performed a speech production task. The speech task was designed as a question-answer-scenario to elicit target words in three different focus structures: background, broad focus and contrastive focus. The target words were disyllabic girl names (CV.CV) with word stress on the first syllable (Table 1). To control for segmental context, we alternated open and closed vowels in the vicinity of the target syllable, e.g. in the carrier sentence *'Der Opa hat der **Mila** gewunken'* (*The grandpa had waved to Mila).*

Table 1: *Target words consisting of six different vowels and controlled segmental make-up.*

preceeding article	**[dea]**	**preceeding article**	**[di:]**
open – closed – open vowel structure	[mi:la, li:na]; [me:la, le:na] [mu:la, lu:na]	closed – open – closed vowel structure	[mo:li, lo:ni], [ma:li, la:ni]

To capture the marking of prosodic prominence in the acoustic domain, syllable durations of the target words in all focus conditions as well as the tonal height of nuclear rising pitch accents in the F0 trace for the broad and contrastive focus were measured. As the vocalic element is the main domain of changes within prosodic prominence marking, articulatory measures of the tongue dorsum movement during the production of V1 in the stressed syllable of the target words were carried out. The articulatory results were related to the task-dynamic-model of speech (Browman & Goldstein, 1986). Therefore, the following variables were used: i) gestural duration, ii) displacement, iii) peak velocity and iv) stiffness (peak velocity/displacement; Munhall et al., 1985, see Figure 1). Stiffness is an abstract control parameter that is related to the relative speed of a movement. Results of the stressed closed vowel /i:/ and of the open vowel /a:/ will be presented. In addition, patients' motor impairment was evaluated using the 'Unified Parkinson's Disease Rating Scale' (UPDRS III; 18 items; the higher the score the stronger the motor impairment; Goetz et al., 2018) in med-OFF and med-ON condition. The test scores of both conditions were compared for investigating the influence levodopa had on the patients' motor function. The dopa response indicates the percentage of motor improvement from OFF to ON condition.

***Figure 1:** Movements of tongue body and articulatory measurements.*



Preliminary results from six patients show that all patients responded well to the levodopa intake (Table 2). When comparing med-OFF with med-ON, the acoustic syllable durations decrease under medication, while the tonal range of the rising F0 movements increases with med-ON (Figure 2). Furthermore, changes on the articulatory level are observable: Under medication, we find i) shorter vocalic gesture durations, ii) higher displacements and iii) higher maximal velocities (Figure 3). The stiffness values of vocalic movements decrease in the unaccented syllables (background), but increases in accented syllables (broad, contrastive focus) in med-ON condition (Figure 3). We found evidence for prominence marking in med-ON and med-OFF condition. Figure 2 shows that the syllable durations in accented syllables (broad, contrastive focus) are longer than in unaccented syllables (background). However, syllable durations between contrastive focus and broad focus productions do not differ. Patients also mark prosodic prominence by producing a rising F0 contour. The tonal height of this rise is higher in contrastive focus condition compared to broad focus condition to increase prominence (Figure 2). The comparison of maximum divergent focus structure (background vs. contrastive focus) on the articulatory level raise evidence that patients with med-ON produce longer, faster and stiffer vocalic movements in prominent positions. Interestingly, the same patients produce longer but slower and less stiff V1 movements in prominent positions with med-OFF.

***Table 2:** Patients' characteristics and evaluated motor functions as well as dopa responses.*

PwPD	age	gender	PD duration (years)	UPDRS III med-OFF	UPDRS III med-ON	dopa response (%)
PD01	53	m	4	45	24	47
PD02	54	m	6	44	27	38
PD03	58	m	2	20	11	45
PD04	52	f	6	41	20	51
PD05	56	f	6	45	29	36
PD06	70	f	20	27	14	48

These preliminary results demonstrate that levodopa has an influence on acoustic and articulatory parameters of speech production in patients with PD. Under medication, patients produce shorter syllable durations related to a higher articulation rate in general. Further, they produce shorter, larger and faster tongue body movements in med-ON condition compared to med-OFF. This points to the fact that levodopa improves not only gross motor skills but also speech motor skills. Moreover, our data confirms that patients with PD can mark prosodic prominence with and without medication. In line with Thies et al. (2020), they increase syllable duration, modulate F0 and change vowel articulation. However, the prosodic adjustments are less efficient in med-OFF condition. This is especially the case for the changes in vowel articulation, where we even found a shrinking and therefore a hypo-articulation of the vowel space under prominence in med-OFF. The patients seem to rescale the vocalic movement under prominence, but in the opposite direction in med-ON and med-OFF. Therefore, strategies of prominence marking differ across medication conditions.

Patients with Parkinson's disease mark prosodic prominence by changing phonetic parameters such as F0, syllable duration and vowel articulation from background to broad focus and contrastive focus condition. Moreover, this data set suggests that levodopa has an influence on patients' speech performance by producing sounds more efficiently and precisely and by improving the modulation of prosodic prominence markers in med-ON condition.

Figure 2: *Acoustic syllable durations (left) and tonal height of F0 movements in accented syllables (right).*

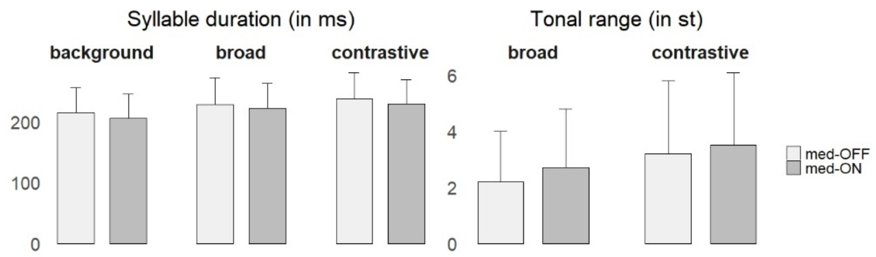
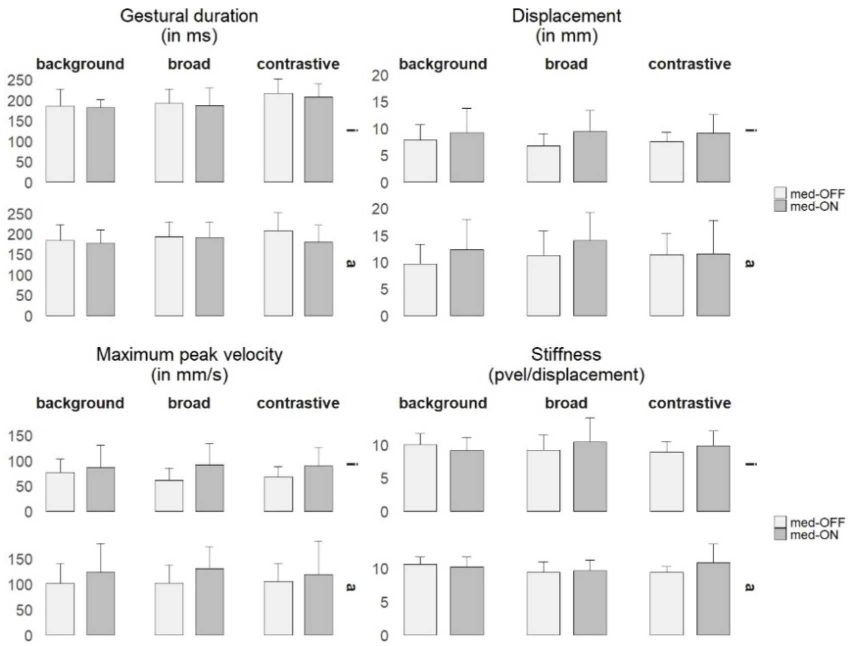


Figure 3: *Articulatory measures of the task-dynamics applied to the V1 vowels /i:/ and /a:/.*



References

Duffy, J. R. (2013). *Motor Speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences.

Goetz, C. G., & the Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: Scale presentation and clinimetric testing results. *Movement Disorders, 22,* 2129–2170.

Hughes, A. J., et al. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases*. Journal of Neurology, Neurosurgery & Psychiatry, 55*, 181-184.

Katzenschlager, R., & Lees, A. J. (2002). Treatment of Parkinson's disease: levodopa as the first choice. *Journal of neurology, 249*(2), ii19-ii24.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. Phonology, 3(1), 219-252.

Munhall, K. G., et al. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 457-474.

Thies, T., et al. (2020). Prominence marking in parkinsonian speech and its correlation with motor performance and cognitive abilities. *Neuropsychologia, 137*(3),107306.

Effect of aging on speech production

Anne Hermes
Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, France
anne.hermes@sorbonne-nouvelle.fr

Aging as an inevitable process involves changes at different physiological levels, including inter alia increased stiffness and decreased strength of the connective tissue, a decrease in mass and strength of the muscles, sensory losses, decrease in motor function and also a decline in breathing control. All these changes can lead to deficits in movement and posture, involving not only limbs and torso, but also the speech organs. The process of aging is thus expected to have an impact on speech production.

While there exists quite a lot of research in language acquisition, the research currently available on language attrition is particularly scant. In this talk I will present a review on effects of aging on speech production. I will start with an overview of age-related effects on the general motor control system, which can be partly transferred to the speech motor control system. In the following, I will emphasize selected research findings from speech production studies (from acoustics and articulation), trying to disentangle age-related from disease-related effects. I will show that incorporating age-related speech allows to cover repair mechanisms and adaptation to changes of the speech system, and I will discuss what kind of linguistic structures are maintained and what structures are more vulnerable to change. An outlook of this talk will be the claim for a systematic analysis of linguistic patterns in age-related speech informing us about how much variability is linked to the process of aging, and could furthermore convey a transition from aging to disease.

Effects of talker identity on speech intelligibility: a lifespan perspective

Adriana Hanulíková
University of Freiburg, Germany
adriana.hanulikova@germanistik.uni-freiburg.de

Previous research suggests that perceived racial identity may alter comprehension and interpretation of native-accented and foreign-accented speech. Following expectation-based accounts, anticipating a talker's accent based on speaker ethnicity enhances intelligibility of speech if expectations are confirmed (e.g. Babel & Russell 2015; McGowan 2015). In contrast, bias-based accounts such as reverse linguistic stereotyping (e.g. Rubin, 1992; Kang & Rubin, 2009) predict comprehension difficulties due to stereotypical attributions to speaker's appearance. Such stereotyping, however, only occurs when native speech is perceived to be from non-native speakers.

In this study, this apparent contradiction is examined by asking a) how perceived ethnicity shapes speech comprehension across different varieties and tasks, and b) how effects of speaker identity on speech comprehension develop across the life span. Given that stereotypes and biases tend to be more difficult to suppress with increasing age, talker effects on speech comprehension may increase as well.

All together, 172 native speakers of German (non-university population) were tested. High-school teens (n = 72, mean age 14.1; range 12-15), elderly (n = 50; mean age 77.6; range 70-92), and adults (n = 50, mean age 35.9, range 30-45) listened to nonsense utterances (e.g. *the old helmet is fishing*) that were embedded in speech-shaped noise at 0 SRN. Utterances in Standard German, a regional (Palatinate) dialect, and a foreign (Korean) accent were presented over loud speakers along with photographs of white Caucasian or Asian women displayed on a computer monitor. Each accent was presented with each ethnicity. Teens were asked to write down and adults to repeat what the ostensible speakers said. Responses were scored based on % correct word transcriptions/repetitions. In addition, accent ratings, cognitive skills, and hearing ability (for adults) were collected. A generalized linear mixed-effects regression for binary responses was fitted to the intelligibility data. The accent ratings were analysed using cumulative link mixed models.

In line with expectation-based accounts, speech intelligibility increased in case of a match between the talker's accent and perceived ethnicity of the ostensible speaker (see Figure 1). Across all groups, Korean accent was more intelligible when presented along with the Asian guise, and Palatinate accent was more intelligible when presented along with the white Caucasian guise. Although speech comprehension in the standard-accent condition was considerably higher than both foreign and regional accent conditions (with the Korean accent being more intelligible than the Palatinate accent), repetition accuracy did not vary as a function of speaker ethnicity, in contrast to what both accounts would predict. Elderly experienced greater difficulty understanding speech relative to adults and teens; however, talker effects were comparable across groups. In contrast to the transcription task, accent ratings revealed a main effect of ethnicity such that all varieties were rated as more accented in the Asian guise (see Figure 2), in line with previous studies and the bias-based accounts. This effect was most pronounced in elderly and least pronounced in adults.

Overall, the results show that social cues and linguistic information interact and stay relatively stable across the lifespan. However, the effects of talker ethnicity are more likely to emerge when speech is substantially degraded by background noise and unfamiliar accents. In addition, the effects are short-lived as participants adjust to both speakers and their speech during the course of an experiment. In conclusion, while both theories can account for the present findings, they explain different aspects of speaker identity integration. Bias-based theories account well for metalinguistic and postperceptual processes, which are usually assessed by indirect tasks such as accent and comprehensibility ratings. Expectation-based theories, in contrast, are accounted for using direct methods such as transcription tasks that assess what exactly listeners comprehended. The apparent contradiction in previous findings and theories thus stem from different types of measures that tap into distinct aspects of social processing.

***Figure 1**: Transcription/repetition scores per group of participants.*

Teens *Adults* *Elderly*

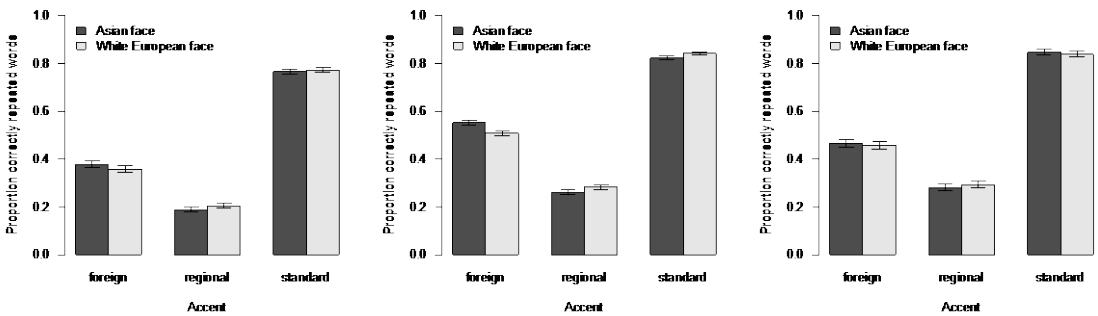
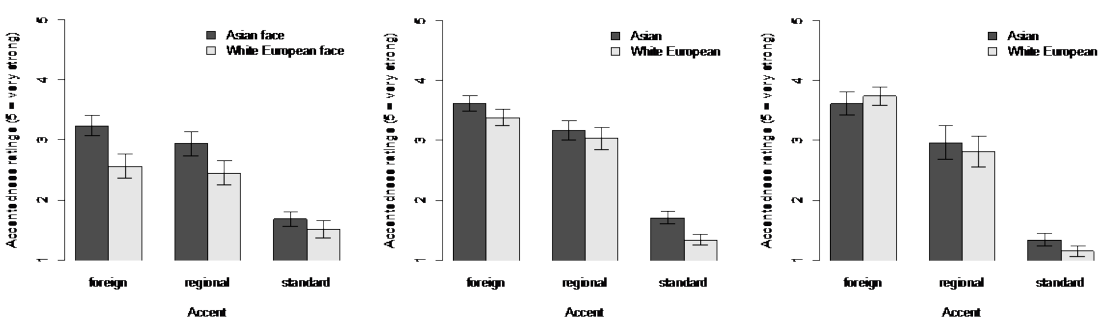


***Figure 2**: Accentedness ratings per group of participants.*

Teens *Adults* *Elderly*



References

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of Acoustical Society of America,* 137(5), 2823-2833.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology,* 28, 441–456.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech,* 58(4), 502-521.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgements of nonnative English-speaking teaching assistants. *Research in Higher Education,* 33(4), 511-531.

Self-repairs in speech: The effect of age and speech task

Judit Bóna
ELTE Eötvös Loránd University, Budapest, Hungary
bona.judit@btk.elte.hu

During spontaneous speech, speakers monitor their own speech, and they often correct it. Self-monitoring can be covert and overt. In case of covert monitoring, the reparandum is not articulated, filled pauses and repetitions may refer to it. In case of overt monitoring, the speech error and its repair both appear in the speech. Levelt (1983) classifies self-repairs into 5 types, depending on their reason: Different information repairs (D-repairs), Appropriateness repairs (A-repairs), Error-repairs (E-repairs), Covert repairs, Other repairs.

Overt speech errors and repairs clearly express self-monitoring. There are two main repair-strategies during speaking: speakers prefer either fluency or accuracy (Seyfeddinipur et al. 2008; Nooteboom & Quené 2017). If the speaker prefers fluency, they speak as long as they find a plan for correction. This means that they can even continue the original utterance even after having articulated the error. If the speaker prefers accuracy, they try to interrupt the reparandum as soon as possible, independent of the time they need for correction. The two strategies are characterized by different durational patterns and other acoustic-phonetic properties.

Speech planning (and self-monitoring) processes are influenced by several factors like speakers' age and speech task (Yairi–Clifton 1972; Duchin–Mysak 1987; Jacewicz et al. 2010). This study investigates the characteristics of self-monitoring of speakers of various ages in two speech tasks: spontaneous narratives and narrative recalls. It analyses three types of overt repairs: the proportion of D-repairs, A-repairs, and E-repairs, and it examines the characteristics of E-repairs in detail.

The main questions of the analysis were the following: 1) What durational patterns and functions characterize the monitoring processes in the analysed age groups? 2) How does the dichotomy of fluency vs. accuracy occur in the self-repairs of the different age groups? Is there a difference in self-repair strategies between the examined age groups?

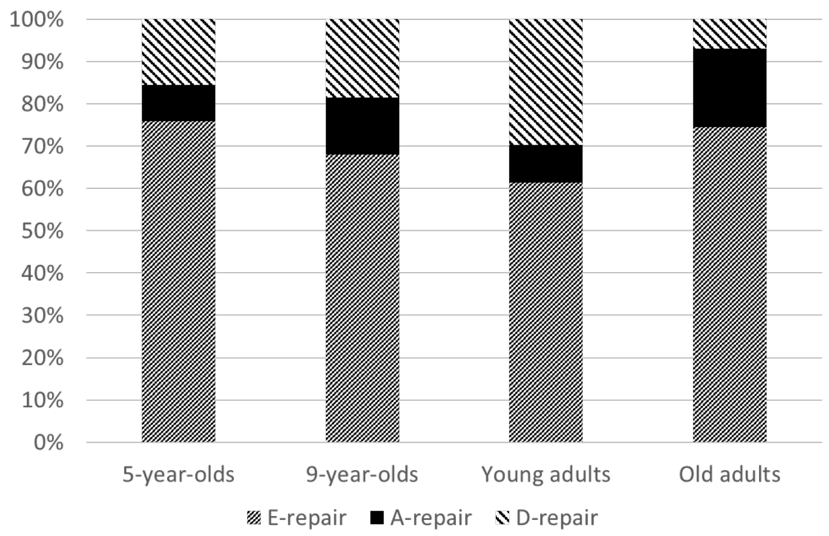
I hypothesized that 1) there are differences between the groups in the proportion of the different types of self-repairs, but error-repair is the most frequent type (compared to A-repairs and D-repairs). 2) Children interrupt their erroneous utterances earlier than (young and old) adults, but they need longer time to correct their errors. This means that children "prefer accuracy", while adults prefer fluency in speech. 3) There are differences between the two speech tasks in the repair-strategies, too.

For the analysis, spontaneous speech samples were analysed in four age groups: in 5-year-old children, 9-year-old children, 20-30-year-old young adults and 65-85-year-old speakers. In each age group there were 20 participants (10 male and 10 female). Speech samples were chosen from the GABI Child Language and Speech Database and Information Repository (Bóna et al. 2019) and BEA Hungarian spoken language database (Gósy 2013). Two speech tasks were analysed: spontaneous narratives and narrative recalls. The two speech tasks required different cognitive and speech planning processes. In spontaneous narratives, speakers spoke about themselves and their own lives, and they could use words and grammatical forms of their own choice. In narrative recalls, speakers had to listen to the heard text, they had to understand it, and recall the structure and the linguistic form of the story. This required more mental effort from the speakers. 5 minutes of spontaneous narratives and 2-4 minutes of narrative recalls were analysed from each speaker.

In the speech productions, self-repairs were annotated and error-to-cutoff times (time from the beginning of the reparandum to the interruption of the original utterance), error-to-repair times (time from the beginning of the reparandum to the beginning of the repair), and editing phases were measured. The types of repairs and the repair strategies were also analysed. Finally, the data of age groups and speech tasks were compared.

Results show that both age and speech task influence the occurrence of repair types. Children and old speakers produce less D-repairs than young adults (Figure 1). In addition, strategies of error-repairs are different in the age groups. According to the quantitative analysis of the interruption points and the durational patterns of the repairs, children "prefer accuracy", while young and old adults prefer fluency during error-repairs. These results contribute to a more accurate understanding of the age characteristics of speech production processes.

Figure 1: *The types of repairs in spontaneous narratives*



Acknowledgement: This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. K-120234 and the Thematic Excellence Program.

References

Bóna, J., Tímea, V., & Viola, V. (2019). GABI – Hungarian Child Language and Speech Database and Information Repository. *The Phonetician*, *116,* 41-52.

Duchin, S. W. & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders, 20,* 245–257.

Gósy, M. (2013). BEA – A multifunctional Hungarian spoken language database. *Phonetician*, *105*, 50-61.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America, 128,* 839–850.

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19-35.

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*(3), 837-842.

Yairi, E. & Clifton, N. F. (1972). Disfluent speech behavior of preschool children, high school seniors, and geriatric persons. *Journal of Speech and Hearing Research, 15,* 714–719.

Aging effects on prosodic structuring in French

Anne Hermes, Angélina Bourbon, Cécile Fougeron
Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, France
{anne.hermes, angelina.bourbon, cecile.fougeron}@sorbonne-nouvelle.fr

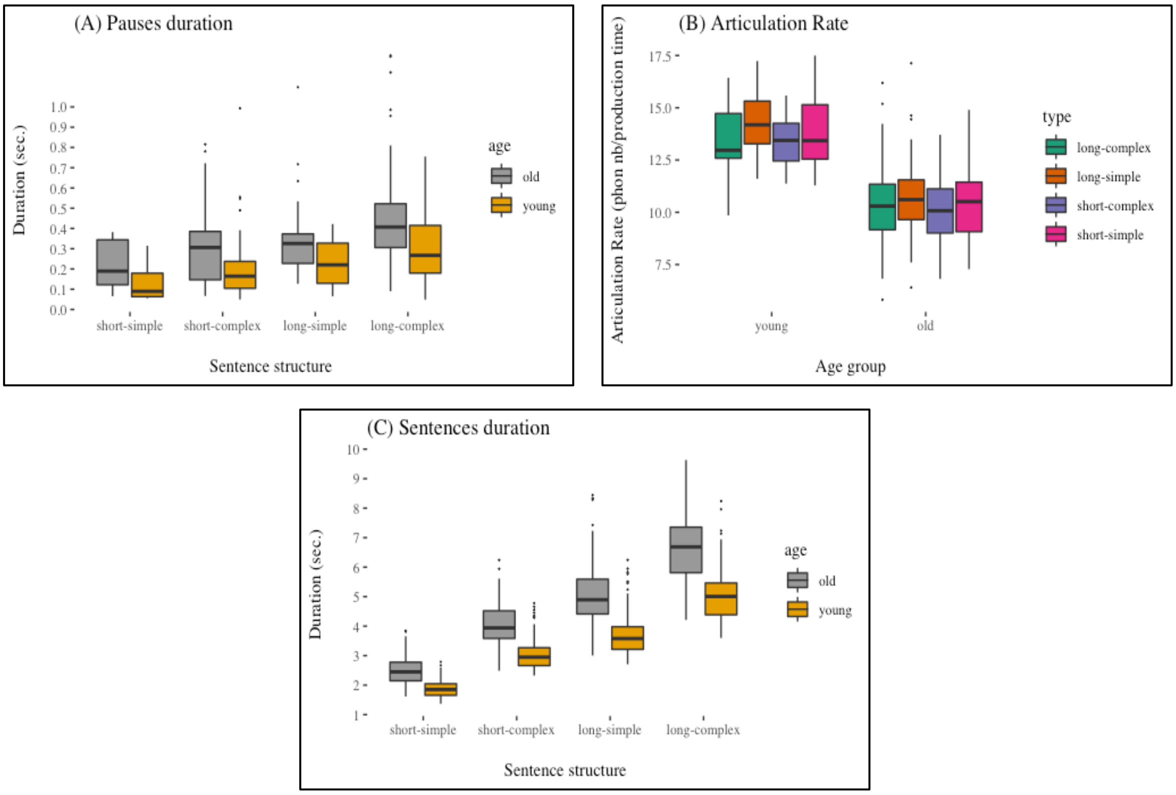
The current study aims to shed light on the effects of aging on prosodic structuring by analysing several durational parameters in sentences differing in length and complexity. This study provides evidence that (i) all age groups produce phrasal breaks at expected syntactic boundaries, (ii) older speakers produced much longer pausal breaks, slower articulation rates and longer sentence durations than the younger ones.

Prosodic structure is essential in speech communication as information needs to be highlighted to convey specific linguistic meaning. This highlighting includes inter alia the planning/structuring and the production of larger prosodic constituents, such as the utterance, and smaller prosodic constituents, such as the word or the syllable. Across life span there is a change in respiration in that with increasing age, the ribcage cannot expand and contract as well during breathing and the diaphragm becomes weakened (Kahane, 1981). This physiological age-related change is expected to affect the control of breathing and thus also prosodic structuring. Studies investigating the effects of aging on speech breathing with respect to prosodic structuring are relatively sparse and limited to English. Evidence has been provided that breathing in older subjects revealed higher lung volume to initiate speech, longer sentences are split into smaller units with less breathing at larger units and more breathing at smaller units (Hoit & Hixon, 1987; Huber, 2008).

We recorded 27 French speakers (18 women and 9 men) aged between 23 and 88 years. The subjects were split into a ‘younger’ (23-44 years) and an ‘older’ group (68-88 years). Speakers produced different sentences which vary in length (long vs. short) and complexity (simple vs. complex). These French sentences were designed on the basis of Fuchs et al. (2013) for German. We recorded 12 different sentences, each repeated three times, i.e. 36 sentences per speaker. The order was randomized. The speech material is presented in Table 1. The annotations of the pauses were done manually in Praat (Boersma & Weenink, 2017), taking as criteria marking of syntactic autonomy with the absence of hesitation as well as other pauses which were not syntactically driven. Pauses were segmented as visual (using the oscillogram and spectrogram) and audible breaks in the speech stream which included inspiratory and physiological phenomena (e.g. breathiness, glottal noise, etc.). The measurement was kept consistent within each speaker; pauses were excluded below 100ms. Analysed variables were as follows: (a) NUMBER OF PAUSES marked as breaks at comma-indicated boundaries and other breaks; (b) PAUSE DURATION; (c) ARTICULATION RATE (syllables/sec) was calculated as the number of syllables divided by the whole sentence duration (excl. pauses) and (d) SENTENCE DURATION is based on the total utterance time including pauses.

First, we looked at the NUMBER OF PAUSES. Both, younger and older speakers, produced clear pauses at the expected, comma-indicated boundaries. Thus, syntactically simple sentences were marked with fewer pauses (young: short=11/long=15; old: short=16/long=71) than complex sentences (young: short=103/long=162; old: short=122/long=164). Especially in the long-simple and short-complex condition older subjects showed a much higher number of pauses. Figure 1A displays the parameter PAUSE DURATION depending on sentence structure and age group: in long-complex sentences, the pauses were longest for both age groups (e.g. pause at first comma: young: short-complex=171ms vs long-complex=228ms; old: short-complex=308ms vs long-complex=420ms). In Winkworth et al. (1994) results showed that lung volume significantly increased before longer utterances compared to shorter utterances, which can be related to what we found acoustically. Additionally, all pause durations were much longer for older than for younger speakers (on average 120ms longer). In order to capture these age-related differences, we analysed ARTICULATION RATE across the different sentence structures and age groups (see Fig. 1B). The results revealed that the speakers with the slowest articulation rate were those who also produced the longest pauses, i.e. the older group. This aging effect was observed for all sentence structures. However, we did not observe an effect for overall articulation rates depending on sentence structure. The lower articulation rate for older speakers in this study seems to be in line with the most robust, reported age-related effect (Ramig, 1983; Verhoeven, 2004). The parameter SENTENCE DURATION, (cf. Fig. 1C) supported the observation made before: the longer and complex a sentence, the longer its duration, with age even more affecting it. Another observation was a higher variability for all measurements across older speakers. This age-related increase in duration variability has been reported before (Smith et al., 1987; Tremblay et al., 2018).

Figure 1: *Results for (A)* PAUSE DURATION, *(B)* ARTICULATION RATE *and (C)* SENTENCE DURATION *depending on age and sentence structure.*



This study applied different durational parameters to shed light on the effect of sentence structure (length and complexity) and age (younger and older group) on prosodic structuring. We were able to provide preliminary evidence that both length and complexity do affect durational patterns in younger and older speakers. A slower speech rate for e.g. long-complex sentences can be explained by increased cognitive demands (Swets et al., 2013). The factor age entailed an additional increase in duration (of e.g. whole sentences, and even pauses). This slowing down could inter alia be explained as a mechanism to make sure to reach the targets properly (Hermes et al., 2018). With respect to cognition, if we consider that working memory is involved in planning speech production, as Swets et al. (2013) do, speakers should store the units to be planned. Thus, these would depend on the speakers' abilities to store them and could explain the need for more breaks/more time for the older speakers. It would be more likely to have smaller prosodic units (e.g. by inserting a higher number of breaks or being 'forced' to perform comma-indicated breaks, see the different results for the number of 'correctly' performed comma-indicated breaks) to ensure specific linguistic meaning. Whether an age-related change in sentence/pause duration and articulation rate for older speakers can be explicitly linked to breathing cannot be solved in this study and will be of main interest in future studies.

Table 1: *Speech material with sentences differing in length (short and long) and complexity (simple and complex).*

Sentences	Structure
Mélanie Dupont a réservé ses vacances. *(Mélanie Dupont booked her holidays.)*	Short and simple
Mélissa Lambert a terminé son dossier. *(Mélissa Lambert finished her file.)*	Short and simple
Mélodie Moreau a préparé son repas. *(Mélodie Moreau prepared her meal.)*	Short and simple
Mélanie Dupont, qui adore le soleil, a réservé ses vacances. *(Mélanie Dupont, who loves the sun, booked her holidays.)*	Short and complex
Mélissa Lambert, qui travaille cet été, a terminé son dossier. *(Mélissa Lambert, who work this summer, finished her file.)*	Short and complex
Mélodie Moreau, qui habite à Montreuil, a préparé son repas. *(Mélodie Moreau, who lives in Montreuil, prepared her meal.)*	Short and complex
Mélanie Dupont a réservé lors d'un froid jour d'hiver ses vacances en Andalousie. *(Mélanie Dupont booked her holidays in Andalusia on a cold winter's day.)*	Long and simple
Mélissa Lambert a terminé la semaine dernière son dossier pour l'Université. *(Melissa Lambert finished her file for the University last week.)*	Long and simple
Mélodie Moreau a préparé durant l'après-midi son repas avec du saumon. (Mélodie Moreau prepared her meal with salmon during the afternoon.)	Long and simple
Mélanie Dupont a dit à sa mère, qui l'a appelée dans la matinée, qu'elle attend avec impatience ses vacances. (Melanie Dupont told her mother, who called her in the morning, she's looking forward to her vacation.)	Long and complex
Mélissa Lambert a dit à son ami, qui l'a appelé dans la soirée, qu'elle est fière d'avoir terminé son dossier. Melissa Lambert told her friend, who called her that night, that she's proud to have completed her file.)	Long and complex
Mélodie Moreau a dit à son père, qui l'appelé sur son portable, qu'elle est en train de préparer son repas. (Mélodie Moreau old her father, who called her on her cell phone, that she's making her dinner.)	Long and complex

This work was partially supported by the Swiss FNS Grant CRSII5_17371/1 as well as by the French ANR Grant ANR-10-LABX-0083.

References

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (version 6.0. 28)[Software].

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, *41*(1), 29-47.

Hermes, A., Mertens, J., & Mücke, D. (2018, September). Age-related Effects on Sensorimotor Control of Speech Production. In *Interspeech* (pp. 1526-1530).

Hoit, J. D., & Hixon, T. J. (1987). Age and speech breathing. *Journal of Speech, Language, and Hearing Research*, *30*(3), 351-366.

Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory physiology & neurobiology*, *164*(3), 323-330.

Kahane, J. C. (1981). Anatomic and physiologic changes in the aging peripheral speech mechanism. *Aging, Communication Process and Disorders*, 21-45.

Ramig, L. A., & Ringel, R. L. (1983). Effects of physiological aging on selected acoustic characteristics of voice. *Journal of Speech, Language, and Hearing Research*, *26*(1), 22-30.

Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech, Language, and Hearing Research*, *30*(4), 522-529.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, *50*(1), 23-51.

Tremblay, P., Deschamps, I., Bédard, P., Tessier, M. H., Carrier, M., & Thibeault, M. (2018). Aging of speech production, from articulatory accuracy to motor timing. *Psychology and aging*, *33*(7), 1022-1034.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and speech*, *47*(3), 297-308.

Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. *Journal of Speech, Language, and Hearing Research*, *37*(3), 535-556.

Referential and inferential production across the lifespan: different patterns and different underlying mechanisms

Raphaël Fargier[1] & Marina Laganaro[2]
[1]Aix-Marseille University, Laboratoire Parole et Langage, Aix-en-Provence, France [2] FPSE, University of Geneva, Switzerland
raphael.fargier@univ-amu.fr

Speaking is a pervasive activity in which we engage throughout our lives. This ability is grounded in general memory and control processes and thus likely displays changes across the lifespan. Referential naming, that is, the ability to produce words in response to pictures, follows a U-shape or bow-shape trend across the lifespan, with worse performance in children and elderly relative to young adults (D'Amico et al., 2001; Laganaro et al., 2015; Newman & German, 2005; Thornton and Light, 2006; Kavé et al., 2010; Valente & Laganaro, 2015). Picture naming has been extensively used as a proxy to language production and thus used to develop theories about the cognitive architecture of language production in healthy human beings, and how it is affected by aging or impaired in pathological conditions (see Indefrey & Levelt, 2004). Picture naming is however not the only way to elicit the production of words, and the mechanisms supporting *inferential* naming ability (e.g. naming from definitions) are far less known.

Referential naming and inferential naming share important features as they both require to select a word from long-term memory and to transform it into articulated speech sounds. Accordingly, word-form encoding is achieved through similar processes in both tasks (Fargier & Laganaro, 2017). However, both tasks also differ in one important respect, which is *how* the word is selected : from a visual input in referential naming and through semantic and/or episodic associations in inferential naming (Fargier & Laganaro, 2017). In naming from definition paradigms, the speaker has to produce a word in response to a given written or oral definition (e.g. "Animal from which we obtain honey" *bee*, "Object that we use to shoot arrows" *bow*), thus the information contained in the definition must be combined in order to retrieve the target word. This great distinction highlights how the two tasks likely rely on different cognitive abilities and which, importantly, may not follow the same trajectory across the lifespan. Here, we sought to characterize the patterns of referential and inferential naming from infancy to elderly, and to identify its underlying cognitive abilities.

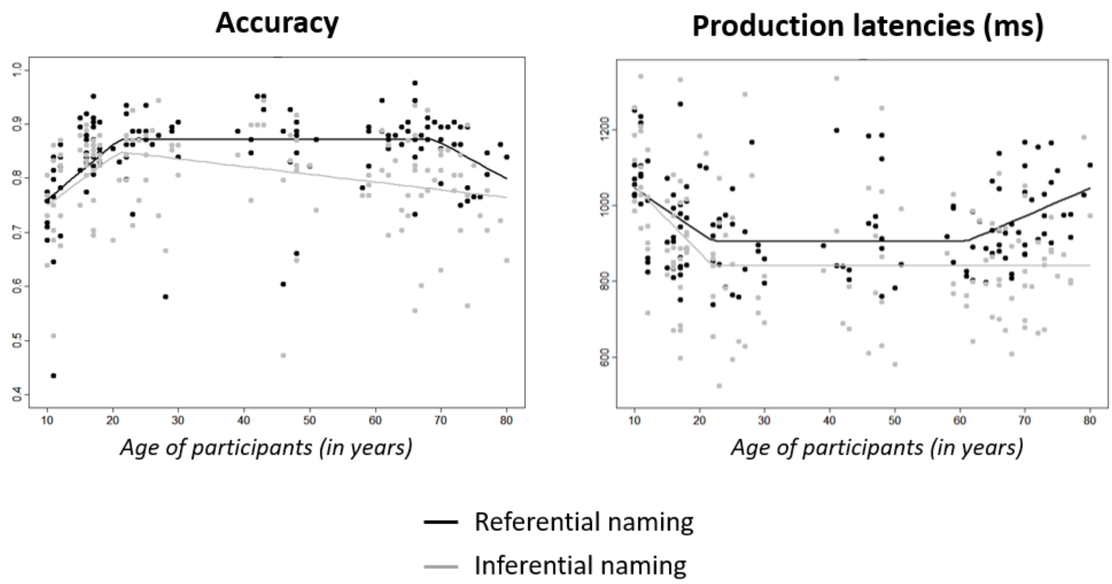
Data were collected in 154 participants ranging from 10 to 80 years old. Participants were spread across defined age-groups including children from 10 to 13 years old (N=21), adolescents from 15 to 18 years old (N=27), young adults from 20 to 30 years old (N=25), middle-aged adults from 40 to 50 years old (N=19), older adults from 58 to 68 years old (N=27) and elderly (>70 years old ; N=23).

Different tasks were conducted: We collected accuracy and production latencies in Referential naming and Inferential naming (i.e. naming from definition) tasks. For Referential naming, we used a picture naming task: 120 black and white drawings and their corresponding modal names were selected from two French databases (Alario & Ferrand, 1999, Bonin et al., 2003). For Inferential naming, we designed a naming from definition task. 108 target words were selected and their corresponding definitions were constructed similarly to previous work (Fargier & Laganaro, 2017) i.e. definitions had similar structure across items and target words could not be anticipated before the last word of the definition was heard. All target words corresponded to words learned before the age of 9, according to the mentioned databases. We also collected measures of Vocabulary size and Digit Span memory (from the Wechsler Adult/Children Intelligence Scale, Wechsler, 2005), Semantic & Phonemic Fluencies (e.g. category names from animals and letter P in two minutes) and Processing speed (Simple and Forced choice Reaction time experiments).

Two main analyses were performed. The first consisted in characterizing the lifespan pattern of Referential and Inferential naming using multivariate adaptive regression splines (package Earth, Milborrow et al. 2013). This approach produces a regression that is allowed to bend at certain knots that mark a change in the behavior of the function. The knots obtained across the lifespan could then be compared between Referential and Inferential naming. The second analysis consisted in multiple regression models with all behavioral outcomes as predictors (under the quadratic assumption) to test for critical effects of these measures on performance across the lifespan.

We found different patterns for Referential naming and Inferential naming with similar increase in performance from childhood to young adulthood, and significant decrement of performance in elderly for Referential naming but not (or rather less evident) for Inferential naming (see Figure 1). Regular analyses of Age effects and non-parametric multivariate adaptive regression splines analyses statistically confirmed this result. Lifespan regressions on accuracy and production latencies showed two inflexion points in Referential naming (around 22 years and around 61 years) reflecting the U-shaped curve. By contrast, the analysis showed only one inflexion point in inferential naming (around 21 years).

Figure 1 : *Lifespan patterns of referential and inferential naming on accuracy and production latencies. Lines correspond to non-parametric multivariate adaptive regression splines obtained with the package Earth. Whereas two knots are observed for referential naming patterns (black lines), only one knot is observed for inferential naming patterns (grey lines).*



Additional analyses used behavioral measures as predictors of language production performance. For accuracy in Referential naming, we observed significant linear (B=0.31, t=3.7, p=0.0003) and quadratic (t=-2.7, p=0.0067) effects of Age, a linear effect of Semantic Fluency (B=0.18, t=2.25, p=0.026) and linear effect of Processing Speed (B=-0.22, t=-2.57, p=0.011). For production latencies, the model indicated significant linear (B=-308, t=-2.5, p=0.01) and quadratic (B=257, t=2.2, p=0.032) effects of Age, linear effect of Semantic Fluency (B=-434, t=-3.67, p=0.0004) and linear effect of Processing Speed (B=459, t=3.7, p=0.0003).

For accuracy in Inferential naming, we found a quadratic effect of Age (B=-0.25, t=-3.1, p=0.002), and significant linear effects of Vocabulary (B=0.28, t=3.6, p=0.0004), of Digit Span (B=0.17, t=2.17, p=0.032), of Semantic Fluency (B=0.2, t=2.45, p=0.016) and of Processing Speed (B=-0.32, t=-3.9, p=0.0002). Finally, for production latencies, we found significant linear effects of Age (B=-860, t=-5, p=1.68^{e}-6), of Semantic Fluency (B=-583, t=-3.54, p=0.0058), and of Processing Speed (B=699, t=4.06, p=8.8^{e}-5) and a marginal effect of Phonemic Fluency (B=-290, t=-1.7, p=0.09).

Overall, these results indicated that variables such as Age, Semantic Fluency and Processing Speed predicted behavior similarly in both tasks: high semantic fluency score and high processing speed related to good naming ability. In inferential naming however, Vocabulary size and Digit Span memory also significantly predicted performance such that good scores were related to better performance in naming.

These results suggest that Referential naming and Inferential naming do not rely entirely on the same mechanisms, and notably that inferential naming likely relies more strenuously on comprehension/memory abilities, which may confer a protective mechanism against cognitive decline. Referential naming by contrast may depend more on control processes that are more sensitive to aging. The approach of combining different language production paradigms constitutes an important step forward to modelling language processing across the lifespan.

References

Alario, F. X. & Ferrand, L. (1999). A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*, 531–552.

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development*, *2*(1), 71-105.

Bonin, P., Peerman, R., Malardier, N., Méot, A., Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers, 35*, 158–167.

Fargier, R. and Laganaro, M. (2017) Spatio-temporal dynamics of referential and inferential naming: different brain and cognitive operations to lexical selection, *Brain Topography,* 30(2), 182-197.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.

Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging*, *25*(3), 719.

Laganaro, M., Tzieropoulos, H., Frauenfelder, U. H., Zesiger, P. (2015). Functional and time-course changes in single word production from childhood to adulthood. *NeuroImage, 111,* 204–214.

Milborrow, S., Hastie, T., Tibshirani, R. (2013). earth: Multivariate Adaptive Regression Spline Models, R package, http://CRAN.R-project.org/package=earth.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech*, *48*(2), 123-156.

Thornton, R., & Light, L. L. (2006). Language comprehension and production in normal aging. In *Handbook of the Psychology of Aging (Sixth Edition)* (pp. 261-287).

Valente, A., Laganaro, M., (2015). Ageing effects on word production processes: an ERP topographicanalysis. *Language, Cognition and Neuroscience, 30:10,* 1259-1272. doi: 10.1080/23273798.2015.1059950.

Wechsler D. (2003) Wechsler Intelligence Scale for Children. 4th Edition. San Antonio, TX: Pearson Assessment; 2008.

Wechsler D. (2008) Wechsler Adult Intelligence Scale. 4th Edition. San Antonio, TX: Pearson Assessment.

Author Index