SPPL2017: Workshop on Speech Perception and Production across the Lifespan

Book of Abstracts

UCL, London, UK, 26-27 April 2017
Division of Psychology and Language Sciences, Chandler House

Sponsored by
Foreword

Although the focus of much research into speech development has been to establish when ‘adult-like’ performance is reached (with young adult speakers taken as a ‘norm’), it is increasingly clear that speech perception and production abilities are undergoing constant change across the lifespan as a result of physical changes, exposure to language variation, and cognitive changes at various periods of our lives. Few studies have examined changes in speech production or perception measures across the lifespan using common materials and experimental designs. Lifespan studies can further our understanding of the extent and direction of these changes for key measures of speech communication and of how these changes interact with cognitive, social or sensory factors. Such knowledge is essential to refine and extend models of speech perception and production.

The workshop will provide an opportunity for interactions between researchers from areas of speech and language sciences research that may be focused on different developmental stages, e.g. early development and ageing. It will also discuss methodological issues, such as how to overcome the difficulty of developing tests that are equally appropriate for children, younger and older adults.

This Book of Abstracts contains 35 accepted papers presented at the oral and poster sessions of the SPPL2017 workshop as well as 4 invited talks and an opening talk.

We are delighted to welcome to London participants from many different countries and would also like to thank all the invited speakers for their participation to this workshop. We hope that this workshop will lead to fruitful exchanges and hopefully even new collaborations. Final thanks must go to members of the SPPL2017 organising committee, to our colleagues who have helped organise this event, to the Economic and Social Research Council (ESRC), UK-SPEECH Group and Division of Psychology and Language Sciences for their financial sponsorship and, of course, to all contributors.

The workshop is organised under the aegis of ESRC project ES/L007002/1 on ‘Speech communication in older adults: an acoustic and perceptual investigation’.

Valerie HAZAN and Outi TUOMAINEN, Co-Chairs SPPL2017

SPPL2017 Organising Committee

Christopher DAVIS (University of Western Sydney)
Bronwen EVANS (UCL)
Valerie HAZAN (UCL) – Co-chair
Esther JANSE (Radboud University Nijmegen/Max Planck institute Nijmegen)
Jeessun KIM (University of Western Sydney)
Merle MAHON (UCL)
Kathleen McCARTHY (UCL)
Caroline NEWTON (UCL)
Outi TUOMAINEN (UCL) – Co-chair
Rosemary VARLEY (UCL)
## PROGRAMME

**WEDNESDAY 26 APRIL 2017**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30-9:20</td>
<td>Registration</td>
</tr>
<tr>
<td>9:20-9:30</td>
<td>Welcome</td>
</tr>
<tr>
<td>9:30-10:15</td>
<td><strong>Opening talk – Valerie Hazan and Outi Tuomainen</strong> (G10) Spontaneous speech production across the lifespan</td>
</tr>
<tr>
<td>10:15-11:15</td>
<td><strong>Oral Session 1 (G10)</strong></td>
</tr>
<tr>
<td></td>
<td>Judit Bóna &amp; Tímea Vakula Disfluent word-repetitions across the lifespan</td>
</tr>
<tr>
<td></td>
<td>Elisa Pellegrino, Lei He, Nathalie Giroud, Martin Meyer, Volker Dellwo Speech rhythm and aging</td>
</tr>
<tr>
<td></td>
<td>Mária Gósy &amp; Valéria Krepsz Stem and suffix durations in words of increasing length across the lifespan</td>
</tr>
<tr>
<td>11:15-11:45</td>
<td><strong>COFFEE BREAK (G15)</strong></td>
</tr>
<tr>
<td>11:45-13:00</td>
<td>Flash presentations and POSTER 1 (G10/B01)</td>
</tr>
<tr>
<td>13:00-13:45</td>
<td><strong>LUNCH (G15)</strong></td>
</tr>
<tr>
<td>13:45-14:40</td>
<td><strong>Keynote 1 – Sandra Gordon-Salant</strong> Factors influencing speech perception in older adults</td>
</tr>
<tr>
<td>14:40-15:40</td>
<td><strong>Oral session 2 (G10)</strong></td>
</tr>
<tr>
<td></td>
<td>Aina Casaponsa, Viktoria Vianeva, Michael Akeroyd, Johanna Barry Listening across the life span: What you have to do determines how you do it</td>
</tr>
<tr>
<td></td>
<td>Sarah Colby, Meghan Clayards, Shari Baum Top-down and bottom-up perceptual learning for speech is maintained in older adults</td>
</tr>
<tr>
<td></td>
<td>Shekeila Palmer &amp; Sven Mattys Statistical learning for speech segmentation: Age-related changes and underlying mechanisms</td>
</tr>
<tr>
<td>15:40-16:10</td>
<td><strong>TEA BREAK (G15)</strong></td>
</tr>
<tr>
<td>16:10-17:05</td>
<td><strong>Keynote 2 – Hayo Terband</strong> Toward a model of speech development and disorders for diagnosis and treatment planning</td>
</tr>
<tr>
<td>17:05-17:45</td>
<td><strong>Oral session 3 (G10)</strong></td>
</tr>
<tr>
<td></td>
<td>Laurianne Cabrera, Christian Lorenzi and Stuart Rosen The development of speech perception in noise: the contribution of temporal-processing</td>
</tr>
<tr>
<td></td>
<td>Chadi Farah, Nuria Esteve-Gibert, Stéphane Roman, Mariapaola D’Imperio Prosodic focus acquisition in French early cochlear implanted children</td>
</tr>
<tr>
<td>19:30 – 22:00</td>
<td><strong>WORKSHOP DINNER (THE PEASANT, 240 St John St, Clerkenwell, London EC1V 4PH)</strong></td>
</tr>
<tr>
<td>Time</td>
<td>Session/Activity</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 9:15-10:10   | **Keynote 3: Mitchell Sommers**  
Listening comprehension across the adult lifespan |
| 10:10-10:50  | **Oral session 4 (G10)**                                                        |
|              | Chris Davis, Jeesun Kim, Outi Tuomainen, Valerie Hazan  
Partner-directed gaze in a communicative task: the effects of age and communication barrier |
|              | Anne Hermes  
Effects of ageing on articulation |
| 10:50-11:20  | **COFFEE BREAK (G15)**                                                           |
| 11:20-12:00  | **Oral session 5 (G10)**                                                        |
|              | Cecile Fougeron, Véronique Delvaux, Cédric Gentrot, Marina Laganaro, Lucie Ménard  
Introducing two databases of spoken French throughout adulthood |
|              | Martin Hagmüller  
Challenges for automatic speech recognition over the life span |
| 12:00-13:15  | **Flash presentations and poster 2 (G10/B01)**                                  |
| 13:15-14:00  | **LUNCH (G15)**                                                                  |
| 14:00-14:45  | **Panel session on challenges of lifespan studies (G10)**                        |
| 14:45-15:40  | **Keynote 4: Paul Foulkes**  
Sociophonetic perspectives on first language acquisition |
| 15:40-16:10  | **TEA BREAK (G15)**                                                              |
| 16:10-17:10  | **Oral session 6 (G10)**                                                        |
|              | Bronwen Evans, Lidia Martin-Alvarez, Elizabeth Wonnacott  
Age-related differences in second-language learning? |
|              | Susanne Brouwer, Simone Sprenger, Sharon Unsworth  
Processing grammatical gender in Dutch: Evidence from eye movements |
|              | Elizabeth Wonnacott, Anastasia Giannakopoulou, Helen Brown, Meghan Clayards  
High or Low? Comparing high- and low-variability phonetic training in adult and child second-language learners |
| 17:10-17:20  | **Final discussion and closing comments (G10)**                                  |
### WEDNESDAY 26 APRIL: FLASH PRESENTATIONS AND POSTER SESSION 1

<table>
<thead>
<tr>
<th>Presenter(s)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anita Auszmann</td>
<td>Phonological development of vowels in Hungarian children</td>
</tr>
<tr>
<td>Gwen Brekelmans, Elizabeth Wonnacott, Bronwen G. Evans</td>
<td>The effect of variability on phonetic training for adults and children</td>
</tr>
<tr>
<td>Dorottya Gyarmathy</td>
<td>Functions of silent pauses in spontaneous speech</td>
</tr>
<tr>
<td>Kathleen McCarthy, Katharine Mair, Katrin Skoruppa, Paul Iverson</td>
<td>Development of speech processing in infancy and its relationship with later language skills</td>
</tr>
<tr>
<td>Joan C. Mora</td>
<td>Individual differences in L2 bimodal input processing</td>
</tr>
<tr>
<td>Leanne Nagels, Roelien Bastiaanse, Deniz Başkent, Anita Wagner</td>
<td>The processing of lexical and contextual information in cochlear implant users</td>
</tr>
<tr>
<td>Tilda Neuberger</td>
<td>Perception boundaries between single and geminate stops in Hungarian children and adults</td>
</tr>
<tr>
<td>Nikola Paillereau &amp; Naomi Yamaguchi</td>
<td>Acquisition of L1-French corner vowels and formant visualisation in the VisuVo software: a longitudinal case study</td>
</tr>
<tr>
<td>Melissa Redford</td>
<td>Orthogonal development of schwa coarticulation and reduction in DET+N sequences?</td>
</tr>
<tr>
<td>Miriam S. Tenderini, Esther de Leeuw, Tiina M. Eilola, Marcus T. Pearce</td>
<td>Bilingual experience of emotional meaning: priming effects between music and the first and second language</td>
</tr>
</tbody>
</table>

### THURSDAY 27 APRIL: FLASH PRESENTATIONS AND POSTER SESSION 2

<table>
<thead>
<tr>
<th>Presenter(s)</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarah Colby, Victoria Poulton and Meghan Clayards</td>
<td>Inhibition predicts lexical competition in older adults’ spoken word recognition</td>
</tr>
<tr>
<td>Viktória Horváth, Viola Váradi, Judit Bóna</td>
<td>Characteristics of filled pauses across the lifespan</td>
</tr>
<tr>
<td>Ella Jeffries</td>
<td>Developments in accentpause between the pre-school and primary school years and the effects of exposure to variation</td>
</tr>
<tr>
<td>Holly Kennard</td>
<td>Production and perception of stress in ‘traditional’ and ‘new’ speakers of Breton</td>
</tr>
<tr>
<td>Valéria Krepsz &amp; Mária Gósy</td>
<td>Auditory-phonetic processing in word recognition across ages</td>
</tr>
<tr>
<td>Anabella-Gloria Niculescu-Gorpin &amp; Monica Vasileanu</td>
<td>Perception and processing of context-dependent anglicisms across the age groups</td>
</tr>
<tr>
<td>Rosie Oxbury &amp; Esther de Leeuw</td>
<td>Style-shifting and its acoustic correlates in a cohort of 11 year old girls from East London</td>
</tr>
<tr>
<td>Stefanie Ramachers, Susanne Brouwer, Paula Fikkert</td>
<td>Native and non-native perception of lexical pitch patterns across the lifespan</td>
</tr>
<tr>
<td>Josiane Riverin-Coutlée &amp; Vincent Arnaud</td>
<td>Regional variation in young adult and older speakers of Quebec French: A pilot acoustic study</td>
</tr>
<tr>
<td>Adrian P. Simpson</td>
<td>Are ejectives on the increase in English? Evidence from a single speaker</td>
</tr>
</tbody>
</table>
# Table of Contents

**Opening talk**

Spontaneous speech production across the lifespan

*Valerie Hazan and Outi Tuomainen*

<table>
<thead>
<tr>
<th>Oral session 1</th>
</tr>
</thead>
</table>
| Disfluent word-repetitions across the lifespan

*Judit Bona and Timea Vakula* |

| Speech rhythm and aging

*Elisa Pellegrino, Lei He, Nathalie Giroud, Martin Meyer and Volker Dellwo* |

| Stem and suffix durations in words of increasing length across the lifespan

*Maria Gosy and Valeria Krepsz* |

**Poster session A**

| Phonological development of vowels in Hungarian children

*Anita Auszmann* |

| The effect of variability on phonetic training for adults and children

*Gwen Brekelmans, Elizabeth Wonnacott and Bronwen G. Evans* |

| Functions of silent pauses in spontaneous speech

*Dorottya Dr. Gyarmathy* |

| Development of speech processing in infancy and its relationship with later language skills

*Kathleen McCarthy, Katharine Mair, Katrin Skoruppa and Paul Iverson* |

| Individual differences in L2 bimodal input processing

*Joan C. Mora* |

| The processing of lexical and contextual information in cochlear implant users

*Leanne Nagels, Rochen Bastiaanse, Deniz Baskent and Anita Wagner* |

| Perception boundaries between single and geminate stops in Hungarian children and adults

*Tilda Neuberger* |

| Acquisition of L1-French corner vowels and formant visualisation in the VisuVo software: a longitudinal case study

*Nikola Paillerou and Naomi Yamaguchi* |

| Orthogonal development of schwa coarticulation and reduction in DET+N sequences?

*Melissa Redford* |

| Bilingual experience of emotional meaning: priming effects between music and the first and second language

*Miriam S. Tenderini, Esther de Leeuw, Tiina M. Eilola and Marcus T. Pearce* |

**Keynote 1**
Factors influencing speech perception in older adults .......................... 40
    Sandra Gordon-Salant

Oral session 2

Listening across the life span: What you have to do determines how you do it .......... 41
    Aina Casaponsa, Viktoria Vianeva, Michael Akeroyd and Johanna Barry
Top-down and bottom-up perceptual learning for speech is maintained in older adults .... 44
    Sarah Colby, Meghan Clayards and Shari Baum
Statistical learning for speech segmentation: Age-related changes and underlying
    mechanisms ......................................................... 47
    Shekeila Palmer and Seen Mattys

Keynote 2

Toward a model of speech development and disorders for diagnosis and treatment planning 50
    Hayo Terband

Oral session 3

The development of speech perception in noise: the contribution of temporal-processing .. 53
    Laurianne Cabrera, Christian Lorenzi and Stuart Rosen
Prosodic Focus Acquisition in French Early Cochlear Implanted Children .................. 54
    Chadi Farah, Nuria Esteve-Gibert, Stephane Romain and Mariapaola D’Imperio

Keynote 3

Listening comprehension across the adult lifespan ........................................ 57
    Mitchell Sommers

Oral session 4

Partner-directed gaze in a communicative task: the effects of age and communication
    barrier .............................................................. 58
    Chris Davis, Jeeseun Kim, Ouli Tuomainen and Valerie Hazan
Effects of ageing on articulation .......................................................... 61
    Anne Hermes

Oral session 5

Introducing two databases of spoken French throughout adulthood .................... 64
    Cecile Fougeron, Veronique Delvax, Cedric Gendrot, Marina Laganaro and Lucie
    Ménard
Challenges for Automatic Speech Recognition over the Life Span ....................... 66
    Martin Hagmueller

Poster session B

viii
Inhibition predicts lexical competition in older adults’ spoken word recognition ......... 69
Sarah Colby, Victoria Poulton and Meghan Clayards

Characteristics of filled pauses across the lifespan .............................................. 72
Viktória Horváth, Viola Váradi and Judit Bóna

Developments in accent perception between the pre-school and primary school years and
the effects of exposure to variation ........................................................................ 75
Ella Jeffries

Production and perception of stress in traditional’ and new’ speakers of Breton ........... 78
Holly Kennard

Auditory-phonetic processing in word recognition across ages .............................. 81
Valéria Krepzs and Mária Gósy

Perception and Processing of Context-dependent Anglicisms across the Age Groups.... 84
Anabella-Gloria Niculescu-Gorpin and Monica Vasileanu

Style-shifting and its acoustic correlates in a cohort of 11 year old girls from East London
Rosie Ozbury and Esther de Leeuw

Native and non-native perception of lexical pitch patterns across the lifespan .......... 90
Stefanie Ramachers, Susanne Brouwer and Paula Fikkert

Regional variation in young adult and older speakers of Quebec French: A pilot acoustic
study ..................................................................................................................... 93
Josiane Riverin-Coutlée and Vincent Arnaud

Are ejectives on the increase in English? Evidence from a single speaker .............. 96
Adrian P. Simpson

Keynote 4

Sociophonetic perspectives on first language acquisition ....................................... 99
Paul Foulkes

Oral session 6

Age-related differences in second-language learning? A comparison of high and low
variability perceptual training for the acquisition of English /i/-/i/ by Spanish adults
and children. .........................................................................................................102
Browen Evans, Lidia Martin-Alvarez and Elizabeth Wonnacott

Processing grammatical gender in Dutch: Evidence from eye movements ............. 105
Susanne Brouwer, Simone Sprenger and Sharon Unsworth

High or Low? Comparing high- and low-variability phonetic training in adult and child
second language learners ................................................................................. 108
Elizabeth Wonnacott, Anastasia Giannakopoulou, Helen Brown and Meghan Clayards
Spontaneous speech production across the lifespan
Valerie Hazan and Outi Tuomainen
Department of Speech, Hearing and Phonetic Sciences, UCL, UK
v.hazan@ucl.ac.uk; o.tuomainen@ucl.ac.uk

The acoustic-phonetic characteristics of speech change across the lifespan due to physical changes in the vocal apparatus, the effect of linguistic experience as well as of cognitive and sensory factors. However, with the exception of a few studies such as Jacewicz et al. (2010), there have been few investigations of speech production carried out across the whole lifespan using common materials, experimental design and speech analysis methods. Also, studies to date have tended to be based on the use of either read materials or spontaneous monologues which are not produced with true communicative intent. The particular speech style that is employed in those conditions is likely to show much less dynamic adaptations than speech produced in interaction with one or more interlocutors, as, in communicative situations, speech production is guided by how well our speech is being understood by our conversational partners (Lindblom, 1990; Garrod and Pickering, 2004).

In a series of three major projects, we have investigated the acoustic-phonetic characteristics of speech produced in good and adverse listening conditions while talkers resolve a picture-based problem-solving task (Diapix: van Engen et al., 2010) with a conversational partner. The first study analysed the speech of 40 young adults (Hazan and Baker, 2011), the second, the speech of 96 children aged 9 to 15 years (e.g., Hazan et al., 2016; Pettinato et al., 2016) and the current project has analysed the speech of 57 older adults aged 65 to 85 years, with 26 further younger adult controls (Tuomainen et al., 2016). Each of these projects has led to the creation of large speech corpora (LUCID, kidLUCID and the forthcoming elderLUCID) which are in the public domain and which each contain many hours of spontaneous speech interactions. The lengthy processing of these corpora involves manual or automatic orthographic transcription, automatic alignment, manual checking of these alignments and the use of Praat scripts to obtain a number of acoustic-phonetic measures. These measures include articulation rate, fundamental frequency mean and range, relative intensity (representing spectral tilt) and segmental measures of vowel space.

In this talk, I will review the main findings of these three major studies. Although the experimental design and conditions varied to an extent across studies, two main questions could be addressed across the age groups. The first was how the characteristics of spontaneous speech produced in good listening conditions varied as a function of talker age and sex. For articulation rate (Figure 1), for example, an analysis of the speech of 198 talkers from the 9-10 year, 11-12 year, 13-14 year, young and older adult age groups revealed main effects of talker sex \([F(1,188)=12.7; p<.001]\) with male talkers \((M=4.10 \text{ syl/sec})\) speaking at a faster rate than female talkers \((M=3.86)\). There was also a main effect of age group \([F(4,188)=18.6; p<.001]\); post-hoc tests showed that the speech of the older adult (OA) group \((M=3.60)\) was slower than all age groups except for the 9-10 year group \((M=3.77)\). Young adults (YA) \((M=4.32)\) did not differ in articulation rate from 11-12 \((M=4.08)\) and 13-14 years groups \((M=4.11)\).
The second question was how talkers at different ages varied in their ability to adapt their speech to counter the effects of interference affecting their conversational partner. Our study with young adults showed that the adaptations made by individual talkers were, to an extent, dependent on the type of interference that was affecting their interlocutor (babble noise or vocoded speech), even though the talkers that we were analysing were not directly hearing the interference (Hazan and Baker, 2011). This suggests that talkers used the direct or indirect feedback from their interlocutors during the interaction to attune their adaptations. Our study with children showed that they too made adaptations under similar conditions, although younger children at least had a tendency to use a strategy of increasing vocal effort (as shown by strong correlations between increases in fundamental frequency and decreases in spectral tilt) rather than using more varied strategies favoured by adults (Hazan et al., 2016). Our ongoing study with older adults is showing a similar trend: older adults with age-related hearing loss tend to increase vocal effort to counter the effects of adverse conditions (again as shown by significant correlations between spectral tilt and fundamental frequency changes) while older adults with normal hearing thresholds and younger adults do not show this tendency. Our analyses are also revealing some differences in the way that men and women adapt their speech in difficult listening conditions.

Although our three major studies used the same task and experimental setting to elicit spontaneous speech in interaction from children, young and older adults, they do not yet constitute true ‘lifespan’ studies. The running of sequential studies over close to a ten-year period means inevitable changes in experimental design that affect the comparisons that can be made across all age groups. Also, in common with many studies of speech production, no data is included for the 25 to 65 year age range despite the fact that significant changes may occur in this period because of changes in language
experience, health issues, etc. There are many challenges involved in designing true lifespan studies of either speech production or perception but these are essential to truly model the impact of talker age and sex on speech communication.

Acknowledgement
The research reported here was funded by three grants from the UK Economics and Social Research Council (RES-062-23-0681, RES-062-23-3106, ES/L007002/1). We acknowledge the contribution of: Rachel Baker, postdoctoral researcher on the project with young adults, Michele Pettinato, postdoctoral researcher on the project with child speakers, Jeesun Kim and Chris Davis (MARCS Institute, Australia) as co-investigators and Doug Brungart, Jay Desloge and Ben Sheffield (Walter Reed, USA) as collaborators on the project involving older adults.

References


Disfluent word-repetitions across the lifespan
Judit Bóna and Tímea Vakula
Department of Phonetics, Eötvös Loránd University, Hungary

Introduction
During spontaneous speech, speakers often produce disfluencies. One of the most frequent disfluencies is word-repetition (Shriberg 1995; Branigan–Lickley–McKelvie 1999). Repetitions can occur at different levels of the speech planning process: they can indicate word-finding problem, difficulty in conceptual planning, or covert self-monitoring (Plauché–Shriberg 1999).

The main parts of the repetitions are the following: the original utterance, the first instance of the repeated word, the second instance of the repeated word and the continuation of the utterance (Plauché–Shriberg 1999). Optional pauses may also occur next to the main parts (Plauché–Shriberg 1999).

The duration of the first and second instances of the repeated words and the occurrence of pauses are related to the function of the repetitions. Heike (1981) suggested two types of repetitions: bridging a gap and hesitating. Plauché and Shriberg (1999) suggested more other functions according to the phonetic features of the repetitions. They found three main types of functions: canonical repetition, covert self-repair, and stalling repetition (Table 1). Canonical repetition corresponds to Heike’s retrospective repeat, while stalling repetition to prospective repeat.

Table 1: The structures of the three types of word-repetitions (examples with ‘the’) ‘+’ = a longer than fluent duration. ‘- ’ = no pause. (Based on Plauché–Shriberg 1999)

<table>
<thead>
<tr>
<th>Type</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical repetition</td>
<td>(Original Utterance) (Possible Pause) the+++ (Long Pause) the (-) (Continuation)</td>
</tr>
<tr>
<td>Covert self-repair</td>
<td>(Original Utterance) (Often Pause) the+ (-) the+ (-) (Continuation)</td>
</tr>
<tr>
<td>Stalling repetition</td>
<td>(Original Utterance) (-) the+ (Possible Pause) the+++ (Possible Pause) (Continuation)</td>
</tr>
</tbody>
</table>

The characteristics of repetitions (like other disfluencies) are influenced by several factors. Speakers’ age is assumed to be one of these factors. The aim of this study is to analyse temporal features, functions and other characteristics of the repeated words in diverse age groups (from preschool children to elderly speakers).

Our hypotheses are that (i) across the lifespan, the temporal factors of repetitions change: a) the pauses between the two instances of the repeated words and b) the ratio of the duration of the two repeated instances. ii) The ratio of the diverse functions differs between the age groups. iii) All speakers mostly repeat function words but the ratio of the repetitions of content words is higher in children and senior speakers’ speech than in that of young adults.

Methods
For this study, spontaneous speech recordings of 140 speakers were selected from two Hungarian speech databases. Speech samples of preschoolers (4–5-year-olds), school children (9-year-olds), and adolescents (13–14-year-olds) were selected from the GABI Hungarian Children Speech Database and Information Repository (Bóna et al. 2014). Speech samples of young adults (20–25-year-olds), middle-aged adults (45–53-year-old), young-old (60–65-year-olds), and old-old (75+year-olds) speakers were selected from BEA Hungarian Speech Database (Gósy 2012). In every age group there
were 20 speakers (10 women and 10 men). They were native Hungarian speakers with normal hearing and without any known mental or speech disorders.

Participants were asked to speak about their own lives, hobbies and families. From the 30-45-minute-long recordings on average, we analysed 7-8 minutes from each speaker. The duration of the components of repetitions (pauses and first and second instances of repeated words) were measured by Praat. The time interval between two utterances was considered as pause irrespective of its duration, and whether it was silent or filled with sound (not words). We analysed the ratio of the first instance of the repeated word and the second instance of the repeated word and the pauses before, between and after them. Functions and the types of repeated words (function word or content word) were examined, too.

Results
The preliminary results confirmed our hypothesis on age-dependent properties of the repetitions in relation to both functions and temporal patterns. There were significant differences both in the pauses and in the ratio of the duration of the repeated instances between certain age groups. Children produced longer pauses between the repeated instances than the other age groups did, while there was no significant difference in this parameter between young adults and older speakers (Figure 1). Functions of repetitions varied among the age groups. Young and middle-aged adults produced canonical repetitions in higher ratio than the other groups, while children and elderly (both 60-65 and 75+) speakers produced more stalling repetitions than the young and middle-aged groups. Children repeated content words in a higher ratio than the other age groups. The differences were more explicit between young children and adults or young adults and elderly speakers than between adolescents and young children or young adults or between young-old speakers and old-old speakers.

Figure 1: Duration of pauses between the first and the second instances of the repeated words in four age groups

Conclusions
Our findings lead to the conclusions that (i) disfluent word-repetitions are good predictors for detecting some age-dependent changes in speech production process reflected by repetitions, (ii) children’s use of repetitions differ from the adults’ ones in a number of properties demonstrating a developing speech planning mechanism, (iii) elderly speakers’ repetitions differ from young
speakers’ ones in function which shows their cognitive changes and different speech planning processes.

References

This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. K-120234.
Speech rhythm and aging

1Elisa Pellegrino, 1Lei He, 2Nathalie Giroud, 1Martin Meyer and 1Volker Dellwo
1Department of Computational Linguistics, University of Zürich, Switzerland
2Department of Psychology, Neuroplasticity and Learning in the Normal Aging Brain, University of Zürich, Switzerland
*elisa.pellegrino@uzh.ch

The changes in speech production from early childhood to old age alter the way individuals speak both segmentally and suprasegmentally (Tompkins, Scharp & Meigh, 2006; Scholtz, 2007). Altered vowel formant frequency patterns, increased jitter and shimmer, increased breathiness, lengthening of vowels and stop consonants, reduced speech rate, decline in amplitude stability (at least in men), and variation in pitch height are the acoustic features typically associated with adult speakers’ voice (Ramig & Ringel, 1983; Amerman & Parnell, 1992; Linville, 2004).

Does aging also alter the rhythmic characteristics of speech? There is so far no definite answer to this question. While there is evidence that pathological aging, associated for example with dysarthrias, determines considerable variation in segment duration and amplitude envelope characteristics between healthy subjects and patients (Liss et al. 2009; Liss, LeGendre & Lotto, 2010; Pettorino, Pellegrino & Busà, 2016), far less research has been done on rhythmic variations due to healthy aging. Preliminary longitudinal and cross-sectional studies on Italian, quantifying rhythm in terms of segmental durational characteristics, showed that the vocalic portion of the utterance (%V) and the interval between two consecutive vowel onset points (VtoV) vary significantly between elderly and young adults (Pettorino & Pellegrino, 2014; Pettorino, Pellegrino & Maffia, 2014).

In the present study we further explored age-related rhythmic variations, analyzing the segmental durational variability of a corpus of speech produced by 60 Zurich German speakers ranging in age between 18 and 81 years. All speakers read 90 Zurich German sentences. The whole corpus was automatically segmented, using Munich Automatic Segmentation System (MAUS), and then annotated on different tiers: segments, consonantal and vocalic intervals. Durational variability across age was quantified through a variety of different rhythmic variables (Ramus, Nespor & Mehler, 1999; Grabe & Low, 2002, Dellwo, 2006): the proportion over which speech is vocalic (%V), standard deviation of consonantal and vocalic intervals (deltaC and deltaV), variation coefficient of consonantal and vocalic intervals (VarcoC and VarcoV), raw pairwise variability index of consonantal intervals (rPVI-C), normalized pairwise variability index of vocalic intervals (nPVI-V) and segment rate (segment/second).

Given that the participants belonged to two main age groups, namely 18-30 and 66-81, we studied rhythmic differences between these groups. To understand which measures separate more consistently the two age groups, we ran mixed-effect models, in which the above mentioned rhythm measures were taken as dependent variables, age group (‘young’ and ‘old’) was entered as fixed factor, while speakers and sentences as random factors.

Preliminary results based on a subset of the corpus described above (520 sentences = 26 subjects (16 young and 10 old) * 20 sentences), manually corrected after the automatic segmentation, have shown that there is a significant effect of age group for the percentage over which speech is vocalic (%V) as well as for the consonantal and vocalic variability measures (deltaC, deltaV, rPVI-C) (table 1).
Table 1: Results from Mixed-effect models for segment rate and four rhythm measures

<table>
<thead>
<tr>
<th>Rhythm measure</th>
<th>df</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Segment rate</td>
<td>4, 5</td>
<td>22.379</td>
<td>0.001</td>
</tr>
<tr>
<td>2. %V</td>
<td>4, 5</td>
<td>21.094</td>
<td>0.001</td>
</tr>
<tr>
<td>3. delta C</td>
<td>4, 5</td>
<td>7.0111</td>
<td>0.01</td>
</tr>
<tr>
<td>4. delta V</td>
<td>4, 5</td>
<td>31.286</td>
<td>0.001</td>
</tr>
<tr>
<td>5. r-PVI-C</td>
<td>4, 5</td>
<td>6.832</td>
<td>0.01</td>
</tr>
</tbody>
</table>

As shown in figure 1 (left), the subjects aged over 65 spoke at significantly slower rate than the group of young adults, but paused similarly to them. Both groups, indeed, typically read sentences without silent pauses, except for complex sentences for which old speakers always produced a silence between the main and the subordinate clauses. Old speakers also presented higher %V (fig. 1 right), as well as an increase in the durational variability of vocalic and consonantal intervals, both when the variability was calculated on the whole utterance (fig. 2 left and centre) and when it was computed between two consecutive intervals (r-PVI-C) (fig. 2 right).

Figure 1: Boxplots of speech rate in segment/sec (left) and vowel percentage (right) for age group

Figure 2: Boxplots of vocalic durational variability in deltaV (left), consonantal durational variability in deltaC (centre) and in r-PVI-C (right) for age group

Conversely, no significant differences were found between the two-age groups for the consonantal and vocalic variability measures which were normalized for speech rate (VarcoC, VarcoV, n-PVI-V) (table 2).

Table 2: Results from mixed-effect models for rate normalized rhythm measures

<table>
<thead>
<tr>
<th>Rhythm measure</th>
<th>df</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Varco C</td>
<td>4, 5</td>
<td>0.6397</td>
<td>0.423</td>
</tr>
<tr>
<td>2. Varco V</td>
<td>4, 5</td>
<td>1.9366</td>
<td>0.164</td>
</tr>
<tr>
<td>3. n-PVI-V</td>
<td>4, 5</td>
<td>0.2136</td>
<td>0.643</td>
</tr>
</tbody>
</table>
Changes in the physical characteristics as well as in the neural control mechanisms of the articulators might play a significant role in the observed age-related rhythmic change. The slowing of speech rate, (rate segment), the increase in vowel duration (quantified in %V) as well as in consonantal and vocalic variability (deltaC, r-PVI-C, deltaV), indeed, might relate to the generalized slowing in motor function in both the peripheral and central nervous system, as well as to the degenerative changes in the laryngeal and supra-laryngeal systems (i.e. atrophy of laryngeal, pharyngeal, mastication and facial muscles, degeneration of the temporomandibular joints) (Welford, 1990; Schotz, 2007) that inevitably alter the anatomy and hence the movements of the articulators, responsible for the rhythmic organization of speech (Dellwo, Leeman & Kolly, 2015).

References
Introduction
Duration is a physical parameter of spoken language in that words exist in time. There is ample reason from previous research to suppose that duration is considered to reflect two interrelated properties of language: word length and lexical frequency (Greenberg et al. 2003). A great many studies discussed various factors that influence the duration of a word and the variability therein (Losiewicz 1995; Bell et al. 2009; etc.) including the effect of the increasing number of syllables within a word on the duration of the syllables. Individual syllables in a word become shorter as word length increases (e.g., Lehiste 1972; Tily et al. 2009). Age is also acknowledged to be of considerable importance when spoken words are considered. Young children and elderly speakers were reported to show longer durations for words than groups of young adults in several studies (e.g., Smith 1992; Kent 2000). The differences are explained by changes in the effectiveness of speech motor control, which is underdeveloped in children and declines in the elderly. With advancing age, speech changes in its precision, fluency, and communicative effectiveness. For words, produced in spontaneous utterances across ages, the temporal interrelations of stems and suffixes might carry cue information in an agglutinating language about the speakers’ lexical access, speech planning of the articulation of lexemes as well as about understanding the mechanisms underlying the production of spoken words.

In this study we seek to explore the internal temporal patterns of the words of various lengths across the lifespan (in terms of a cross-linguistic analysis from the age of 7 to 80). The core question of this study is whether there is a morphologically conditioned shortening of stems and suffixes in Hungarian, a language with a rich morphology, on the one hand, and whether this shortening phenomenon exists across ages, on the other. Our current hypotheses are that (i) reduction of stems (e.g., ház ‘house’) and suffixed words (e.g., házban ‘in house’) will occur as word length increases (equalization tendency), (ii) this reduction will show similarities between children and elderly people and will be different between them and young speakers, (iii) suffixes will not show durational changes irrespective of word length and age.

Methodology
Seventy speakers were selected to form seven groups (two groups of children /mean ages: 7 and 14 years/, teenagers /17 years/, young adults /25 years/, middle-aged adults /50 years/, young old adults /65 years/, and old adults /80 years/). Each group consisted of 10 speakers (with an equal number of males and females). More than 46 hours of Hungarian spontaneous speech material (from the BEA Hungarian speech database, see Gósy 2012 and the GABI Hungarian children’s database, see Bóna et al. 2014) was carefully hand-labelled using Praat (Boersma–Weenink 2014). Subjects were asked to speak about their family, life, hobby. Stems, suffixes and suffixed words were marked by one of the authors while the other author checked each word (with an agreement ratio of 98%). The word boundaries (between acoustically distinct regions in the signal) were identified in the waveform signal and spectrogram display via continuous listening to the words.

Suffixed verbs and nouns with similar distribution (about 9,000 items) were selected according to the following criteria: (i) stems consisted of various numbers of syllables from 1 to 5, together with suffix syllables to 2 to 6, (ii) 5 frequent monosyllabic suffixes (-ban/-ben ‘in’, -nak/-nek ‘for’, -val/-vel ‘with’, -tam/-tem ‘1sg past’, and -nak/-nek ‘3pl’ were selected that indicated grammatical relationships, (iii) all suffixes were the last syllables of the words, (iv) all words occurred in the middle of phrases (in order to avoid phrase-final lengthening), (v) the suffixes occurred in similar ratios across stems and speakers. Durations of both the stems and suffixes were taken by measuring
the duration between the onsets and the offsets of the stems, the suffixes and the whole words according to common acoustic-phonetic procedures. A specific script was written to obtain the values automatically. All data were transformed into z-scores using the mean and the standard deviation calculated for each speaker to eliminate the influence of variance in speech rates of speakers.

To test statistical significance, linear regression analysis, repeated measures ANOVA and the Mann–Whitney test were used, as appropriate (using SPSS 19.0 version). Measured durations of stems, suffixes, and suffixed words were dependent variables while number of syllables of the stems, word class, age and gender were the independent factors. The confidence level was set at the conventional 95%.

Results
As expected, the longest articulation of the suffixed words irrespective of word length were found with the 7-year olds followed by the old speakers while the fastest articulation was produced by the young speakers. Duration of the suffixed words (range of mean values: 420–1460 ms) showed significant differences depending on both word length and age. No significant differences were found, however, between 14-year-olds and teenage speakers. The increasing number of syllables in the words had an effect on the duration of the syllables. Statistical analysis confirmed significant differences in the reduction of words depending on age. Speakers aged between 14 and 25 years produced words containing more than two syllables substantially shorter than all the other speakers. The length of words produced by middle-aged and young old speakers showed a slight reduction: their measured values fell between those of young and old speakers. 7-year-olds and old speakers articulated suffixed words with practically no reduction.

Durations of the stems (range of mean values: 180–1240 ms) also showed significant differences depending on both the number of syllables and age (Fig. 1). Monosyllabic stems are produced with similar durations across ages with the exception of 7-year-olds and old speakers. As the length of stems increases the durational differences in the production of the stems become significantly different across ages. The largest differences in the durations of both stems and suffixed words were found in those containing 5 and 6 syllables, respectively. The factor of ‘word class’ and ‘gender’ had no effects on durations in either stems or suffixed words.

Figure 1: Durations of stems across syllables and ages
Suffixes had an average duration of about 200 ms across ages with the only exception of 7-year-olds whose mean value was found to be 370 ms. Significant differences were found between children and all the other speakers on the one hand, and old speakers and all the other speakers, on the other. Suffix durations were very similar between ages 14 and 50. The length of stems did not have any significant effect on suffix durations.

Conclusions
This study revealed significant differences in the durations of stems and suffixed words depending both on word length and age. The equalization tendency of the duration of the syllables, however, showed more complex patterns that had been assumed. No significant differences were found in this study in the durations of suffixes after age 14. Our hypotheses were basically confirmed. The measured changes in the reduction of syllables across ages seem to be closely interrelated with the changing nature of speech motor control depending on age (Kent 2000). Immature speech motor control (particularly over phonetic and phonological encoding) might explain the lack of syllable reduction of words in 7-year-olds, while declines of speech motor control (particularly over lexical access) seems to be responsible for a similar phenomenon in old speakers. However, speech motor control does not seem to explain the relatively stable durations of suffixes across the lifespan. Invariable duration in suffixes may be supposed to be a consequence of the agglutinative character of Hungarian, in which coordination of stems and suffixes takes place relatively early in the process of language acquisition.

Our findings will also be discussed from the perspective of information-smoothing theory (e.g., Kuperman–Bresnan 2012), which might provide further explanation for the greater stability found in the articulation of suffixes. We think that it is safe to say that the measured discrepancy in durations between stems and suffixes is a consequence primarily of language-specific traits satisfying the demands of both the speakers and the listeners. Finally, we conclude that word duration is influenced by stem length, and reduction tendency in word production does not apply to all age groups.

References
The children gradually acquire their mother tongue during the first few years. In their speech at first quantitative development (the development of speech is very impressive and fast in the first three years then after six years mainly qualitative changes can be observed (Gósy 2005). The investigation of vowels during childhood is especially important, because it provides useful information about the process of language acquisition, the acoustic and phonological changes of speech production. In childhood the physical structure of the body is continuously changing, thereby the length, shape and volume of the vocal tract are also changing which affects the acoustic structure of the vowels (Gósy 2004). Children have to learn to distinguish different sounds from each other and also to produce them. The acquisition of native language phonology is not completely adult-like until the teenager years. The perceptual ability of distinction precedes the production.

In Hungarian there are 14 vowels which are organised into seven phonological pairs (this research is focusing on only 3 pairs (based on frequency): [i]–[iː], [o]–[oː], [u]–[uː]). Vowel [i]–[iː] and [u]–[uː] are quantitatively different qualitatively almost identical (short ones are slightly less high). Vowel [o] and [oː] are qualitatively less similar (the long members are higher than the short ones). Studies carried out on spontaneous speech during the past decades (Hegedűs 1941, Kassai 1979, Menyhárt 2003, Olasz 2006) proved that the durations of phonemically long vowels are indeed longer than short vowels. However, the temporal interrelations of the long and short vowel pairs seem to be questioned presently. The duration of phonemic vowel pairs is believed to be more overlapped today in Hungarian than they were in the past (Siptár–Törkenczy 2000, Mády et al. 2008). Previous findings confirmed that the front high, unrounded short–long vowels show great durational overlap (Mády–Reichel 2007) based on laboratory conditioned experiments. On the basis of auditory impression Nádasdy and Siptár (2001) assumed that the difference in duration is decreasing between the long and short vowels. Gósy and Beke (2010) found that the duration of short and long vowels is overlapped and the duration of vowels is highly changeable based on the quality of the vowel, the phonetic position. However, they found significant differences in the objective durations between phonemically short and long vowels.

The present study aims to investigate the effect of phonological length on vowel production in the speech of Hungarian children between the ages of 7 and 15. In this study, we aimed to answer several questions: (i) what kind of differences can be observed in formant structure and duration of the phonologically short and long vowels between the different age groups and (ii) in which age can we find similar formant and duration values to adults.

Our hypotheses were as follows: (i) with the increase of the age, the duration of vowels is decreasing. (ii) The difference between the formant structure of phonologically short and long vowels is getting bigger as the children are getting older. (iii) At the age of 15 the formant and duration values of phonologically short and long vowels become similar to adults.

One hundred typically developing monolingual Hungarian-speaking children participated in this study. The analysis was cross-sectional and included five age groups: 7-, 9-, 11-, 13- and 15-year-old children. In every age groups there were 10 girls and 10 boys. None of them had any hearing disorder. Their intelligence fell within the normal range. The control group consisted of 20 adult speakers (10 males and 10 females). They were chosen from BEA (Gósy 2012; Neuberger et al. 2014) Hungarian spontaneous speech database. Speech material consisted of spontaneous speech samples. We analysed the duration and the first two formants (F1, F2) of [i]–[iː], [u]–[uː] and [o]–[oː] phonological vowel pairs.
The labelling was carried out using Praat 5.2 (Boersma–Weenink 2011). We created a reference database (the dataset was measured manually) to find the proper formant tracker. We measured the data formant tracker made by Geoffrey Morrison. We deleted the outliers (which were at least 2 standard deviations away from the mean. We normalized the data with the method made by Lobanov. The statistical analysis (Generalized linear mixed model = GLMM) was conducted by the means of SPSS 13.0.

At the age of 7 we found significant difference in formant structure between the phonological vowel pairs expect in the case of vowel [o]–[oː]. In the spontaneous speech of 9-year-olds the statistical analysis proved significant differences between the phonologically short and long vowels in the case of [o]–[oː]. The results demonstrated that in the spontaneous speech of 11-year-old children there were significant differences between the formant structure of phonological vowel pairs only in the cases of [o]–[oː]. At the age of 13 we found significant differences in the cases of phonological pairs between the short and long vowels (except [i]–[iː] and [u]–[uː]). By the age of 15 the formant structure of the vowels are getting similar to adults’. In cases of adults we found significant differences between [i]–[iː] and [u]–[uː] but not in [o]–[oː].

The duration of phonologically short members of the vowel pair in 7-year-olds is significantly shorter than the long ones difference both in case of [i]–[iː] and [o]–[oː] but not in [u]–[uː]. The investigation of duration of 9-year-olds vowels brought the following results: in the cases of the phonological pairs the short vowels are statistically shorter that the long ones. In the spontaneous speech of 11-year-old children the phonologically short vowels are shorter than the long ones. At the age of 13 the duration of the short vowels is significantly shorter than the long ones. By the age of 15 the duration of the vowels are getting similar to adults’. The duration of vowels shows shortening and in the cases of the formant structure tendentious changes can be also observed as the children are getting older. In adults we found significant difference between the short-long pair of investigated vowels.

The results of our research help to describe the acoustic-phonetic features of vowels of typically developing children, and can be used in diagnostic procedures for speech disorders.

References
The effect of variability on phonetic training for adults and children
Gwen Brekelmans¹, Elizabeth Wonnacott¹ & Bronwen G. Evans²
Department of Language & Cognition¹ and Department of Speech, Hearing and Phonetic Sciences²,
University College London, United Kingdom

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, 1991; Logan, Lively & Pisoni, 1993) using high variability phonetic training (HVPT), this method has become the standard paradigm. Key to the HVPT method is that learners hear multiple instances of phonemes in multiple phonetic contexts, as opposed to low variability (LV) training where they hear tokens spoken by one talker. Many previous studies have shown HVPT is an effective method for non-native speech sound training in adults (e.g. Nishi & Kewley-Port, 2007; Iverson & Evans, 2009), but few have used HVPT for non-native phoneme learning with children (e.g. Shinohara, 2014; Giannakopoulou, Uther & Ylinen, 2013). Moreover, no studies since Logan et al. (1993) have directly compared the effect of HV versus LV training.

The present study investigates how adults and children differ in their L2 speech learning using the HVPT paradigm, and whether they both show the expected HV over LV advantage. In two experiments, native English speaking adults and children were taught Dutch vowels in a single training session, during which they received either HV or LV input. In both experiments, computerised training was used in which participants identified the correct member of a minimal pair. In Experiment 1, geometrical shapes were used as a stand-in for orthographic representations of the vowels (given that these were naïve learners of Dutch with no knowledge of its orthography). In Experiment 2, however, pictures were used, making the task more akin to vocabulary learning.

Experiment 1: Participants were 48 8-year-old children and 48 adult native speakers of English. Participants were randomly assigned to 2 experimental conditions; HV and LV training. In the HV condition, stimuli were produced by 4 talkers, whilst in the LV condition, stimuli were produced by only 1 talker. For all tasks, stimuli consisted of 12 CVC three-way minimal pairs (all non-words) recorded by 6 different talkers, contrasting the three Dutch vowels: <au> /ɑu/, <eu> /ø:/, and <ui> /œy/. For each condition, talkers were counterbalanced across tasks. The 2 talkers used in the pre- and post-test were not used in training, and were rotated between the discrimination and identification task. The order of the 4 talkers in training was rotated across versions so that each talker occurred in every block across the different versions. Each participant completed one experimental session of about 30 minutes, which consisted of a pre-test, training, and post-test. The pre-test consisted of a discrimination task, after which participants were familiarised with the target vowels. Participants then completed four blocks of training with trial-by-trial feedback in which they mapped the vowels in the CVC contexts to fixed geometrical shapes presented on screen. The post-test comprised a discrimination task (identical to pre-test), followed by a vowel identification task akin to the training task but using both trained and novel items (and with no feedback).

Results: Adults outperformed children in all tasks. In training, while both groups were above chance, only adults showed reliable improvement through the session (beta = 0.19, SE = 0.03, z = 5.81, p < .001); while children remained stable across blocks performing just above chance (beta = 0.01, SE = 0.03, z = 0.32, p = .747; see Figure 1). Neither group showed a main effect of condition, indicating that the variability manipulation did not affect learning (adults: beta = 0.05, SE = 0.14, z = 0.37, p = 0.715; children: beta = -0.02, SE = 0.08, z = -2.24, p = 0.080). In the discrimination task, although both groups were above chance at pre-test, there was no improvement at post-test in either training condition (adults: beta = -0.12, SE = 0.16, z = -0.80, p = 0.422; children: beta = -0.18, SE = 0.11, z = -
1.61, p = 0.107). In the identification task taken at post-test, adults were well above chance and were more accurate on items they were trained on (beta = 0.19, SE = 0.09, z = 2.09, p = .037), while children were only barely above chance with no effect of item novelty (beta = 0.10, SE = 0.09, z = 1.15, p = .248). Neither group showed an effect of variability (adults: beta = 0.36, SE = 0.20, z = 1.83, p = 0.067; children: beta = -0.16, SE = 0.09, z = -1.69, p = 0.092).

**Figure 1:** Accuracy scores in training for Experiment 1 plotted over the 4 blocks for adults (left) and children (right), comparing HV and LV. Error bars show 95% CI, the dashed line indicates chance.

In sum, though both groups showed some learning of the relationships between the symbols and the Dutch vowels, there was no benefit of high over low variability input and neither training condition led to improvement in discrimination. However learning was very weak in child learners. Experiment 2 uses a less abstract learning task to explore whether this would boost children’s learning.

**Experiment 2:** The methodology used in Experiment 1 was adapted so training was more similar to word learning. Participants were 48 children and 48 adult native speakers of English. Stimuli consisted of 12 monosyllabic two-way minimal pairs in CVC contexts containing the same Dutch vowels as before. All stimuli were real words, recorded by the same talkers as before. The 30-minute experimental session was again a pre/post-test design, and was nearly identical to Experiment 1. The pre-test consisted of a discrimination task (same as Experiment 1), followed by four blocks of training during which participants were asked to map the stimuli to clipart pictures depicting the word, for which they received trial-by-trial feedback. The post-test comprised the discrimination task (as in the pre-test), followed by a picture identification task akin to training but using both trained and non-trained picture pair combinations. The critical manipulation was once again talker variability in training, with participants randomly assigned to either a HV or LV condition. Talkers were counterbalanced as in Experiment 1.

**Results:** Performance in training was overall higher than in Experiment 1 (see Figure 2). Although adults again outperformed children, both adults and children were well above chance and both showed improvement through the session (adults: beta = 0.20, SE = 0.03, z = 5.79, p < .001; children: beta = 0.08, SE = 0.03, z = 3.01, p = .003). Both groups showed stronger learning in the LV condition where they were trained on a single talker, though this HV/LV difference was only reliable in children (beta = 0.22, SE = 0.08, z = 2.79, p = 0.005). For the identification task, adults again outperformed children though both groups were above chance. Critically however, there was no effect of training type, with participants performing equally after both HV and LV training (adults: beta = -0.08, SE = 0.26, z = -0.30, p = 0.761; children: beta = 0.14, SE = 0.17, z = 0.81, p = 0.416). This was true both for stimuli with the trained and untrained talker, suggesting that there was no benefit of variability even for untrained items. As in Experiment 1, participants were above chance in the discrimination task even at pre-test, however there was no increase in performance following training (adults: beta = 0.23, SE = 0.16, z = 1.41, p = 0.159; children: beta = 0.08, SE = 0.10, z = 0.75, p = 0.454).
In sum across the two experiments, we saw no benefit of HV over LV phonetic training for either children or adults, including no benefit for untrained items. In addition, although participants learned associations between shapes/pictures and the novel vowels, we also did not see any effects of either type of training on their discrimination of those vowels. An additional finding was that learning was much greater in a word learning context than in a more abstract task, in particular for the children who learnt very little in the latter. Moreover, adults outperformed children on all tasks, suggesting that in this study, children did not benefit from increased plasticity (cf. Shinohara, 2014). However, the lack of both any effect of variability in training and of any benefit for generalisation might be due to the brevity of the training session: in general, studies that have used the HVPT paradigm have been multiple-day training studies. This could also explain the lack of change in discrimination abilities. Future plans involve expanding the second experiment to investigate effects of input variability, i.e., single vs. multiple phonetic contexts, single vs. multiple talker, on learning in a two-week training study with Dutch learners of English. Results will have implications for L2 teaching in particular, with regard to the type of exposure from which learners might benefit most.

References

**Functions of silent pauses in spontaneous speech**
Dorottya Gyarmathy
Research Institute for Linguistics, Hungarian Academic of Sciences, Hungary
Our speech is occasionally interrupted by pauses of various length. They are an essential part of human speech. Until now it is not clarified, what can be actually considered as pause, how differently pauses can appear in speech, what are their acceptable minimum and maximum durations, and what functions they can carry on in speech. Until the first half of the 20th century researchers examined the pauses mainly from rhetorical aspects, they analysed the relationship between punctuation of written texts and their spoken realisations (Mátray 1861; Hevesi 1908; Simonyi 1903; Lindroth 1933). As early as in 1877 Sweet was the first scientist who mentioned pauses as parts of the language system. He connected pauses of speech with breathing, and called “breath group” the section of speech uttered with one act of breathing out. Following Sweet’s idea, various researchers discussed the problem of pauses of speech, such as Hungarian Balassa (1886), German Viëtor (1894), and Danish Jespersen (1904). Previous studies connected pauses of speech with either breath-taking, or punctuation (Weiske 1838; Bieling 1880).

From the second half of the 20th century empirical researches showed that silent pause is the most common phenomenon in the spontaneous speech, and carries many different functions in spontaneous speech (Boomer 1965; Goldman-Eisler 1958; Hargreaves–Starkweather 1959; Levin et al. 1967; Tannenbaum et al. 1967; Verzeano–Finesinger 1949; Misono–Kiritani 1990; Gósy 2000, 2003; Menyhárt 2003; Markó 2005; Bóna 2007; 2013; Neuberger 2014). Speakers need pauses to breath, plan what they are going to say or negotiate turn-taking. Esposito et al. (2007) conclude that pauses in speech are typically multi-determined phenomena; they have socio-psychological, communicative, linguistic and cognitive reasons. It is known from previous research that silent pauses caused by speech planning difficulties differ from those occurring at syntactic boundaries (Boomer 1965; Lounsbury 1965; Szende 1976). Silent pause can be defined as disfluency pause if it solves speech planning difficulties, gives opportunity to self-repair or supports lexical access (Gósy 2007).

This study focus on the analysis of the temporal structure of silent pauses in Hungarian spontaneous speech. We hypothesized, that 1. silent pauses differ from each other according to their functions, 2. various types of silent pauses in terms of frequency and duration shows different patterns, 3. duration of silent pauses is defined also by the syntactical position.

Our research is based on narratives of the BEA Hungarian spontaneous speech database (Gósy 2012) and and of the HunEng-Database. We used the recordings of 10 adult subjects (5 males and 5 females, their mean age was 27.4 years) and 10 teenagers (5 males and 5 females, their mean age was 16.3 years), and analysed all occurrences of their silent pauses. The spontaneous speech material we analysed was 71 minutes long. Each occurrence of silent pauses was annotated by Praat, version 5.4.21 (Boersma & Weenink 2013). We defined their duration manually. For statistical analysis, General Linear Mixed Model (GLMM) was used (SPSS 20.0).

We differentiated silent pauses (S) occurring either in phrases or at phrase boundaries and pauses functioning as editing phases (E). The former pauses were called as syntactical silent pauses while the latter ones were called as editing phase silent pauses. Silent pauses were categorized into the following categories: 1. phrase boundary pauses, 2. within phrase pauses, 3. end of phrase pauses, 4. utterance onset pauses. The subcategories of the editing phases depend on the type of disfluency surrounding the silent pause.

We analysed the temporal properties of occurrences of silent pauses in the two main categories (syntactical silent pauses vs. editing phase silent pauses). 87.84% of all pauses were syntactical silent pauses while 12.16% of them were editing phase silent pauses. Comparing the average durations of instances in the two categories, we found statistically significant difference between them. The
mean duration of the editing phase silent pauses was 494 ms, whereas that of the syntactical silent pauses was 559 ms, on average (Figure 1).

**Figure 1: Durations of instances of syntactical silent pauses and editing phase silent pauses (means and standard deviations)**

The subcategories also differ from each other in terms of duration and occurrence. The most common types (54%) were “phrase boundary” pauses, their mean duration was 587 ms, whereas the durations of the less frequently used type, the “utterance onset pauses” amounted to 899 ms, on average in each age group. In terms of statistical analysis, the difference is significant: $F(5, 47) = 12.376; p < 0.001$ (Figure 2).

**Figure 2: Durations of the subcategories – results of adults (means and standard deviations)**

Our results confirm all the three hypotheses, the temporal data supported that the function of silent pauses had a main effect on both their occurrences and their durations, as well.

**References**
Simony, Zs. (1903). Iskolai helyesírás. [School Orthography]. Budapest.
Development of speech processing in infancy and its relationship with later language skills
Kathleen McCarthy¹,², Katharine Mair¹, Katrin Skoruppa³ & Paul Iverson¹
¹University College London, ²Queen Mary University of London, ³University of Neuchâtel

In our recent work, we created detailed perceptual maps of infant vowel development using the Acoustic Change Complex (ACC) within EEG. More specifically, we showed that vowel perception development seems to be driven by the low frequency aspects of the speech signal. Such findings could be explained by early auditory experience of low frequencies in the womb (Kisilevsky et al., 2009), and possibly explain the developmental trajectory of vowels being acquired earlier than higher frequency consonant sounds such as fricatives (e.g., Nespor, Pena & Mehler, 2003). The aim of the present study was to explore this hypothesis further by detailing individual differences in vowel and fricative perception throughout the first year of life. A secondary aim was to investigate the relationship between this early auditory processing and later language outcomes.

Eighty monolingual English infants (4-5, 7-8 and 10-11 months old) took part in the study. Vowel and fricative sensitivity was assessed using the ACC within EEG, which for infants typically evokes a positivity about 150-200 ms after a spectral change. The ACC was measured for three vowel pairs and three fricative pairs, that were presented in runs of five per pair before switching to a new random pair. ERPs were averaged across epochs for each pair, with the magnitude of the response for each pair being used as a comparison measure. Each infant was tested for an average of 18 minutes, resulting in around 250 trials per pair. In addition, the infants’ language skills were assessed at 16, 20, and 24-months using the Oxford CDI.

Preliminary results show individual differences for vowel and fricative perception. Infants initially show a greater sensitivity to the vowel contrasts, with fricative sensitivity increasing with age. Further analysis will be conducted to explore which spectral features are driving the infants’ perceptual sensitivity. These findings display a more detailed pattern of the speech perception than has been shown before, and demonstrate how the ACC can be used to explore developmental trajectories. The results will be discussed with respect with the infants’ later language outcomes, and the possible application of this method in a clinical setting.

References
Individual Differences in L2 bimodal input processing
Joan C. Mora
Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, Spain
mora@ub.edu

The low perceptual salience of function words in spoken English pose a processing challenge for English learners of syllable-timed languages like Spanish that lack syllable reduction and are likely to transfer their L1 syllabic rhythm when speaking English, producing unreduced vowels in unstressed syllables (Flege & Bohn, 1989). Bimodal input processing of text and sound, as in reading-while-listening or watching captioned video, facilitates the integration of the orthographic and auditory forms of words enhancing speech processing and comprehension (Mitterer & McQueen, 2009). Research on multimodal input in language learning has shown the benefits of watching captioned video for L2 listening comprehension (Vanderplank, 1988), L2 vocabulary acquisition (Montero Pérez et al., 2013), speech perception (Mitterer & McQueen, 2009) and speech segmentation (Charles & Trenkic, 2015), but little is known about its effects on the processing of function words (Krejtz et al., 2015) or how individual differences in speech-related cognitive skills affect speech processing in highly attention-demanding speech processing contexts such as watching L2 captioned video. Individual differences in the use of executive control in speech processing (working memory, attention and inhibition) has been shown to explain a substantial amount of variance in L2 phonological attainment (Darcy et al., 2016; Golestani & Zatorre, 2009; Lev-Ari & Peperkamp, 2013, Link et al., 2009), but their role in the processing of L2 speech in captioned video has not been investigated yet with a focus on cross-language differences in speech rhythm (syllable- vs. stress-timed).

The current study examined the role of individual differences in inhibitory control and attention in 43 L1-Spanish EFL learners’ processing of L2-English speech during bimodal input exposure through captioned video. The EFL learners (and native British English controls, n=19) watched 6 short film clips in English with intra-lingual subtitles (captions) and answered true/false comprehension questions on their content while their eye gaze behaviour was recorded on a Tobii T120 eye-tracker. For each one of the video clips, participants were randomly assigned to one of three viewing conditions: (A) standard: on-screen image + audio + caption text; (B) reading-while-listening: audio + caption text without background image; and (C) reading-only: caption text only; so that each participant watched two of the six video clips in each one of the 3 conditions (A, B, C) and each video clip was presented in each one of the 3 conditions to approximately one third of the participants (14-15 L2E and 6-7 NE). Areas of Interest (AOIs) were drawn for every function and content word in the captions and to obtain standard eye-tracking metrics (fixation counts and durations) while controlling for word length. The role of individual differences in visual inhibitory control was assessed through a flanker task (Costa et. al., 2009), whereas attention control was measured through a novel attention-switching task using an alternating runs paradigm (Monsell, 2003). In addition, we assessed inter-learner differences in segmentation skills through a word-spotting task (McQueen, 1996) and in L2 proficiency through an elicited imitation task (EIT, Ortega et al., 2002). A selection of five of the thirty sentences produced in the EIM were analysed for speech rhythm, using calculated from vowel (V) and consonant (C) durations: varcoV (Dellwo, 2006), Vnpvi (Grabe & Low, 2002) and Vcci (Bertinetto & Bertini, 2008).

The eye gaze data analyses revealed that function words (F) were fixated less often than lexical words (L) in all three viewing conditions by all participants. In condition A captions (n=69) received
more fixations from L1-Spanish learners than L1-English controls, whereas speakers’ mouths were fixated similarly by both groups. In all conditions both function \((n=872)\) and lexical \((n=676)\) words were fixated more often by learners than by controls. Because there was a main effect of word length on the mean number of fixations by word in all conditions (Figure 1) further analyses and individual eye-gaze measures were based on 1-syllable function \((n=255)\) and lexical \((n=120)\) words only.

**Figure 1:** Mean number of fixations per word by participant group.

![Graph showing mean number of fixations per word by participant group.](image)

Eye gaze behaviour differed as a function of L1 group, viewing condition and word type. In conditions A (standard) and C (read-only) L1-Spanish learners total duration fixation (dwell time) was similar for 1-syllable function and lexical words (Figure 2), whereas in condition B (read-while-listen) fixation duration was much longer on lexical than function words (unlike L1-English controls) suggesting that in the absence of background images the L2 audio led participants to focus more on lexical than function words.

**Figure 2:** Dwell time (total fixation duration) by condition, word type and group for L1-Spanish learners.

![Graph showing dwell time by condition, word type and group.](image)

We then examined the relationship between L1-Spanish learners’ proficiency, segmentation, inhibition, attention and speech rhythm measures and their eye gaze data for 1-syllable function words. Higher proficiency learners had better segmentation skills \((r=.482, p=.003)\) and fixated less often \((r=.401, p=.014)\) and with shorter first-fixation durations \((r=.482, p=.003)\) on function words in condition A (standard) than lower proficiency learners. Learners with stronger inhibitory control obtained longer first-fixation durations \((r=-.375, p=.022)\) in condition B (read-while-listen) and longer
dwell times ($r=-.436, p=.007$) in condition C (read-only). Slower mean response latency on switch trials in the attention switching task were related to poorer inhibitory control ($r=.376, p=.022$), but no association was found between attention control switching costs and any of the eye gaze measures. Learners’ speech rhythm was found to be stress-timed (i.e. L2-like) rather than syllable-timed (L1-like) on the varcoV, Vnpvi and Vcci measures (with indices $>55$), but the more stress-themed their speech rhythm was the longer they fixated on lexical words in condition A (Vnpvi: $r=-.581, p<.001$; Vcci: $r=.314, p=.044$). Overall these results reveal an important effect of L1-Spanish learners’ proficiency on the processing of bimodal input and suggest a limited role for segmentation and cognitive skills. However, clear processing differences emerged as a function of viewing condition and word type.

References
The processing of lexical and contextual Information in cochlear implant users

Leanne Nagels\textsuperscript{1,2*}, Roelien Bastiaanse\textsuperscript{2}, Deniz Başkent\textsuperscript{1}, and Anita Wagner\textsuperscript{1}

\textsuperscript{1}Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, the Netherlands
\textsuperscript{2}Center for Language and Cognition Groningen (CLCG), University of Groningen, the Netherlands, leanne.nagels@rug.nl

Introduction
Although speech perception is an elaborate process that demands the rapid identification of acoustic features, it is often effortless for normal-hearing listeners in optimal conditions. When noisy environments degrade the acoustic information in the speech signal, listeners rely more on additional linguistic cues, such as contextual or lexical information (Mattys & Wiget, 2011). It is unclear how listeners with a cochlear implant (CI), who process a spectrotemporally degraded speech signal, integrate acoustic and contextual information, and whether this process is inherently effortful for them. In the present study we investigate the time-course of speech recognition and integration of contextual and lexical information in CI users by means of eye-tracking, and the processing effort involved therein by means of pupillometry.

Lexical competition is the process in which the listener considers the words in the mental lexicon that overlap with the speech signal. An unclear or noisy speech signal will increase listeners’ uncertainty about their interpretation of the signal, and also the number of lexical candidates that are activated. Previous studies have shown that the timely integration of contextual and lexical information reduces the time course of word recognition in silence (Dahan & Tanenhaus, 2004), and also in degraded speech (Wagner, Pals, De Blecourt, Sarampalis, & Başkent, 2016). In addition, there are several lexical factors that influence the time course of word recognition, such as lexicality, neighbourhood density (Vitevitch, Luce, Pisoni, & Auer, 1999), age of acquisition (AoA) (Turner, Valentine, & Ellis, 1998), and word frequency (Dahan, Magnuson, & Tanenhaus, 2001).

A cochlear implant (CI) is a prosthetic device that enables deaf individuals to partially regain the ability to perceive speech. However, the improvement in speech perception is limited by the degraded speech signal that is received via the device. Wagner et al. (2016) studied the effect of verb-based thematic constraints in normal-hearing (NH) listeners in CI simulations using an eye tracking paradigm. The results of Wagner et al. (2016) showed that verb-based thematic constraints reduced the time course of lexical competition for natural and CI simulated speech. However, the integration of semantic information was delayed and lexical competition was prolonged in the CI simulated speech. In addition, simultaneous recordings of listeners’ pupil dilation showed that processing degraded speech increases processing effort, while integrating contextual information decreases processing effort for natural speech. The present study investigates how CI users, who through experience adapted to the degraded speech signal, process contextual and lexical information, and whether this reduces their processing effort.

Methods
Two eye-tracking experiments were conducted with twelve postlingually deaf CI users (mean age: 58.1 years; age range: 31-71 years) and eleven NH listeners (mean age: 59.5 years; age range: 31-71 years), matched on age and gender, using an EyeLink 500 head-mounted eye tracker.

In experiment 1, the processing of contextual information was measured by means of verb-based thematic constraints (Dahan & Tanenhaus, 2004; Wagner et al., 2016). A visual world paradigm with drawings of a target (e.g. lamp), a phonological competitor (lamb), a semantic competitor (mirror) and an unrelated distractor (dragon) were used. The drawing sets were matched to a verb that provided thematic constraints (e.g. to break), which were coherent with the target noun and the
semantic competitor. Recordings were made of sentences in which the main verb either preceded the target noun, and thus provided thematic constraints, or followed the target noun without providing thematic constraints. Sentences were presented auditorily concurrently with the display of the four drawings on a computer screen. Participants were instructed to click on the corresponding drawing of the noun that was mentioned in the sentence. Listeners’ gaze fixations to the drawings and their pupil dilation were measured while listening to the sentences.

The aim of experiment 2 was to investigate the processing of four types of lexical information, namely, lexicality (word-nonword distinction), neighbourhood density, AoA, and frequency. An auditory lexical decision task was conducted with recordings of fifty words and fifty nonwords that were balanced for syllable length, frequency, and neighbourhood density using the Dutch CLEARPOND database (Marian, Bartoletti, Chabal, & Shook, 2012). AoA ratings for words were retrieved from Brysbaert, Stevens, De Deyne, Voorspoels, and Storms (2014). Participants were instructed to press a yellow key if they heard a non-existing word and a red key if they heard an existing word.

Results
In experiment 1, gaze fixations of correct response trials within the time window of 0 to 2000 milliseconds after the target onset were analysed using growth curve analysis (Mirman, 2014) in R (R Core team, 2013). The time curves of gaze fixations towards the competitors were modelled as fourth-order orthogonal polynomials with random effects for participants. The variables for group, condition, and the three-way interaction between group, condition, and the terms describing the polynomial were added to the model as fixed effects. Model comparison was done using the Likelihood Ratio test to determine the contribution of individual fixed effects to the fit of the model. A significant interaction between group and condition on all terms of the polynomial was found for gaze fixations towards the phonological competitor \( \chi^2(5) = 294.18, p<0.001 \), and semantic competitor \( \chi^2(5) = 529.433 p<0.001 \).

Figure 1: Proportion of gaze fixations

This shows that CI users and NH listeners were both able to restrict the number of considered lexical candidates based on the contextual information that was provided. Lexical competition primarily took place between the target and semantic competitor in the constraining condition and between the target and phonological competitor in the neutral condition (Figure 1). However, the time course of lexical competition was prolonged in CI users relative to NH listeners. In addition, more individual variation was found among CI users than NH listeners. Individual-level analysis of the pupil dilation data reveals that CI users who had a larger overall increase in pupil dilation, also showed a slower resolution of lexical competition. Figure 2 displays the percentage of increase in pupil dilation normalized to the baseline of 200 ms preceding word onset per participant and trial, averaged across participants. Also, CI users had a smaller overall increase in pupil dilation than NH listeners who were presented with CI simulations (Wagner et al., 2016), which indicates that they adapt to the degraded speech through experience with their hearing device.

Figure 2: Pupil dilation normalized per participant and trial

In experiment 2, two generalized linear mixed-effects models were used in R (R Core team, 2013) to analyse the accuracy and reaction time data. Accuracy scores were modelled using a binomial distribution and reaction times were modelled using a Gaussian distribution with a log link function.
The variables for group (NH vs CI), lexicality, frequency, neighbourhood density, and an interaction between group and lexicality were included as fixed effects. Model comparison based on the Likelihood Ratio test show that the addition of fixed effects for group, ($\chi^2(1) = 31.15, p < 0.001$), lexicality, ($\chi^2(1) = 18.16, p < 0.001$), neighbourhood density, ($\chi^2(1) = 2.39, p < 0.001$), and the interaction between group and lexicality, ($\chi^2(1) = 33.89, p < 0.001$), significantly improved the model fit for accuracy scores. The addition of fixed effects of group, ($\chi^2(1) = 11.59, p < 0.001$), lexicality, ($\chi^2(1) = 88.95, p < 0.001$), and an interaction between group and lexicality, ($\chi^2(2) = 21.03, p < 0.001$) significantly improved the model fit for reaction times. A tendency towards significance was found for AoA on accuracy ($\chi^2(1) = 3.25, p = 0.071$), and reaction time ($\chi^2(1) = 3.67, p = 0.055$). We found that CI users had lower accuracy scores and longer reaction times than NH listeners. Furthermore, significant effects of lexicality and neighbourhood density were found. More errors were made with nonwords and high neighbourhood density words and nonwords, especially by CI users who demonstrated a lexical bias. CI users may rely more on lexical information and use more top-down restoration mechanisms due to the less reliable information in the degraded speech signal.

Conclusion
Overall the results demonstrate that CI users can timely integrate contextual and lexical information and benefit from it, but the lexical competition patterns that they display are different from NH listeners. The integration of semantic information is delayed in CI users relative to NH listeners due to the degraded speech signal, which prolongs the time course of lexical competition. Finally, pupillometry data shows that there is a relation between processing effort in CI users and their ability to resolve lexical competition.

References
Perception boundaries between single and geminate stops in Hungarian children and adults
Tilda Neuberger
Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary, neuberger.tilda@nytud.mta.hu

Introduction
There has been much research indicating that a specific speech sound is characterized by wide variability in timing of articulation (e.g., Rosen, 1992). Thus, the realisation of singleton and geminate consonants shows considerable overlapped durations in many languages (Pickett et al., 1999; Hirata & Whiton, 2005; Neuberger, 2015). However, despite the wide variability, it is also confirmed by production studies that the primary acoustic attribute that distinguishes geminates from singletons is duration (Ham, 2001; Ridouane, 2010). Human listeners can identify phonemes – and distinguish them by quantity – successfully, irrespective of speaker or speech condition. In speech perception research, it is an essential question how listeners discriminate phonological categories of quantity (singleton and geminate) along a continuous durational scale of the phonetic realisations. This perceptual task seems to be more difficult for children than for adults. Typically developing Hungarian-speaking children under the age of 7 years were found to have major problems in consonant length discrimination tasks (Gósy, 2006; Jordanidisz, 2015). It was also found that at the age of 7 and 8 years, children become able to precisely identify consonants across quantity categories.

Many studies investigated the perceptual boundary between the two phonemic categories (e.g., Lisker, 1958; Hankamer et al., 1989; Pickett et al., 1999; Amano & Hirata, 2010). Listener responses to incrementally manipulated durations allow one to determine the point at which a singleton percept shifts to a geminate percept.

The aim of this paper is to examine perception boundaries between two quantity categories (singleton and geminate) as well as to reveal the role of closure duration (CD) as a perceptual distinction between Hungarian single and geminate stops. The main question was what kind of similarities and differences can be found between children’ and adults’ perceptual shifts from singleton to geminate. It was hypothesized that (i) the perceptual boundary between singletons and geminates would be situated at different CD values in adults than in children (children’s boundary values are presumably higher than those of adults), (ii) the original quantity of the manipulated stops would affect all listeners’ responses regardless of age.

Method
A binary discrimination test of stops with systematically manipulated closure duration was conducted to test the above-mentioned hypotheses. The stimuli contained VCV and VC:V nonwords with Hungarian voiceless stops and their geminate counterparts (the consonants were bilabial, alveolar and velar stops; the vowel was identical [i] in all cases). Closure duration of original singletons was artificially lengthened in 10 ms steps, while CD of original geminates was shortened in 10 ms steps. Incremental manipulation of CD as well as perception test was carried out using Praat software (Boersma & Weenink, 2015). Participants were selected from three age groups: a) 7-year-olds children (N = 16), b) 8-year-old children (N = 16), c) young adults (age: 18–27 years; N = 44). Their task was to listen to the samples and make a binary decision about whether the heard consonant was long or short, for example [iti] or [it:i]. All children (N = 32) and adults (N = 44) did the same two-alternative forced-choice task.

For evaluating the binary data, listeners’ responses to each item were summarized. Percentage of geminate responses at each duration was measured in the case of all stop consonants. Perception boundaries were measures at the 50% rate between the identification of a singleton versus a geminate. In addition, the two sets, i.e. originally singleton and originally geminate stimuli, were analysed separately. The reason for having two sets of stimuli, i.e., shortened geminates and
lengthened singletons, was to observe any possible differences in listeners’ short or long responses for original singletons and original geminates with equal closure duration. Fitting a logistic function to the sets and plotting response curves was carried out using MATLAB 2015b. Statistical analysis (binary logistic regression, GLMM) was conducted by means of SPSS 20.0 software.

Results

Binary logistic regression showed that closure duration had a main effect on listeners’ short-long responses both in adults \( [F(1, 6067) = 1317.391; p < 0.001] \) and in children \( [F (1, 1926) = 248,253; p < 0.001] \).

Results showed that perceptual shifts differed between children and adults, but it depended on the place of articulation of stops. Children’s and adults’ responses were similar in case of alveolar and velar stops, but showed differences in the case of bilabial stop. We also found age-specific differences between 7-year-olds and 8-year-olds; ‘age’ had a main effect on short-long responses, which was confirmed by statistical analysis \( [F (1, 1926) = 11,102; p = 0.001] \).

Response curves (which represent the percentage of ‘long’ responses at different closure durations) formed typical S-shaped curves in adults, but a linear function was observed in children (Figure 1). This result suggests that adults’ responses were more categorical than those of children’s. Response curves and of stimuli created from original singletons and original geminates were not identical in the case of the velar consonant. In children and adult, for velar stops, the perceptual boundary of originally geminate consonants appeared earlier than in the case of responses to stimuli created from original singletons. The displacement of 50% boundary closure durations between \([k]\) and \([k:]\) was 20 ms in children and 15 ms in adults (Figure 2).

Conclusion

Examination of native listeners’ quantity discrimination was conducted with the intention of obtaining detailed information about the relationship between acoustic and perceptual cues of consonant length in Hungarian.

Both hypotheses were confirmed by the data. We found age-specific differences in the perceptual boundaries between singletons and geminates. The original quantity of the manipulated velar stops affected both children’s and adults’ responses. This suggests that acoustic cues other than closure duration may play a role in the identification of quantity contrast.

Results supposed to shed light on age-specific changes of the processing of speech perception. Findings may contribute to speech therapy or educational practice (development of articulation or reading, writing, spelling skills), and may be useful for second language learning or teaching.

Figure 1: Response curves for voiceless velar stops with respect to listeners’ age
Figure 2: Adults’ response curves for voiceless velar stops with respect to the original length of stimuli

References
Acquisition of L1-French corner vowels and formant visualisation in the VisuVo software: a longitudinal case study

Nikola Paillereau & Naomi Yamaguchi
Laboratoire de Phonétique et Phonologie (CNRS & Université Sorbonne Nouvelle Paris 3)
nikola.paillereau@mac.com, naomi.yamaguchi@univ-paris3.fr

In the course of the first language (L1) development, productions of speech sounds are more variable in children than in adult speakers. This variability can be explained by several factors: from birth to adulthood, children's vocal tracts experience significant changes (Vorperian et al. 2005); their speech motor control is developing, and they shift from babbling, which contains sounds that may be unrelated to their language, to the target phonological sound system.

Anatomical changes are reflected in acoustic measurements: as the vocal tract lengthens, the vowel formants decrease. According to Vorperian and Kent's (2007) synthesis of numerous acoustic data on English vowels, changes in F1 and F2 values are particularly noticeable in corner vowels of children aged between 1 and 4.

In French, corner vowels [i, a, u] are good examples of Daniel Jones' cardinal vowels (Vaissière, 2011). French isolated [i, a, u] are defined by converting formants that are amplified: F3 and F4 in [i] (F3 is maximal, contrary to English [i] which aims at maximal F2), F1 and F2 in [u] (with the lowest possible values of both formants) and [a] (with high F1 and low F2). Formants of French vowels have been recently described in productions of adult speakers (Gendrot and Adda-Decker 2005, Paillereau 2016), but they need to be investigated in infants, toddlers and young children in a longitudinal perspective (Rvachew et al., 2006).

The aim of this study is to show how formant values of [i, a, u] change according to the chronological age. Of particular interest are coarticulatory effects and variability due to vowels' prosodic positions. Results are demonstrated in a piece of software called VisuVo (Paillereau 2015, 2016), whose innovation resides in the possibility to generate, from large databases, original figures in real-time and to dynamically compare data from different recording sessions or of different speakers.

Methodology: The speech material consists of French vowels [i, a, u], spontaneously produced in different phonetic environments, i.e. in isolation (0), and in symmetrical labial (pVp), coronal (tVt) and palato-velar (kVk) contexts, as well as different positions within an accentual phrase, i.e. initial (I), median (M) and final (F). Speaker A is a French male toddler, raised up in a French monolingual family in the Parisian region and without any reported hearing problems. Recordings were made from the age of 1;03 to 4;11 once a month, in naturally occurring interactions. In order to illustrate the evolution of vowel formants as well as the variability due to coarticulation and prosodic position, recording sessions at 3 different chronological ages - 2;01 (session 1), 2;08 (session 2) and 3;08 (session 3) – are presented in this study. The acoustic data are systematically compared to vowels uttered by 10 native French women (M age = 28.7), taken from Paillereau, 2016.

The original corpus was phonetically transcribed by an experienced native French listener and a set of 546 vowels was analysed in Praat (Boersma and Weenink 2017). Formants were calculated with the formula To Formant (burg) and an analysis window at 25 ms. The first five formants were searched within 8000 Hz, but in the case of aberrant values, the frequency range was modified. Formants’ values were taken semi-automatically; the raw data were treated in VisuVo.

Results: Results on the acoustic development of [i, a, u] are as follows:

1. Acoustic changes due to chronological age are not the same for all 3 vowels. When comparing session 3 to session 1, F2 of [u] decreases by 221 Hz, F2 of [i] increases by 496 Hz, and F2 of [a] remains about the same (+ 13 Hz). The F1 value drops by 170 Hz in [i] and by 79 Hz in [a], but it remains about the same for [u] (- 6 Hz). The vowel triangle F1/F2 of Figure 1 (on the left), where productions of French adult women are in purple, and productions of A from session 1 in red and from session 2 in blue, makes it clear that the phonetic realization of an open [a] is much more
adult-like than that of close vowels [i] and [u]. In session 3, F1 and F2 of [a] are at 833 Hz and 1823 Hz respectively; those of adults are at 744 Hz and 1703 Hz. On the contrary, [u] is not adult-like in any session; it is realized with similar F1 and F2 as adults’ [œ] (in session 1) and as adults’ [ɔ] (in session 3), yet with a much higher F3.

**Figure 1:** On the left: F1/F2 of vowels produced by A in session 1 (in red), session 2 (in blue) and by French female adults. On the right: F1/F2 of vowels produced in labial and dental contexts by A in session 1 (in green and blue respectively) and by French adults (in red and purple respectively)

2. Acoustic variability due to coarticulatory effects is more pronounced in speaker A than in adults. Figure 1 (on the right) compares F1/F2 triangles of vowels produced in labial and coronal contexts by speaker A in session 1 to those produced in the same contexts by adult speakers. Whereas A produces [i, a, u] with a much higher F2 in a dental context compared to a labial context (+590 Hz, +542 Hz and +660 Hz respectively); only F2 of [a] and [u] is increased in adults (by 274 Hz and 449 Hz respectively).

3. Variation of vowel formants due to prosodic position is much more pronounced in speaker A than in adults. Figure 2, which is one of the 3 graphs automatically generated by VisuVo, indicates the evolution of formants of the vowel [u] pronounced by A in session 2 (in blue) and by French adults (in red). Formants are traced according to consonantal contexts: pVp in the 1st column, tVt in the 2nd column, kVv in the 3rd column, and RVR (uvular, only for adults) in the 4th column. Each column gives 3 sets of formants according to the prosodic position: I = 1st set, M = 2nd set, F = 3rd set. Horizontal lines (red and blue) that cross the figure correspond to the F-pattern of an isolated [u]. Prosodic positions in which A did not produce [u] are left vacant.

**Figure 2:** Average formants F1, F2 and F3 (in Hz) of [u] uttered in pVp, tVt, kVv and RVR contexts by A in session 2 (in blue) and by French adults (in red). The standard deviation is 1.5

Figure 2 shows that while the prosodic position does not affect adults’ realization of [u], it does affect that of speaker A. In tVt context (2nd column), the central value of F2 of [u] uttered in initial and median positions is at 1770 Hz and 1895 Hz respectively, whereas that of final [u] is at 1326 Hz. Isolated [u] has not a focal character as in adults; the F2-F1 distance in A is at 6.1 Bark, whereas it is
at 4.1 Bark in adults. Speaker A’s realizations are thus the most adult-like in accentual phrase final positions, which are accented.

**Conclusion:** Spectral changes due to chronological age do not concern all vowels equally. Whereas [a] is produced adult-like from the age of 2;01 years, [i] undergoes, between the ages of 2;01 and 3;08 years, important changes in F2, which increases. To some extent, F1 decreases. On the contrary, the acoustic realization of [u] undergoes less important spectral changes. At the age of 3;08 years, A’s [u] is still far from adult speakers’ one, even though the accentual phrase final position enhances its acoustic accuracy. The development of the two high vowels is thus non-linear, namely in their F1 value. These findings do not seem to confirm Vorperiant and Kent’s (2007) statement about a quasi-complete acoustic mastery of all non-rhotic vowels at the age of 36 months, which may be explained by the fact that the authors focused on the acquisition of English vowels that are less focal than French vowels. Coarticulatory effects are stronger in A than in adult speakers. This is particularly illustrated by the acoustic realization of [i], which, in adults, is completely resistant to coarticulation, but which, in A, undergoes important spectral changes, namely at the age of 2;01. The finding that coarticulatory patterns in children do not match those of adults’ was also noticed by Lee et al. (1997), who explain large F2 transitions between a vowel and its adjacent consonants by excessive tongue movements. Finally, whereas variability due to prosodic position within an accentual phrase is not characteristic of adult productions of vowels, speaker A produces final vowels in a more adult-like manner than initial and median vowels. This indicates that acoustic realizations under accentual phrase stress are more adult-like.

These results need to be confirmed with a larger corpus, in terms of participants and longitudinal steps. They would be easily obtained thanks to VisuVo, which captures fine-grained acoustic realisations of vowels in the course of L1 development, and allows an interactive representation of formant evolution over a time period.

**References**


Orthogonal development of schwa coarticulation and reduction in DET+N sequences?
Melissa A. Redford
Linguistics Department, University of Oregon, USA
redford@uoregon.edu

Unstressed vowels in English are said to be “reduced.” Phonetically, schwa and other unstressed vowels are shorter and lower in amplitude than vowels that receive lexical stress or phrasal accents. Although a phonological notion, vowel reduction is conceptually linked to coarticulation in the phonetics literature, even across fundamentally different theories of speech production. So, for example, Lindblom’s theory of duration-dependent undershoot assumes acoustic targets that are realized more or less well depending on the time available for articulation. Thus, Moon and Lindblom (1994) explain that adjacent consonants have a stronger influence on vowel production in longer words compared to shorter words because polysyllabic shortening in longer words decreases the time available for articulators to reach a position compatible with the acoustic vowel target. The same argument of duration-dependent undershoot applies to explain why unstressed vowels are more centralized than their stressed counterparts. Although the theory of Articulatory Phonology assumes spatial targets rather than acoustic ones, it also explains coarticulatory effects with reference to time. In particular, the theory defines coarticulation as the degree to which adjacent gestures are overlapped (coproduced). Further, schwa is viewed as “targetless” in the theory; its acoustic features are derived entirely from gestural overlap (Browman & Goldstein, 1994). In sum, across phonetic theories, unstressed vowel reduction is hypothesized to drive coarticulation by affecting the time available to achieve an acoustic/articulatory goal(s). Following the same logic, incomplete reduction should result in vowels that are less coarticulated with adjacent consonants and vowels because they are longer; a hypothesis that has been used by some to explain weaker coarticulation in disordered speech (Kent & Rosenbek, 1983). Here, we test the hypothesis in speech produced by school-aged children and adults.

Allen and Hawkins (1978) were perhaps the first researchers to observe that young school-age children do not reduce grammatical words in English to the same extent as older children and adults. This observation, which has since been confirmed several times in the literature (e.g., Goffman, 2004), led us to investigate the correlation between schwa reduction and coarticulation in 5-year-old and 8-year-old children’s productions of utterance-medial “the” compared to adult productions. The predictions were (a) that children would produce longer and louder schwa compared to adults, and (b) that schwa would therefore be less influenced by the phonological shape of the following noun in children’s speech compared to adults’ speech.

A total of 36 speakers participated in the study: two groups of 12 American English-speaking children, aged 5 and 8 years old, and one group of 12 American English-speaking adults. All participants produced two repetitions of 18 sentences with determined noun phrases (DP) in phrase-medial position. Sentences were always 9 syllables in length. The DP was always the direct object in the sentence. The preceding verb was in the 3rd person. The determiner was the definite article (i.e., “the”). The nouns were monosyllabic and had the same rhyme, but began either with a labial or a velar consonant (i.e., “bat” versus “cat”). Different verbs were used to vary the metrical context in which the determiner appeared since there is some indication that children may be more likely to produce grammatical words appropriately in a strong metrical context than in a weak context (see, e.g., Gerken, 1996). This manipulation turned out to have minimal effects on schwa reduction and coarticulation, and so is not discussed further here. Once elicited, sentences were coded for disfluencies and the vowels in all fluently produced DP targets were segmented. Duration, amplitude, and formant measures (at vowel midpoint) were extracted and analyzed as a function of age group, metrical context, and target noun onset.

As predicted, absolute schwa duration, measured in milliseconds, varied systematically with age, \(F(2, 32) = 4.33, p = .022\): schwa duration was significantly longer in 5-year-old children’s speech compared to adults’ speech (mean difference = 14.93, \(p = .006\)). Mean schwa durations in 8-year-
olds’ speech were intermediate to those produced by 5-year-olds and adults (5-year-olds’ M = 61.79, SD = 18.30; 8-year-olds’ M = 54.37, SD = 18.79; adults’ M = 47.02, SD = 6.39), and not significantly different from either. In addition, there was a statistical trend in the data towards an effect of age on relative schwa duration, F(2, 33) = 3.19, p = .054, with mean comparisons indicating a significant difference between 5-year-olds’ and adults’ productions (mean difference = −.72, p = .012; see Figure, top left). On average, the /æ/ in “cat” and “bat” was 3 times as long as schwa in 5-year-olds’ determiners (M = 3.12; SD = .77) and nearly 4 times as long in adults’ determiners (M = 3.82; SD = .03). These results are consistent with the expectation that younger children reduce grammatical words less than older children and adults.

The analysis on schwa amplitude strengthened the conclusion that “the” was less reduced in younger children’s speech compared to older children’s and adults’ speech. In particular, the analysis on relative vowel amplitude indicated an effect of age, F(2, 44) = 8.48, p = .001 (see Figure, top right). Post hoc tests indicated that 5-year-olds’ productions were significantly different from 8-year-old productions (mean difference = −.036, p < .001) and from adults’ productions (mean difference = −.037, p < .001). The /æ/ in “bat” and “cat” was often slightly lower in amplitude than the schwa in “the” in DPs produced by 5-year-olds (M = 0.99; SD = .04). In contrast, /æ/ was almost always louder than schwa in DPs produced by 8-year-olds (M = 1.02; SD = .05) and adults (M = 1.02; SD = .03). A soft-then-loud pattern is what we would expect if “the” is reduced relative to the noun it determines.

The results from the analyses on vowel formants were ambiguous with respect to the prediction that children’s less reduced schwas would also be less influenced by the phonological shape of the following noun. The analysis of normalized F1 values indicated that younger children’s schwa was more closed overall than adults’, F(2, 33) = 4.99, p = .013. In the context of the open vowel of the noun, this could indicate weaker vowel-to-vowel coarticulation in children’s speech compared to adults’ speech. It could also indicate stronger influence of the following consonant on children’s schwa production in DPs compared to adults’ productions. The analysis of normalized F2 values (Z3-Z2) indicated that schwa productions were influenced by the shape of the following noun, F(1, 164) = 856.65, p < .001. There was also an effect of age group, F(2, 33) = 4.75, p = .015, on schwa production: children’s schwa productions were somewhat less fronted overall than adults’ schwa productions, and schwa was more fronted before “cat” than before “bat” (see Figure, bottom center). The interaction between age group and noun was also significant, F(2, 164) = 5.97, p = .003. Whereas all 3 groups of speakers produced “the bat” phrases with normalized F2 values that were different for schwa and the subsequent /æ/, only children’s normalized F2 for schwa was the same their normalized F2 for /æ/ in “the cat” phrases. These results could indicate stronger vowel-to-vowel coarticulation in children’s production of “the cat” compared to adult productions of “the cat.” The normalized F2 values in adults’ schwa production before “cat” suggests that these were produced in with an even more advanced tongue position than /æ/, which is perhaps consistent with the palatalized /k/ expected before a front vowel context in this dialect. So, adults’ schwa productions may have been more influenced by the subsequent consonant than by the next vowel in “the cat” phrases. What is more certain is that children and adults differed in how they coarticulated schwa with upcoming material, but not necessarily in the degree to which they did so.

Overall, the results indicate that children do not reduce schwa in grammatical words to the same extent as adults. They may nonetheless coarticulate schwa in these words with an upcoming content word to the same degree as adults, albeit very differently than adults. In particular, the results on F2 suggest different dominant influences of upcoming phonological material on schwa production in children’s speech compared to adults’ speech: children’s “the” productions may be more influenced by a subsequent stressed vowel than adults’ productions. This particular result is compatible with the view that the guiding representations in children’s speech are less differentiated at the segmental level than adults’ representations (e.g., Goodell & Studdert-Kennedy, 1993). It is inconsistent with the hypothesized relationship between reduction and coarticulation as.
a trade-off between time and target achievement, which focuses on the role of sequential target attainment in motor control.

References

Figure: The determiner’s normalized vowel duration (top left), amplitude (top right), and mean F2 values (bottom center) are shown as a function of speaker group.
Bilingual experience of emotional meaning: priming effects between music and the first and second language

Miriam S. Tenderini¹, Esther de Leeuw²; Tiina M. Eilola¹; Marcus T. Pearce³

¹ School of Biological and Chemical Sciences, Queen Mary University of London, England
² School of Language, Linguistics and Film, Queen Mary University of London, England
³ School of Electric Engineering and Computer Science, Queen Mary University of London, England

Emotion words seem to differ from neutral words already at a very early stage of cognitive and neural processing (Kissler, 2006). However, it is unclear whether emotional meaning in a second language (L2) is processed in a similar way like in the first language (L1) and which factors may explain individual differences in the perceived emotional weight of the speaker L2 (Dewaele, 2008). By examining priming effects between music and emotion words in the L1 and L2 across bilinguals who have learned their L2 at different ages in life, this project opens a door to investigate connections between the perception and processing of emotion in music and language. The affective priming paradigm allows the investigation of early, partly unconscious processing stages of emotional information at the behavioural as well as neurophysiological level (Fazio, 2001, Goerlich et al., 2012). Music stimuli used as primes allow us to present single words from both languages in the same controlled emotional context.

The focus of this research was to investigate rapid processing of emotion words in the L1 and L2 of German-English bilinguals. Furthermore, we examined the effect of predictor variables such as age of L2 acquisition (AOA), length of residence in the UK (LOR) and amount of L1 and L2 use in different contexts of everyday life. This was achieved using two cross-modal affective priming experiments examining whether single emotion words in the bilinguals’ L1 and L2 interfered in a similar manner with the processing of affect conveyed by music. The first experiment examined music priming responses to the word stimuli while the second examined words priming responses to the musical stimuli. Reaction times (RTs) and electrophysiological neural responses (especially the N400 component) were analysed with respect to valence congruence of the stimulus (congruent and incongruent stimulus pairs) and language (L1, L2).

In Study 1 (music priming words in L1 and L2) participated 25 German-English late unbalanced bilinguals (16 females, 9 men, mean age = 37.08 ± 10.71 years). They learned English after the age of 8 years (mean AOA = 11.04 ± 1.79 years). At the time of the experiment, they had been living in the UK for at least 1 year (mean LOR = 7.30 ± 6.74 years) and used German and English to different degrees in everyday life in social contexts, at home and at work (with a dominance of English). Language proficiency was self-assessed on a six-point scale from (‘1’ = ‘very poor’ to ‘6’ = ‘native like’). The mean index of proficiency was 5.11 ± .81 (proficiency scores averaged for speech comprehension: 5.12 ± .73, speech production: 5.08 ± .76, reading: 5.32 ± .75 and writing: 4.92 ± .99).

In Study 2 (L1 and L2 words priming music) participated 25 other German-English late unbalanced bilinguals (16 females, 9 males, mean age = 31.61 ± 7.52 years). This group learned English as well after the age of 8 years (mean AOA = 9.91 ± 1.68 years) and had been living in the UK for at least 1 year (mean LOR = 6.43 ± 4.87 years) at the time of the experiment too. They used German and English to different degrees in everyday life in social contexts, at home and at work (with a dominance of English). Language proficiency was as in the previous study self-assessed on a six-point scale from (‘1’ = ‘very poor’ to ‘6’ = ‘native like’). The mean index of proficiency was 5.52 ± .56 (proficiency scores averaged for speech comprehension: 5.57 ± .59, speech production: 5.49 ± .51, reading: 5.65 ± .49 and writing: 4.39 ± .66).
In both studies, single words (L1/L2) with a positive (e.g. ‘friend’) or negative (e.g. ‘war’) valence were presented together with musical excerpts with positive or negative valence (happy and scary excerpts from Vieillard et al., 2008). These music-word pairs were either congruent in their valence (both stimuli had a positive or both a negative valence) or incongruent (a positive stimulus was paired with a negative one). Words were presented visually on a screen while music was presented via speakers. As explained above, Study 1 examined responses to word targets primed by music while Study 2 examined responses to music targets primed by words. In Study 1, musical excerpts were presented for 1000 ms. The presentation of the target stimulus was terminated by the participant making a rating by pressing a key on the response box. In Study 2, word primes were presented for 200 ms followed by music targets which were presented up to 1000 ms if no response was given earlier. The stimulus onset asynchrony was set to 200 ms in both studies. Both L1 and L2 words were tested in separate blocks, i.e. L1 then L2 or L2 then L1, counterbalanced across participants. Participant were exposed to the language of the forthcoming block by watching a short movie and listening to the recorded instructions in the respective language. Prime – target pairs were created and presented in a randomized fashion. Each stimulus was presented once in the first half and once in the second half of one block. Both, the German and the English block consisted each of 198 trials.

RTs were recorded in milliseconds starting with the target onset.

Electrophysiological data was recorded in reference to the right mastoid at a rate of 500 Hz from 64Ag/AgCl electrodes. Impedances were kept below 5 kΩ. EEG signals were amplified with a NetAmps 300 (EGI) amplifier. Data was pre-processed and analysed in matlab (R2014b) with the toolboxes EEGlab (13.6.5b) and fieldtrip (20170123). The data was filtered offline with a .5 Hz high pass filter and a 45-60 Hz cleanline filter. The data was then re-referenced to the left mastoid. Before the re-referencing the left mastoid was subtracted by half of the recorded voltage. The critical time window for the ERP analysis (N400 component) was between −100 and 1 000 ms with baseline correction performed relative to the 100 ms pre-stimulus activity. The N400 component was analysed with respect to the congruence of the stimulus pair and language (L1, L2).

For Study 2 (words priming music), it was hypothesized that in the L1 a difference would be observed between congruent and incongruent stimulus pairs such that congruent pairs show faster RTs, and less negative N400-amplitude than incongruent pairs. In priming studies, incongruent pairs have been found to produce slower RTs and a more negative N400 amplitude (e.g. Sollberger et al, 2003; Steinbeis & Koelsch, 2011). The central question of interest here is whether this differs between the L1 and L2. The effect was hypothesized to be weaker in the L2. However, it was also predicted that the amount of L1 and L2 use in different contexts of everyday life, as well as the LOR, would also be influential. For example, those bilinguals with a higher everyday L2 use would show differences in RT and N400 amplitude between congruent and incongruent stimulus pairs that are more comparable to the L1 than individuals with less L2 use. Given the largely shared musical culture between Germany and the United Kingdom, it was hypothesised that the results of Study 1 would show no such difference in priming strength of music for words in L1 and L2.

RTs of responses in accordance with the predefined valence of the stimuli were analysed with a 2 (negative prime, positive prime) × 2 (negative target, positive target) × 2 (German, English) repeated-measures ANOVA with prime, target, and language as within-subject factors. In both studies two participants were excluded due to too many incorrect answers (n = 1) or technical problems with the recording (n = 3). Preliminary results for RTs only indicate a three-way interaction between language (L1, L2), word-valence and music-valence in Study 2 (L1 and L2 words priming music), $F (1,22) = 5.188, p = .033, \eta^2_p =.191$. However, no such effect of language was found in Study 1 (music priming L1 and L2 words), $F (1,22) = .905, p =.352, \eta^2_p =.040$. When analysing the interaction between prime and target in the two languages in Study 2 separately L1 words clearly primed musical targets [F (1,22) =
30.464, \( p < .001, \eta_p^2 = .581 \) while this interaction with music is weaker in the L2 \([F (1,22) = 5.446, p = .029, \eta_p^2 = .198]\). These behavioural results suggest decreased integration of the emotional information communicated by L2 compared to L1 words. Conversely, music has a consistent emotional priming effect across L1 and L2. Analysis of both behavioural and EEG data is underway and full results will be presented at the conference.

References
As people age, they often experience considerable difficulty accurately understanding speech, but there is large variability in performance. This variability derives, in part, from different listener attributes that may pre-dispose the listener to relatively good or poor performance. Other factors that contribute to the variability in speech understanding performance of older listeners include the talker’s speaking style and the characteristics of the listening environment. Thus, a framework for considering the interaction between listener, talker, and situational attributes is critical for unraveling the nature of the older listener’s difficulty in understanding speech.

The first part of this presentation will review the relative importance of three listener attributes that are thought to contribute to speech understanding performance of older listeners in degraded listening conditions: hearing sensitivity, auditory temporal processing abilities, and cognitive capacity. Undoubtedly, the principal source of older listeners’ difficulty understanding speech is reduced signal audibility that accompanies age-related hearing loss. Predictions of speech understanding in noise based on audibility-based methods will be reviewed. However, other factors have been shown to contribute to speech understanding problems of older listeners as well. One of these factors is an age-related decline in the ability to process critical timing information in sound. The importance of auditory temporal processing for recognizing brief speech sounds as well as perceiving supra-segmental aspects of spoken messages will be considered. Another factor is age-related decline in cognitive abilities, particularly working memory and processing speed. Cognitive theories of aging that appear to be relevant for interpreting older listeners’ performance in difficult speech understanding tasks will be discussed as well.

The second part of the presentation will emphasize selected research findings from speech perception studies that compare older and younger adults’ performance for understanding speech presented in different listening environments (noise, reverberation), as well as for different types of talkers (those who speak at a fast rate or with a foreign accent). The relative importance of hearing sensitivity, auditory temporal processing abilities, and cognitive capacity in relation to performance for these types of challenging listening conditions will be highlighted.

The presentation will culminate in a discussion of the implications of these findings for everyday speech understanding in common listening situations. Additionally, the importance of listening experience and new directions for focused training toward improving older listeners’ ability to process speech in noise and rapid or foreign-accented speech will be highlighted.
Listening across the life span: What you have to do determines how you do it
Aina Casaponsa1, Viktoria Vianeva1, Michael Akeroyd1, Johanna Barry1,2
1 MRC Institute of Hearing Research, University of Nottingham, Nottingham, United Kingdom; 2 Nottingham University Hospital Trust, Nottingham, United Kingdom; 3 Department of Linguistics and English Language, Lancaster University, United Kingdom
a.casaponsa@lancaster.ac.uk

Introduction
There is considerable evidence to suggest that as people age, their ability to perceive and understand speech declines when listening with noise in the background. There is also evidence to suggest listeners compensate for these declines in perception by relying more on underlying cognitive abilities (Besser et al. 2015), or cues from linguistic context (e.g. Pichora-Fuller 2008).

It is still unclear what influences the apparent age-related downward trajectory in ability to listen in noise although hearing loss, and declines in cognitive abilities like working memory have been implicated (Heinrich et al. 2016). Most findings about age-related changes in listening are based on studies using tasks that have been specifically sensitized to assess the role of cognition or language in supporting listening. These studies typically omit a control condition where task performance in quiet is assessed. This makes it difficult to untangle effects specific to age-related changes in perception from effects specific to the task design.

Here, we compare outcomes from two different test designs regarding factors influencing speech perception abilities in older and younger adults when listening either in quiet or against a background of 4-speaker babble. Our aim was to assess the impact of task design on conclusions about how listening changes across the lifespan. One task (Cole et al. 1980; Roebuck et al. in revision) involved listening to speech over an extended period of time (continuous listening task) and detecting embedded speech errors (mispronunciations). The second task (BESST-UK, Barry et al. 2014) followed a commonly used task design and involved repeating out loud short sentences presented in isolation. In both tasks, linguistic context was modulated to assess the role of language in supporting listening. Non-verbal IQ and serial and working memory were measured to assess the role of cognition.

Assuming the two task are sensitive to the same influences, we predicted equivalent performance across the groups when listening in quiet, but poorer performance in the older adults when listening in noise. Consistent with Pichora-Fuller (2008), we predicted increased reliance on linguistic cues in the older adults for both tasks, as well as a direct association (positive correlation) with working memory.

Methods
Two groups of participants with normal hearing completed both the continuous listening and the BESST-UK tasks. The older adults (n = 32) were aged between 55-74 years (M = 64 years) and their audiograms indicated better-ear pure tone averages (PTA) between 2.5 – 41.5 dB (M = 15.8 dB). The younger adults (n = 32) were aged between 18 – 38 years (M = 23 years) and had better-ear PTAs between -3.6 – 15.7 dB (M = 2.3 dB). A cross-design was used such that half the older and younger adults (16 per age group) completed one task in quiet (either BESST-UK or continuous listening task; 65dB HL), and the other in 4-speaker babble (0dB SNR; 65dB HL) and vice versa. This design was chosen to allow us to collect data for each participant in both noise and quiet, while minimising priming effects due to task repetition.

The continuous listening task was a simple story lasting 16 minutes which was read by a male speaker. Participants were requested to press a button on a button box as fast as possible on
hearing a mispronunciation (n=108). The mispronunciations were either highly related to the preceding context or not (i.e. predictable versus unpredictable).

In the BESST-UK task, participants heard a short sentence in noise or quiet and immediately repeated what they heard. Only the final word was scored. As with the continuous listening task, this final word was either predictable or not from the preceding context.

A pure tone audiogram was obtained for all participants, and they all additionally completed tests of non-verbal IQ (WASI matrices), serial memory (digit span forwards) and working memory (digit span backwards).

**Results**

Results from the continuous listening task show an overall similar pattern of results between young and older adults. As expected, both groups were significantly worse at detecting mispronunciations in babble compared with quiet \[ \chi^2 = 36.41, p<.001; \chi^2 = 13.09, p<.001 \]. Reaction times showed a similar pattern across the task [older: F = 48.10, p<.001; younger: F = 48.81, p<.001], suggesting similar processing effects including a drop in arousal / attention over time.

Consistent with our primary prediction, older adults missed significantly more mispronunciations than younger adults in the babble condition only (percent correct detection in noise: older adults = 73%, younger: 82%; p = 0.02). A significant effect for predictability was also found for the older adult group only (p = .03). Correct mispronunciation detection was influenced by hearing abilities (Figure 1), but not by nonverbal-IQ or working memory (digit span backwards) (Figure 2). These results suggest that age-related declines in hearing loss, rather than age-related changes in cognitive function primarily contribute to age-related difficulty in perceiving speech in noise and in quiet in a continuous listening task.

**Figure 1:** Correlations between target detection (%) and hearing levels for younger and older adults in the story task in quiet and in babble conditions. Asterisks (*) indicate significant correlation (p < .05).

Results from the sentences in noise test (BESST-UK) also indicated greater difficulty listening in noise in the older adults (older adults = 76%, younger: 91%; p < .001). In contrast with the continuous listening task, there were marked predictability effects in both groups (all p < .001). Moreover, in addition to PTA, performance on this task correlated with working memory (Figure 2), and non-verbal IQ in the older adult group. These results suggest task-related effects contribute to the emergence of associations between cognition and speech-in-noise intelligibility measures.
**Discussion**

Studies investigating how listening abilities change across the life span report a mixture of findings. Evidence has been put forward suggesting both greater reliance by older listeners (e.g. Pichora-Fuller 2008) on language cues, as well as no difference in the use of these cues in older and younger listeners (Dubno et al. 2000). Evidence has also been provided suggesting an increased reliance on cognitive skills like working memory in older listeners when listening in noise (Besser et al. 2015).

We used two different listening tasks apparently designed to capture differences in speech perception in noise, but based on different task requirements i.e. detection versus recall. Our findings highlight the important role that task design plays in influencing study outcomes. The continuous listening task involving error detection suggested changes in listening were primarily related to age-related hearing loss. By contrast, the more cognitively demanding task involving recall of sentences presented in noise suggested that cognitive abilities are also implicated in age-related changes in speech-in-noise listening.

It is possible, that task specific influences may confound current understanding about the role that cognitive and linguistic abilities play in supporting listening across the life span.

**References**


Top-down and bottom-up perceptual learning for speech is maintained in older adults

Sarah Colby1,3, Meghan Clayards1,2, & Shari Baum1
1 School of Communication Sciences & Disorders, McGill University, Canada; 2 Department of Linguistics, McGill University, Canada 3 Corresponding author: sarah.colby@mail.mcgill.ca

Introduction. Perceptual flexibility is necessary for speech perception, as listeners must adapt to variation among speakers. We focus on two types of perceptual learning that have previously been studied in regards to speech perception: lexically-guided learning and distributional learning. Lexically-guided learning relies on the top-down influence of lexical knowledge to update phonetic categories after exposure to novel or ambiguous speech stimuli. Distributional learning, on the other hand, requires the listener to track relevant acoustic cues from a presented distribution to update their phonetic categories. Both of these types of perceptual learning have been suggested to be forms of implicit learning, with distributional learning updating the listener’s prior beliefs about a category based on the input (Kleinschmidt & Jaeger, 2015) and lexically-guided learning relying on error signals generated by lexical representations (Guediche, Fiez, & Holt, 2016). Older adults have previously been shown to adapt to ambiguous speech stimuli, with no difference found between older and younger adults when learning was driven by top-down processes, as in lexically-guided learning (Scharenborg & Janse, 2013). Given that older adults seem to show an increased top-down bias compared to younger adults (Mattys & Scharenborg, 2014), the question remains whether older adults will remain flexible when top-down information is not advantageous. We expected older adults to show comparable learning to younger adults in the lexically-guided (top-down) learning task, but that differences might emerge between the age groups when lexical information cannot be relied upon, as in the distributional (bottom-up) learning task. Older adults are known to employ top-down strategies to compensate for some of the known age-related declines in speech perception (Pichora-Fuller, 2008), and thus may have more difficulty in the latter task.

Method. A group of younger (n=31, ages 18-29) and older (n=27, ages 63-86) adults completed the two perceptual learning tasks, as well as several cognitive tasks, including measures of vocabulary, attention-switching, and working memory. The perceptual learning tasks were comprised of a pretest 2AFC task on an /ɛ/-/ɪ/ continuum, an exposure phase, and a posttest identical to the pretest, but repeated for 3 blocks. Exposure included both clear and ambiguous tokens designed to shift the category boundary between the two vowels towards more responses for the ambiguous end. In the distributional task, these clear and ambiguous tokens were presented in minimal pairs. For example, the /ɛ/-ambiguous condition presented clear tokens of /ɛ/, like ‘pen’, along with ambiguous tokens of /ɪ/ that fell between ‘pin’ and ‘pen’, and vice versa for the /ɪ/-ambiguous condition. Crucially, the lexically-guided task provided additional lexical information not present in the distributional task (e.g., the /ɛ/-ambiguous condition presented clear /ɛ/ tokens in words like ‘upset’, along with ambiguous tokens of /ɪ/ in words like ‘violin’, where ‘upsit’ and ‘violen’ are not real words). All participants completed both learning tasks over two sessions to allow for within-subject comparisons, and participants were assigned different exposure conditions and different voices for the two learning tasks to ensure learning in one task would not influence learning in the second (Table 1 outlines the counterbalancing of the learning tasks).

Results. Separate mixed effects logistic regressions were run for the lexically-guided and distributional learning tasks. The outcome measure was learning consistency (whether a 2AFC response was consistent with the expected shift; e.g., responding /ɛ/ for the /ɛ/-ambiguous condition is learning consistent). Learning consistency increased from the pretest to the posttest (Distributional: β=0.72, z=6.11, p<0.001; Lexically-guided: β=0.63, z=3.6, p<0.001), and thus participants showed learning in both tasks. We also find a significant effect of Exposure type for both
Table 1: Counterbalancing of learning tasks and example stimuli. ‘?’ indicates the ambiguous stimulus.

<table>
<thead>
<tr>
<th>Order</th>
<th>Task</th>
<th>Voice</th>
<th>Ambiguous Condition</th>
<th>Task</th>
<th>Voice</th>
<th>Ambiguous Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Lexically-guided</td>
<td>Male</td>
<td>/ɛ/: up?t – violin</td>
<td>Distributional</td>
<td>Female</td>
<td>/ɪ/: pen – p?n</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Female</td>
<td>/ɛ/: up?t – viol?n</td>
<td></td>
<td>Male</td>
<td>/ɛ/: pen – p?n</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Female</td>
<td>/ɛ/: up?t – viol?n</td>
<td></td>
<td></td>
<td>/ɛ/: up?t – viol?n</td>
</tr>
<tr>
<td>D</td>
<td>Distributional</td>
<td>Male</td>
<td>/ɛ/: p?n – pin</td>
<td></td>
<td>Female</td>
<td>/ɛ/: up?t – violin</td>
</tr>
<tr>
<td>E</td>
<td>Distributional</td>
<td>Male</td>
<td>/ɛ/: p?n – pin</td>
<td></td>
<td>Female</td>
<td>/ɛ/: up?t – violin</td>
</tr>
<tr>
<td>F</td>
<td>Distributional</td>
<td>Female</td>
<td>/ɛ/: p?n – pin</td>
<td></td>
<td>Male</td>
<td>/ɛ/: up?t – violin</td>
</tr>
<tr>
<td>G</td>
<td>Lexically-guided</td>
<td>Male</td>
<td>/ɛ/: up?t – violin</td>
<td></td>
<td>Male</td>
<td>/ɛ/: up?t – violin</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>Female</td>
<td>/ɛ/: p?n – pin</td>
<td></td>
<td></td>
<td>/ɛ/: up?t – violin</td>
</tr>
</tbody>
</table>

tasks (Distributional: β=3.62, z=8.19, p<0.001, Lexically-guided: β=3.27, z=6.37, p<0.001), which points towards a bias in our data to respond with /ɛ/. Recall that learning consistency is scored in relation to Exposure type, so both exposure groups show an /ɛ/ bias, but for the /ɛ/-ambiguous group this is reflected as learning-consistent behaviour, and for the /ɪ/-ambiguous group it appears learning-inconsistent (the difference between blue and red lines in Figure 1). This /ɛ/ bias clouds the learning effect for the /ɛ/-ambiguous group, as a high number of /ɛ/ responses in the pretest leaves them with little room to increase /ɛ/ responses. For the distributional learning task there were two marginally significant interactions between Age group and Block, suggesting that older adults learned less overall (Pre vs. Posttest: β=0.41, z=1.76, p=0.07) and ‘unlearned’ more (i.e., showed less learning consistency) in the last two blocks (Post 2 vs. Post 3: β=0.34, z=1.38, p=0.07). However, this was qualified by a significant three-way interaction among Exposure type, Age group, and Block (Post 2 vs. Post 3), which indicates that the unlearning was most pronounced for the older adults in the /ɪ/-ambiguous exposure group (β=1.01, z=2.15, p=0.03; Figure 1A). For the lexically-guided learning task, there was no clear evidence of a difference between age groups (Age group x Block: β=-0.2, z=-0.56, p>0.05); however, there was evidence that older adults in the /ɪ/-ambiguous group showed increased learning compared to the younger adults, while those in the /ɛ/-ambiguous group showed decreased learning (Age group x Block x Exposure type: β=2.33, z=3.31, p<0.001; see Figure 1B). Taken together, we find that older adults maintain both types of perceptual flexibility, but a ceiling effect in the /ɛ/-ambiguous exposure group may be masking any clear differences in learning between the age groups.

We were also interested in investigating the individual differences in older adults’ cognitive functions that may affect lexically-guided and distributional learning. Two additional models were run on only the older adult posttest data. Neither model found evidence for an effect of

Figure 1: Average learning consistent behaviour by age group, exposure type, and block for (A) the distributional learning task and (B) the lexically-guided learning task.
performance from the other task (i.e., distributional learning did not predict lexically-guided learning, and vice versa; Distributional: $\beta$=-0.30, $z$=-0.85, $p$>0.05, Lexically-guided: $\beta$=0.02, $z$=0.04, $p$>0.05), suggesting that these are two distinct types of learning. Older adults with poorer attention-switching control showed more learning-consistent behaviour in the lexically-guided learning task ($\beta$=0.90, $z$=2.07, $p$=0.03), as previously shown by Scharenborg, Weber, & Janse (2015). This result supports the idea that maintaining task-level attention is beneficial to lexically-guided learning, while switching attention to the signal is detrimental (McAuliffe & Babel, 2016). Thus, the older adults with poorer attention switching, who maintained task-level attention, showed more learning-consistent behaviour than those who switched more between different levels of attention. Additionally, participants with larger vocabularies showed more learning-consistent behaviour in the lexically-guided learning task ($\beta$=1.39, $z$=3.49, $p$<0.001), suggesting that a more robust lexical network may facilitate learning in this kind of task. Hearing sensitivity was the only factor that predicted learning in the distributional learning task ($\beta$=1.15, $z$=2.47, $p$=0.01); we found that older adults with poorer hearing showed more learning. Older adults with poorer hearing may be more flexible in this kind of bottom-up task to accommodate their less reliable auditory input, while those with better hearing will not adapt their category boundaries in response to input that is too different from their existing category. No effect of hearing sensitivity was found for the lexically-guided learning task, supporting the idea that older adults have established strategies to compensate for degraded sensory input, and are able to perform similarly when additional top-down information is available.

**Conclusion.** The current study provides evidence that older adults remain perceptually flexible in both top-down and bottom-up tasks, but that learning in each task is distinct. As reported by Scharenborg & Janse (2013), we find that older adults perform comparably to younger adults in a lexically-guided learning task, and we extend this to show that older adults also maintain their perceptual flexibility in a distributional learning task, where they are required to update phonetic categories without reliance on top-down context. While both tasks require implicit perceptual learning, their underlying mechanisms are different and could have been affected by aging in different ways. We have also identified several factors that play a role in either the distributional or lexically-guided learning of older adults. An individual’s lexically-guided learning is predicted by vocabulary size and attention-switching control, while distributional learning is predicted by hearing sensitivity. This highlights the idea that different mechanisms are important for different types of perceptual learning, and likely reflect the resources that are called upon for each specific task.

**References**


Statistical learning for speech segmentation: Age-related changes and underlying mechanisms
Shekeila D. Palmer, Sven L. Mattys
University of York, United Kingdom, shekeila.palmer@york.ac.uk

Research on language acquisition has shown that infants and young adults are able to extract words from an unfamiliar speech stream based solely on the transitional probabilities between syllables within the stream (cf. Saffran, Aslin, & Newport, 1996). However, little is known about how this aptitude for ‘statistical learning’ (SL) changes over the lifespan, or about how it interacts with general age-related cognitive decline. One of the reasons why it is difficult to predict how SL abilities might change over the lifespan is that there is little agreement with regard to the mechanisms that underpin this form of learning. Since SL is known to operate in infants, it is usually assumed to be an automatic learning mechanism that requires few cognitive resources (e.g. Fiser & Aslin, 2001; Saffran et al., 1996; 1997; Turk-Browne, et al. 2005). From this perspective, one might expect that SL abilities should be relatively resilient to cognitive ageing. Indeed, within the ageing literature, it is often argued that, while cognitive functions such as working memory, executive control, and processing speed show age-related decline, implicit and automatic processes remain comparatively stable throughout life (e.g. Fleischman et al., 2004; Fleishman, 2007; Jennings & Jacoby, 1993).

However, recent evidence suggests that speech segmentation by SL may not be fully automatic (Palmer & Mattys, 2016; Torro, Sinnett, & Soto-Faraco, 2005). For example, Palmer and Mattys (2016) observed that SL performance was inversely related to speech rate: The slower the stream, the better SL performance. However, the benefit associated with slowing down the rate was eliminated when participants were asked to perform a concurrent visual 2-back task while listening to the speech stream. This suggests that SL performance is supplemented by an active processing or maintenance mechanism, which operates more effectively when processing time is increased, but is disrupted when central resources are depleted by a concurrent cognitive load. Since the level of disruption was independent of load type (phonological vs. non-phonological 2-back task), Palmer and Mattys (2016) hypothesised that this mechanism may involve domain-general executive resources, in particular, working memory updating. Given that working memory updating is known to show marked age-related decline (De Beni & Palladino, 2004; Van der Linden, Bredart, & Beerten, 1994), this process may be less efficient in older adults, thereby leading to an age-related decrease in SL performance.

The aim of the current study was twofold. First, we aimed to assess the effect of age on speech segmentation by SL and, second, we wanted to gain further insight into the processes that underlie performance in this task. Following Palmer and Mattys (2016), we assume that SL in adults includes an implicit (or incidental) component, which operates automatically, and a more explicit (or active) component, which relies on working memory. Since older adults generally show a decline in working memory resources, we hypothesised that they may show weaker SL performance compared to their younger counterparts. However, when the resources required for working memory processes are depleted by a cognitive load task, the difference in SL performance between younger and older adults should be attenuated, since SL becomes more dependent on incidental learning mechanisms. Furthermore, given that slowing down the speech rate seems to benefit performance in young adults, presumably by allowing more time for working memory operations to contribute to learning, we expect that slow speech is less likely to benefit older adults. This is because older adults may be less able to capitalise on the increased processing time afforded by the slower rate.

Method
During the experiment, groups of young (18 – 25 years), middle-aged (40 – 50 years), and older (60 – 81 years) adults completed four SL-based speech segmentation tasks, using four different artificial
languages, across four separate testing sessions. Across the four testing sessions, we manipulated speech rate and cognitive load, with each participant performing a speech segmentation task on streams played at two different rates (normal vs. slow), both with and without cognitive load (a concurrent visual 2-back task). The speech streams were always made up of 4 mixed length words (two 3 syllable words and two 4 syllable words). Learning was assessed using a two-alternative forced-choice task in which words from the stream were pitted against either nonwords (easy W-N trials), which never appeared in the stream, or part-words (harder W-P trials), which occurred across word boundaries. Finally, in order to gain greater insight into the processes that underlie SL, and to help account for any age differences observed, all of our participants completed a battery of neuropsychological tests designed to measure working memory, executive function, vocabulary knowledge, and non-verbal reasoning. Hearing Thresholds were also measured.

**Results and Discussion**

The results for the statistical learning task are shown in Figures 1-3. In general our data indicate that speech segmentation by SL is remarkably resilient to age-related decline, though some age effects were visible in the more challenging conditions, namely on W-P trials and when SL was performed under cognitive load. Since the part-words used as foils on W-P trials were heard during learning, albeit less frequently than the words, it could be argued that W-P trials constitute a stronger test of SL than W-N trials, in which the nonword foils were never heard before. Indeed, succeeding on W-P trials is contingent on acquiring distinct and precise representations of the words in the stream, whereas succeeding on W-N trials only requires a general gist of familiar-sounding sequences. Therefore, although older adults appeared unimpaired in SL when performance was considered across all trial types, age deficits were visible when more sensitive measures of statistical learning were used. Likewise, although the interaction between age and cognitive load was not significant for W-P trials, the older adults were the only group who did not perform above chance on these trials in the cognitive load conditions. The young and middle-aged groups both continued to learn under cognitive load, albeit to a lesser extent than in the no load conditions. These findings suggest that older adults have more difficulty using transitional probabilities to form accurate word representations than younger adults, and that this age-related deficiency is exacerbated under cognitive load. Stream rate was not a significant predictor of performance in this experiment.

A hierarchical regression analysis of the cognitive test data indicated that performance on W-P trials was mostly predicted by memory updating. Since the older adults performed worse on the memory updating task than the other age groups, this would seem to explain, at least in part, why they had more difficulty with W-P trials. Interestingly, performance on W-P trials was not related to forward or backward digit span, indicating that it is specifically the WM updating function that matters for these more difficult trials. This provides more concrete evidence for Palmer and Mattys’ (2016) suggestion that executive resources are recruited during SL. It was proposed that working memory...
updating may assist SL by removing and replacing erroneous syllable grouping from working memory, leading to the more accurate representations required to distinguish words from other familiar sounding sequences.

Performance on the easier W-N trials was less affected by age. This finding is consistent with the results of our regression analysis which revealed that working memory storage capacity (as measured by forward digit span), rather than working memory updating, was the strongest predictor of performance on these trials, since older adults performed no worse than the young adults on measures of working memory capacity. We argue that performance on W-N trials relied more on a general feeling of familiarity with test sequences than on computation of transitional probability. The only condition in which age effects were observed on W-N trials was the Slow rate, cognitive load condition, where again the older age group performed at chance level. We suspect that this may be due to the age-related binding deficit which has been widely reported in the ageing and memory literature (e.g. Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000; Naveh-Benjamin et al., 2003). Within the context of SL, this binding deficit would manifest as an increased difficulty in grouping recurring adjacent syllables into units, which would be magnified by the slower rate and become observable under cognitive load, where the resources required to actively maintain units held in memory are depleted.

References
Speech constitutes the primary channel of human social interaction, yet speaking can be considered the most complex skill humans perform. For most people, speech production is effortless and generally error-free, and most children acquire speech skills relatively automatically with little difficulty. However, some children struggle with the acquisition of speech production skills and require sustained and intensive treatment (Campbell, 1999). Children that suffer speech-language disorders are at increased risk for social-emotional and behavioural problems (e.g., Conti-Ramsden & Botting, 2004; Van Daal, Verhoeven, & Van Balkom, 2007), as well as for delayed development of language, literacy and other academic skills (e.g., McCormack, McLeod, McAllister, & Harrison, 2009). These issues tend to threaten employment and occupational opportunities in adulthood (e.g., Felsenfeld, Broen, & McGue, 1994; Snowling, John, Adams, Bishop, & Stothard, 2001). Accurate diagnostic methods and effective intervention programs are thus of crucial importance to limit the short- and long-term impact of speech-language disorders on the individual. Unfortunately, the classification of paediatric speech disorders and treatment planning remains problematic.

The development of clinical instruments requires fundamental knowledge about speech processing and its development. Over the years, a variety of models have been presented in the literature to describe the process of speech production, each with their own specific approach, scope, and theoretical basis (for an overview, see e.g., Terband, Maassen, & Maas, 2016a, 2016b). Traditionally, the scientific discussion on which model gives the best account of all the different speech phenomena focuses on the differences between models (which is of course an important part of the scientific endeavour); however -and especially important from a clinical perspective-, it should be noted that the different models also show similarities and overlap. First and foremost, all models of speech production adopt a hierarchical structure of control, similar to any model of complex motor performance. Furthermore, and more specifically, all models agree that the production of speech involves a preparatory psycholinguistic process of producing a sequence of one or more word forms (a phrase) stored in some short-term memory (buffer), followed by a process that calculates (process of encoding, transcoding, planning, programming) the speech movements that must be executed to articulate the sequence (phrase). (Terband et al., 2016a, 2016b)

A number of different systems for the classification of paediatric speech disorders has been proposed throughout the years, in which subtypes are differentiated on the basis of their theoretical distal or proximal cause (for an overview, see e.g., Terband et al., 2016a; Waring & Knight, 2013). The definitions used in these classification systems generally refer to speech production processes, and accordingly, a variety of methods of intervention has been developed aiming at different parts of the speech production process. In general, the following categories or subtypes are distinguished (e.g., ASHA, 2007; Dodd, 2014):

- Consistent phonological disorder: predominantly phonological processes;
- Inconsistent phonological disorder: phonological and/or sensorimotor processes;
- CAS: predominantly sensorimotor processes;
- Phonetic articulation disorder: isolated phonetic processes; and
- Dysarthria: execution processes.

However, diagnostic classification is primarily based on a description of behavioural speech symptoms, rather than on assessment of the underlying deficits. This raises a fundamental problem, as core impairments at different levels of speech development and different parts of the speech production chain cannot be clearly distinguished from each other at the symptom level.
Based on the similarities between the different speech production models, Figure 1 (left panel) presents a cascade style hierarchical model of the linguistic, phonetic, and sensory-motor functions involved in speech processing. In speech acquisition in infants and children, however, the situation is more complicated as the different functions are not prespecified in the infant brain (see Figure 1, right panel). Rather, there is a gradual emergence of the adult system (e.g., Bishop, 1997; Karmiloff-Smith, 2006) and the different functions and representations develop partly simultaneously and in interaction. The effects of developmental interaction between the different levels on individual development of speech production are largely unknown, and individual children may vary widely in these developmental interaction patterns. The diagnostic dilemma is that the ability to investigate the characteristics of subtypes of paediatric speech disorders requires ‘pure’ cases selected on the basis of unambiguous/clear-cut criteria. These criteria can only be defined and made available as a result of research (into a priori undefined/undetermined cases).

**Figure 1** (Terband et al., 2016a, 2016b). Left panel: Adult model of speech processing, adapted from Levelt (1989) and Van der Merwe (1997), displaying the sensory-motor and memory functions involved in speech production and perception. Right panel: Infant model. Processing functions are yet to be established and will emerge gradually during development (cf. Levelt, 1998; Maassen, 2002).

In this talk, I will discuss a process-oriented approach to diagnosis and treatment planning of paediatric speech disorders that allows us to break through this circularity (Terband et al., 2016a, 2016b). Its three-pillar scientific basis is formed by (1) individual rather than group data, (2) longitudinal rather than cross-sectional data, and (3) utilization of auxiliary technologies (simulations with computational models). The core of this approach comprises three important notions:

1. Although the behavioural symptomatology of paediatric speech disorders is not completely clear, it is possible to precisely define a specific core problem in terms of processes. (A focus on underlying processing deficits rather than classification based on symptoms.)
2. Developmental interaction between processes: a specific underlying impairment on one level or domain also affects the development on adjacent levels or domains. (A focus on process profiles with degrees of involvement.)
3. The speech production system and –disorder develop/evolve in time. (A focus on changing profiles.)
Based on these three notions, a model of differential diagnosis and treatment planning for childhood speech disorders is proposed, that—besides “fluency disorder”—comprises two general diagnostic categories labelled “speech delay” and “developmental speech disorder”. Within these categories, treatment goals are formulated on the level of processes. This process-oriented approach to diagnosis and treatment planning holds important advantages. In contrast to diagnostic classification based on a description of behavioural symptoms, it offers direct leads for treatment aimed at the specific underlying impairment tailored to the specific needs of the individual that is evaluated, and adjusted in the course of the speech disorder. The approach is illustrated with an example.

References
The development of speech perception in noise: the contribution of temporal-processing
Laurianne Cabrera¹, Christian Lorenzi², Stuart Rosen¹
¹University College London, London, UK
²Ecole Normal Superieure, Paris, France
laurianne.cabrera@gmail.com

Temporal cues (e.g., amplitude modulation, AM) play a crucial role in speech intelligibility for adults (1). How the ability to track these temporal cues develops and interacts with the development of speech perception is however unclear. Although aspects of AM processing appear to be mature as early as 3 months of age (2), children’ ability to detect AM continues to improve until 10 years of age (3). The present study explores whether the development of AM processing is related to sensory or cognitive factors and how this ability relates to speech intelligibility in noise during childhood.

AM masking and temporal integration were measured to characterize sensory (AM encoding) and cognitive (memory and attention) mechanisms. Eighty-three normal-hearing children from 5 to 11 years and 40 adults completed three 3IAFC adaptive tasks. The first task assessed AM sensitivity using 3 modulation rates (4, 16, 32 Hz) and AM masking using 3 carriers varying in their inherent AM fluctuations (tones, narrowband noises with small inherent AM fluctuations and noises with larger fluctuations). The second task assessed temporal integration, the effect of increasing the number of AM cycles (duration) on AM detection using tones modulated at 4 or 32 Hz. Finally, thresholds for consonant identification were measured in speech-shaped noise using fricative consonants (/f/-/v/-/ʃ/-/ʒ/-/s/-/z/).

Results show differences in AM sensitivity between younger children and adults only when small fluctuations are introduced in the carrier. This suggests that sensory mechanisms constraining AM processing are mature by 5 years, but that some higher-level factors limiting AM processing (e.g., internal noise) reach adult level around 8 years. Results also show that AM sensitivity improves with increasing number of AM cycles for all ages. Interestingly, AM sensitivity significantly improves with age until 11 years when it is measured using two AM cycles only.

Regarding the intelligibility of speech in noise, age and vocabulary scores but also temporal integration scores and AM masking for noises with small AM fluctuations seem to contribute to the variability in performance. Thus, central factors such as internal noise and echoic-memory/attention capacities for AM may play a crucial role in the development of speech intelligibility in noise.

References
Introduction
Since the early 1990s, cochlear implants (CIs) have been gaining popularity and their criteria for candidacy are getting broader, especially in severely to profoundly hearing-impaired patients (Bruijnzeel, Ziylan, Stegeman, Topsakal, & Grolman, 2016). Early cochlear implantation, before the age of 18 months, is considered nowadays, by many centres, as one of the best indicators of success (Cuda, Murri, Guerzoni, Fabrizi, & Mariani, 2014). The acquisition of appropriate speech prosodic structure by children is a very important element in the dynamics of communication and interaction in order to transmit the real meaning of the message produced. Yet, CI implanted children display an array of speech production and perception problems, both at the segmental and at the prosodic level.

As for the prosodic level, the major problems reported in the speech patterns of CI children concern rate, loudness and laryngeal quality (Lenden & Flipsen, 2007). Lyxell et al (2009) reported lower scores in prosody in CI children when compared to normal-hearing (NH) children at the level of word and phrase production. Intonation performance has also been found to be lower in CI children as compared to NH children. Performance however increased with prolonged use of CI (Peng, Tomblin, Spencer, & Hurtig, 2007; Peng, Tomblin, & Turner, 2008).

Despite a certain amount of knowledge of prosody production in CI children, none of the existing studies have evaluated the specific use of prosody to structure information in the discourse (i.e., focus vs. background information), especially in the French speaking population. Hence, the aim of this study is to evaluate prosody production in early implanted, prelingually hearing-impaired children in comparison to age-matched NH children and adults, to determine whether and when they acquire prosodic focus structuring.

Methods
This is a cross-sectional study of prelingually hearing-impaired French speaking children, without any other anomalies, who received CI before the age of 18 months between 2009 and 2012 at La Timone Children’s Hospital in Marseille, France. Three groups of participants were included. Twenty profoundly prelingually hearing-impaired early CI children, aged between 4 and 7 years. Twenty NH age-matched children and 20 NH caregivers of the CI children.

The speech production task consisted in playing a computer-based semi-structured game, where children interacted with the caregiver in the presence of the experimenter. We elicited sentences containing words that were contrasted in the discourse, like in the sentence *Il faut prendre le BONNET orange* “You need to take the orange HAT” (as opposed to another hat with a different colour) where capital letters indicate the prosodically focused item. Narrowly focused elements could either be the object or the adjective within a Noun Phrase. The same Noun Phrase could also be entirely focused in the Broad Focus condition. CI children were tested during their routine CI fine-tuning session. Caregivers had to interact with the children in order to produce the same words. Every session was recorded using a high definition microphone in a soundproof room. Recordings were analysed using Praat (Boersma, P; Weenink, 2016). Prosodic characteristics of the target utterances (pitch accents and boundary tones) were labelled and compared between the different groups. Moreover, three raters who were blind as to the group membership of each participant were asked to determine the focus pattern of each recording, and scored each recording between 1 and 10 according to the supposed pattern of focus of the target sentence.
Results were interpreted according to both chronological age and CI use (also called hearing experience or hearing age).

Results
The prosodic focus pattern of the different participants was scored by the 3 raters. Our preliminary results showed, a mean score of 6.5 for the CI group, 7.5 for the NH children and 9 for the adults. As for the prosodic characteristics of the produced sentences, and especially of the targeted words, we have found that CI children start using prosodic cues to emphasize discourse elements at the age of 5 years, and start to have adult-like patterns at the age of 7. However, NH children start to use prosodic cues earlier, at the age of 4, and start to become adult-like at the age of 6. Furthermore, we analysed the focus pattern of the CI children using their hearing age, which corresponds to the duration of the use of the CI device. Our preliminary results showed that the mean hearing age at which they acquired prosodic focus capacities was around 4.5 years.

Discussion
In this study, our findings showed that early implanted French-speaking children acquire prosodic cues of speech production somewhat later compared to their NH peers. These results are not completely in line with the data found in the literature, showing that CI children have difficulties in acquiring prosodic aspects of speech in production as well as in perception (Chin, Bergeson, & Phan, 2012; Green, Faulkner, & Rosen, 2004; Peng et al., 2007, 2008). These differences might be mainly due to 2 important elements. First, no study has yet evaluated prosodic focus in speech production in French speaking CI children. Second, the majority of studies does not distinguish between early and late implanted children. In fact, as it has been recently shown, the earlier the child is implanted, the more similar is his/her speech production to NH children (Bruijnzeel et al., 2016).

The importance of the age at implantation is mainly due to the neuroplastic and neurolinguistic dynamics in child development. It is optimally beneficial to offer CI during the critical period of cortex neuroplasticity in which speech and language are developed and mastered (Melo & Lara, 2012; Sharma, Dorman, & Spahr, 2002; Sharma, Gilley, Dorman, & Baldwin, 2007). In our study, using hearing age (or hearing experience) as a criterion in the analysis yields better results. This is consistent with other studies showing that the longer the hearing experience, the better the resulting speech production pattern. Hearing age is hence a better criterion which should be used when comparing CI children to their NH peers (Chin et al., 2012; Peng et al., 2008).

Furthermore, we found other variables that might influence speech production in CI children. Having bilateral CIs yields better hearing results, thus better speech results. Also, the influence of caregivers can play an important role on the prosodic production of their CI children. The parents’ partnership in the rehabilitation of their CI children render better results in their speech production performance (Melo & Lara, 2012). In other words, the absence of a regular rehabilitation follow-up and a good caregiver investment in this program can put the child in serious difficulties in terms of speech and language production performance.

It is important to mention certain limitations to our study. First, the small sample sizes can limit the extrapolation of the results to a larger population. Second, the fact that testing CI children after having their CI fine-tuning session can reduce their performance in terms of attention and cooperation, but increases their hearing performance optimising their speech production.

Conclusion
The present study is the first to evaluate French speaking early CI children prosodic focus in speech production and is an important step towards understanding the rehabilitation needs of these
children for a better communication. Our results suggest that CI children can use prosody to emphasise discourse elements, though later than NH children. Despite the delay, CI children quickly catch-up with their NH peers’ performance curve and become eventually adult-like, especially hearing experience is taken into consideration. Finally, their performance is highly influenced by their caregivers’ speech patterns and their post-CI rehabilitation follow-up.

References
Listening comprehension across the adult lifespan
Mitchell Sommers
Department of Psychology, Washington University, USA

Although age-related declines in perceiving spoken language are well established, the primary focus of research has been on perception of phonemes, words, and sentences. In contrast, relatively few investigations have been directed at establishing the effects of age on understanding extended spoken passages. Moreover, most previous work on age and spoken language has used extreme-groups designs in which the performance of a group of young adults is contrasted with that of a group of older adults. Little if any information is therefore available regarding changes in listening comprehension across the adult lifespan. Accordingly, the goals of the current investigation were to determine whether there are age differences in listening comprehension across the adult lifespan and, if so, to identify the sensory and cognitive declines that contribute to impaired comprehension.

The study used a cross-sectional lifespan design in which a minimum of 50 individuals in each of seven decades, from age 20-90 (a total of 433 participants), were tested on three different measures of listening comprehension. In addition, we obtained measures of auditory and cognitive functioning from all participants and used structural equation modeling (SEM) to compare models varying in the relative importance of sensory and cognitive abilities in predicting comprehension performance across the lifespan. In contrast to the well-documented declines in hearing thresholds and cognitive abilities, performance on the listening comprehension measures remained relatively stable until age 65-70, after which significant declines were observed. Follow-up analyses indicated that this same general pattern was observed across three different types of passages (lectures, interviews, narratives). SEM revealed that increased hearing thresholds and reduced working memory produced direct negative effects on listening comprehension. However, the best fitting model included a component that measures crystalized intelligence and that showed a positive relationship with age and listening comprehension. These findings suggest declines in sensory and cognitive factors are associated with age-related changes in listening comprehension, but that the negative consequences of such declines are partially offset by increases in crystalized intelligence.
Partner-directed gaze in a communicative task: the effects of age and communication barrier

Chris Davis\textsuperscript{1,2}, Jeesun Kim\textsuperscript{1}, Outi Tuomainen\textsuperscript{3} and Valerie Hazan\textsuperscript{3}

\textsuperscript{1}The MARCS Institute, Western Sydney University; \textsuperscript{2}HEARing CRC; \textsuperscript{3}Speech Hearing and Phonetic Sciences, UCL, UK

When people engage in a task that involves looking away from their communicative partner, they will nonetheless occasionally make eye contact to maintain what Lindblom (1990) called the phenomenon of communicative empathy. Lindblom suggested that patterns of eye contact help interlocutors establish communicative goals and to take into account the receiver’s point of view. More concretely, eye contact is important for scheduling turn-taking and pausing as well as for understanding the attentional disposition of the interlocutor. Looking at the speaker becomes even more important when there is a barrier to spoken communication, e.g., when speech occurs with background noise or the listener has hearing difficulties. Under such conditions, seeing the speaker (auditory-visual, AV speech) provides a considerable boost to speech intelligibility (Sumby & Pollack, 1954).

The current study used the Diapix task (Van Engen et al, 2010) to examine patterns of face-to-face engagement (i.e., eye contact when speaking) for young and older adults. Note that the task does not require that the participants actually look at each other. That is, the task is based on the familiar game of ‘spot the difference’; the two participants are given subtly different versions of a cartoon-style picture and the task requires them to cooperate to find all of the differences without either seeing the other person’s version. We tested young and old participants to examine whether the frequency and duration of looks to the interlocutor would vary as a function of whether or not there was a communication barrier.

We expected that gaze frequency would vary depending on the clarity of the spoken message based on a small scale study that examined gaze interaction in quiet and noise conditions in the map task (Mixdorff, Pech, Davis, & Kim, 2007). Here it was found that in noise the frequency of mutual gaze was more than double in babble noise than in silence. When speech is presented in noise, older adults get a similar visual speech intelligibility benefit as younger adults (Tye-Murray et al, 2016). If, when there was a communication barrier, the perceiver simply looked to the speaker to obtain a visual speech benefit, then we would expect that the frequency of gazes to should be similar for both age groups. However, it may be that additional factors may determine partner-directed gaze. For instance, in a recent study that examined communicative interaction in a card matching task it was found that older adults had lower rates of partner directed gaze than did younger ones (Lysander & Horton, 2012). This pattern of results was explained in terms of the possible influence of the older adult’s cognitive capacities, i.e., due to poorer short-term-memory older adults needed to examine the to-be-described cards rather than looking their partner. This explanation would predict that the frequency of the older adult’s looks to their partner would be lower in the no barrier condition; and this pattern may also occur in the more difficult barrier conditions (as these may place extra load on planning and memory).

Method

56 single-sex pairs of native Southern English adult talkers between the ages of 19 and 84 years were recruited to participate in the DiapixUK task (Baker & Hazan, 2011). Each participant was assigned a role of a primary talker (‘Talker A’) or a secondary talker (‘Talker B’). Primary talkers were divided into two age groups: ‘younger adults’ (YA) between 19-26 years of age (13 F, 8 M; Female Mean age =21.5 years, Male Mean age = 20.5 years) and ‘older adults’ (OA) between 65-84 years of age (21 F, 12 M; Female Mean age= 71.6 years, Male Mean age = 75.4 years). Secondary talkers were always younger adults (between 18-30 years of age) of the same sex as the Talker A. One ‘No Barrier’ (NB) and three barrier conditions were tested. In the NB condition, both talkers heard normally. In the
‘hearing Loss’ (HLS) condition, Talker B had simulated severe-to-profound hearing loss (see below). In the BAB1 condition, Talker B heard Talker A in 8 talker babble noise (that varied in SNR, see below). In the BAB2 condition, both Talkers heard each other in 8 talker babble noise (at 0 dB SNR). For each task, pairs always began with a NB condition and the remaining three barrier conditions were counterbalanced across groups.

In the HLS condition, Talker B heard Talker A via a real-time hearing loss simulator modelling a profound sensorineural loss at levels 40-50-60-90-90 dB at frequencies 250-500-1000-4000-8000 Hz; (HeLPS, the Hearing Loss and Prosthesis Simulator, Zurek & Desloge, 2007). In BAB1 condition, Talker B heard Talker A in 8-talker babble noise that was similar in degree of communicative difficulty to the HLS condition. The SNR for the BAB1 condition was individually set using an adaptive procedure to equate HLS performance on the Modified Rhyme Test (MRT). In the DiapixUK task, participants were told that the pictures contained 12 differences. Talker A was leading the conversation and was instructed to do most of the talking, whereas Talker B was mainly required to ask questions and make suggestions. They were given 10 minutes to find these differences (one picture per condition). Auditory and video recording were made of Talker A (the auditory recording with an Eagle G157b lapel microphones and the video with a 640 x 480 (VGA) camera at 30 fps) and the data transferred to ANVIL AV annotation software (Kipp, 2014). For the data presented in the study, the number and duration of partner directed gazes were annotated.

Results and Discussion

Figure 1 shows the Average number of partner directed gazes for the old and young participants as a function of communication condition. Overall there was an effect of communication condition, F1(3,150) = 50.17, \( \eta^2_p = 0.5 \), but no effect of Age, F1(1,50) = 0.34, p = 0.57, or Sex, F1(1,50) = 2.37, p = 0.13. Analysis of the older participant data revealed an effect of Barrier type (Mean BAB1 = 51.1; mean BAB2 = 77.4; Mean HLS = 68.2), F1(2,62) = 8.70, p = 0.001, \( \eta^2_p = 0.21 \) and also an effect of Sex (Female Mean = 76.2; Male Mean = 54.9), F1(1,31) = 5.16, p = 0.03, \( \eta^2_p = 0.14 \); there was no effect of Sex for gaze duration, F1(1,31) = 0.41, p = 0.72. For the younger participants, there was an effect of Barrier type (Mean BAB1 = 59.6; mean BAB2 = 72.6; Mean HLS = 75.0), F1(2,38) = 4.67, p = 0.02, \( \eta^2_p = 0.14 \) and no effect of Sex (Female Mean = 70.2, Male mean = 68.0), F1(1,19) = 0.02, p = 0.89. As predicted, there was an effect of communication condition on partner-directed gaze frequency (with more gazes in the barrier condition; and an effect within barrier type). For the older participants, there was an effect of the sex of the interlocutor, with older males making fewer partner-directed gazes than older females. This pattern has an interesting parallel in the auditory analysis, where it was found that unlike older female speakers, older adult males made no adjustments to their speech for the benefit of the other talker or in the HLS condition (Tuomainen & Hazan, 2016).
Figure 1: Average number of partner-directed gazes for the old and young participants as a function of communication condition (whiskers = SE).

References
In 2015 nearly 19 % of Europe’s population was over 65. This demographic change is one of the major challenges faced by the social, biological and health sciences. Ageing entails changes at several physiological levels, including the central nervous system, the musculoskeletal system, the skeletal system, the cardiovascular system and the respiratory system (Jacobs-Condit & Ortenzo, 1985). Crucially, increasing age affects motor control, involving a slowing down of movements and deficits in the coordination of these movements. Although ageing has been reported to lead to a general slowing down and a high degree of variability in speech production, very little is known about how it affects speech motor control.

The most striking effect of ageing on motor control in general is that movements are slowed down (Cooke et al. 1989; Seidler et al. 2002). This process of slowing down crucially affects the entire structure of the movements. Further, motor control in ageing individuals also entails a high amount of variability in limb coordination, owing to a decrease in accuracy, which in turn results in coordination deficits (Brown, 1996; Cooke et al., 1989). These age-related deficits play an important role in motor control, since “[c]oordination is a part of most tasks of daily living and therefore it is essential to understand breakdowns in control and regulation” (Ketcham & Stelmach, 2004:6). Under time pressure older individuals perform better on simultaneous movements (in-phase, both hands go up and down in synchrony) than on alternating movements (anti-phase, hands go up and down in opposite directions).

Coordinating the movement of the articulators (the oral, or supraglottal, system) to produce speech is referred to as speech motor control. The oral system with its articulators almost exclusively involves fine motor control with the millimetre precision and split-second timing needed to perform this highly complex task. As in motor control in general, a commonly reported effect of ageing on speech is that the tempo is slower (Amerman & Parnell, 1992; Ramig, 1983). This reduction in speech rate is often measured in terms of words, syllables or phonemes per second, but certain sounds and structures are more compressible than others, hence slowing down cannot take place homogeneously (Fletcher, 2010). Our knowledge of how ageing affects the oral system is limited by the fact that most studies are primarily based on acoustics, making an analysis of articulatory coordination patterns difficult to carry out. It is unlikely that age-related speech rate reduction compares to a deliberate rate reduction in younger individuals (e.g. when attempting to speak clearly), much like a slower walking tempo due to ageing is not the same as an intentionally slower walking tempo at a younger age (Kang & Dingwell, 2008). Mefferd & Corder, 2014 state that “relatively little is known about how ageing affects the speech motor system, and its potential contribution to a slowed rate of speech” (Mefferd & Corder, 2014:347). Only few is known about how ageing affects the coordination of articulators and whether or not coordination deficits appear (cf. motor control in general).
This study investigates ageing in speech motor control rather than estimating from the acoustic signal only. The project focuses on healthy ageing subjects rather than on subjects with speech and language disorders. This problem is addressed by analysing articulatory coordination patterns in ageing individuals using electromagnetic articulography (EMA, AG 501, cf. Fig. 1) in order to track the small and fast movements of the speech articulators. In this analysis on speech of ageing individuals, I will analyse gestural coordination patterns within the framework of Articulatory Phonology (Browman & Goldstein 2000).

We recorded 4 older speakers (aged 70-80 years) and 4 younger speakers (aged 20-30 years) of German with an Electromagnetic Articulograph (AG 501). We put sensors on upper and lower lip, chin, tongue tip, tongue blade and tongue body to track the movements of the articulators. Three additional sensors served as references for dynamic head movement corrections (on the bridge of the nose, behind the left/right ear). The speech material contained disyllabic words with simple onsets (e.g. /bina, pina, dina, tina/) which were repeated 7 times. These target words were embedded in a carrier sentence “Er hat wieder _ gesagt.” (‘He said _ again’). We identified articulatory landmarks including onset, peak velocity and target for consonantal and vocalic movements, identified at zero-crossings in the respective velocity and acceleration traces. We calculated temporal and spatial parameters within a mass-spring model, such as (i) the gestural activation interval, (ii) peak velocity, (iii) displacement and (iv) stiffness.

Preliminary results for 2 older speakers (72, 74 years) and two younger speakers (23, 29 years) indicate that ageing affects temporal as well as spatial parameters of the oral system.

Figure 1: Articulatory recordings with Electromagnetic Articulograph (AG 501).

Figure 2: Averaged values (left) and density plots (right) for peak velocity of consonantal gesture for young vs. old speakers.
Fig. 2 displays averaged values and the respective density plots (peak of the curves represents the mode of the distribution) for the peak velocity of the consonantal gesture before it reaches its target for the labial closure for /b, p/ or the alveolar closure for /d, t/. The results reveal first indications that older speakers are slower in articulating the consonantal gesture and thus, do differ from the younger ones (cf. Fig. 2).

The preliminary results reveal that both temporal and spatial parameters are affected by age. Thus, this direct observation of the articulators will provide insights into age-related effects on speech motor control.

References
Introducing two databases of spoken French throughout adulthood
Cécile Fougeron1, Véronique Delvaux2,
Cédric Gendrot1, Marina Laganaro3, Lucie Ménard4
(in alphabetical order except corresponding author)
1 LPP CNRS/Sorbonne Nouvelle, France; 2 FNRS & UMONS, Belgique; 3 Université de Genève, Suisse;
4 Université de Québec à Montréal, Canada
Corresponding author: cecile.fougeron@univ-paris3.fr

Our understanding of life-span changes in the speech of adults is quite sparse. This abstract describes the construction of two recent databases of spoken French documenting the evolution of speech throughout adulthood via longitudinal data on 10 speakers and cross-sectional data on 400 speakers. Both databases include a variety of linguistic material and speech conditions in order to assess multiple speech dimensions.

a) In the PATATRA (Parole AdulTe A TRavers les Ages) project, pursued at the Laboratoire de Phonétique et Phonologie in Paris, the objective is to document variation in the speech of a selected set of individuals over a 10-year period. Longitudinal recordings of 10 French speakers (4 male, 6 female speakers) are made every year in the same conditions (sound booth, recording material, sonometer calibration...). The speakers are all permanent members of the laboratory, aged from 30 years to 60 years old at the beginning of the project in 2013. To date, four to five time points are available per speaker. Audio and EGG signals are recorded for five speech tasks: the reading of a French version of the North Wind and the Sun, the reading of three repetitions of monosyllabic French words with /i,a,u/ in comparable consonantal contexts, five minutes of spontaneous conversation with the experimenter (a colleague), two glissando on a /a/ vowel (at a lowest and then highest intensity possible), three maximum phonation time performances on a sustained /a/ vowel. Recordings are completed by a self-assessment of the speaker’s voice quality on the day of recording, by a French version of the Voice Handicap Index questionnaire (Jacobson et al. 1997), and by information about potential smoking and drinking habits and ENT infections. Recordings are orthographically transcribed and phonetically aligned using ASR. Manual corrections of the alignments are done when necessary for the acoustic analyses. These analyses consist in a quantification of the changes over time in (speech and articulation) rates, pause time and F0 (computed over the text reading and the conversation); in vowel space (computed on the word list and the text); in open quotient; in tonal and intensity dynamic field (glissando); and in maximum phonation time.

b) The second database is a subcomponent of the MonPaGe project (Fougeron et al. 2016) which general goal is the development of a computerized speech screening protocol primarily designed for clinical practice. Directed for French-speaking patients, this protocol is in the process of validation and a set of reference values based on healthy speakers is under construction. For this purpose, a large number of speakers have been recorded in Paris, Geneva, Mons and Montreal in order to cover several French varieties. This reference database includes 400 French speaking adults, half male and half female, encompassing five age groups with 80 speakers per groups: [18-39], [40-49], [50-59], [60-75], [75+]. These cross-sectional data thus form one of the largest French database available for observing normal changes in the speech of adults as a function of age, and especially in older adults.

The material recorded with the MonPaGe speech protocol is primarily designed for the assessment of patients presenting signs of speech motor disorders, and it therefore covers multiple aspects of speech. Speech production is assessed through a variety of tasks: repetition of pseudo-words, reading of words and sentences; automated production of the days of the week; diadochokinetic
tasks (fast and precise repetitions of alternating speech movements); semi-spontaneous production in a picture description task. The speech material covers different speech dimensions: expressive and linguistic prosody, speech and articulation rate, voice quality on sustained /a/, maximum phonation time on a sustained /a/. The production of all French consonants and vowels is also tested according to their position in the word and the level of phonetico-phonological complexity of the utterance (structural pattern, length, planning difficulty). Recordings have been made in a quiet environment, at a 44100Hz sampling rate and with professional or near professional audio equipment depending on the recording site.

The analysis of this large database is currently in progress and first results will be presented at the workshop.

Acknowledgments
This work is partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (reference: ANR-10-LABX-0083). The authors thank all the participants of the MonPaGe project, collaborators and students, in the different sites.

References
Challenges for Automatic Speech Recognition over the Life Span
Martin Hagmüller
Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria,

In recent years, automatic speech recognition (ASR) has become more and more a commodity, as it permeates ever more situations in everyday life. Starting from early speech dialogue systems for telephone self-service applications, we have come to the situation that with a smart phone almost everybody has an ASR front-end in their pocket. For many people this works satisfactorily, e.g. for texting, and command and control. But when looking beyond the group of digital natives to the whole age range from children to very old people, ASR recognition performance drops considerably. Kwon (2016) reports that the word error rate (WER) increases from approx. 5% for young adults to 25% WER for elderly people. In recent years, ASR performance has significantly improved, resulting in WERs below 10% for challenging tasks (see e.g. the results of the recent Chime 4 challenge (Chime, 2016)). These results hold for the young speakers that are easy to recruit. Thus, for members of this age there exists plenty of training material for the ASR system, while others are neglected, because they are more difficult to get hold of.

The most recent efforts to record an Austrian German database with a wide demographic coverage were undertaken around the change of the century. When we compare the data from the largest Austrian speech corpus (Baum, 2000), that aimed at a well-balanced demographic distribution over age we still see large discrepancies between the representation between the corpus and the real age distribution in Austria (see Table 1). The group between 16 and 30 years comprises almost half of the corpus, whereas they not even constitute 20% of the population.

Table 1: Distribution of speakers over age groups and sexes for the speech corpus SpeechDat-AT compared to Austrian population statistics. The data is taken from the SpeechDat-At documentation and Statistics Austria, the Austrian governmental statistics organisation.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Speechdat</th>
<th>Population Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percent</td>
</tr>
<tr>
<td>0-15</td>
<td>33</td>
<td>1.7%</td>
</tr>
<tr>
<td>16-30</td>
<td>994</td>
<td>49.7%</td>
</tr>
<tr>
<td>31-45</td>
<td>590</td>
<td>29.5%</td>
</tr>
<tr>
<td>46-60</td>
<td>341</td>
<td>17.1%</td>
</tr>
<tr>
<td>61-99</td>
<td>42</td>
<td>2.1%</td>
</tr>
<tr>
<td>Total</td>
<td>2000</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Nevertheless, not only people from this age group will be using services that are based on ASR. As an increasing number of people will be able to maintain an autonomous life in old age, the need for smart approaches to support them is obvious. One of those tools is a voice controlled home automation system that helps in case of possible motor impairments. Other applications involve a care robot that uses the speech communication channel for interaction. On the other end of the age scale, but not limited to, are ASR supported learning system that need excellent recognition rates. Voice interactive personal assistants that started to get on the market now, need to improve on ASR to go beyond the currently limited interaction to a true natural speech conversation for all age groups.

The impact of the different times in a life on ASR can be grouped into three topics, (1) voice acoustics, (2) vocabulary and (3) specific use of language. Those relate in terms of ASR to the (1) acoustic model (2) lexicon and (3) language model, respectively.
(1) Voice acoustic is the most obvious difference between the age groups. Especially during childhood and adolescence the voice undergoes dramatic changes (Lee, S., Potamianos, A., & Narayanan, S., 1999). Among others, one major issue is the increasing vocal tract length when growing up, another one the high fundamental frequency.

There are also a few challenges that ASR has to cope with in the processing of elderly peoples’ speech. Several changes in voice and articulation degrade speech in old age (see Table 1 in Young, V., & Mihailidis, A. (2010)). In addition, older people are more likely to suffer of communication disorders that affect the recognition performance of ASR systems.

The acoustic model in an ASR system aims to capture the variability of voice production. For a successful ASR system this either means lots of annotated data to learn the model or to use a specific model for each age group. The former is the method of choice for deep neural network recognizers and is only possible for groups that have access to a lot of real life data, like e.g. Google. The later needs an age classification algorithm to use the correct model. The age classification can be done implicitly by applying different acoustic models and choosing one with the highest confidence.

(2) Vocabulary also changes over the lifespan as it increases with education and life experience, though the growth rate decreases with age (Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M., 2015). Therefore child speech could seem to be the more, easy. But the limited vocabulary is ‘compensated’ by word creations or wrong pronunciation (Gray, S. S., Willett, D., Lu, & J., Pinto, J., 2014).

During the middle age range vocabulary depends a lot on the profession and education.

(3) Language use also changes a lot during lifespan. For young people this is well documented in the literature on interaction in social media. The age of the author can be estimated from transcripts with a small amount of word tokens (Peersman, C., Daelemans, W., & Vaerenbergh, L. V., 2011). The issue of the language model for ASR is highly critical for adolescence, where models trained on adult speech might not reflect at all the use of language of a 16 year old. The language model would even be different for different youth sub-cultures.

ASR performance depends to a large extend on the amount of labelled data available for training. While creating large databases for younger people already involves a lot of effort, for older people this is even more difficult. E.g. it might not be possible to bring the subjects to a high quality recording studio; the use of technology is not as natural as it is for digital natives.

Research on changes in speech and language production during the lifespan is beneficial for research in ASR. But I also believe that ASR can be a tool for speech and language production and perception research by using this technology to test models and hypotheses.
References


Inhibition predicts lexical competition in older adults’ spoken word recognition
Sarah Colby1,2, Victoria Poulton1, & Meghan Clayards1,3
1 School of Communication Sciences & Disorders, McGill University, Canada; 2 Department of Linguistics, McGill University, Canada; 3 Corresponding author: sarah.colby@mail.mcgill.ca

Introduction. Spoken word recognition requires listeners to integrate top-down and bottom-up sources of information, while inhibiting irrelevant lexical competitors as speech unfolds. Using an eyetracking paradigm, McMurray, Munson, & Tomblin (2014) found that adolescents with Specific Language Impairment show a different pattern of looks to phonological competitors than typically developing adolescents: increased competition across both clear and ambiguous speech tokens. The authors suggest this pattern is reflective of differences in lexical competition (top-down information), rather than speech perception (bottom-up information), which would manifest as more competition from ambiguous tokens. As we age, uncertainty becomes more pronounced in speech perception; it becomes harder to recognize words in noise (Helfer & Freyman, 2014) and to inhibit similar sounding high-frequency lexical competitors (Revill & Spieler, 2012). Similar to McMurray, Munson, & Tomblin (2014), we are also interested in contrasting lexical competition and speech perception in older and younger adults, because older adults have been shown to have weaker encoding of some phonetic contrasts (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012; Bidelman, Villafuerte, Moreno, & Alain, 2014) and also show greater lexical effects than younger adults (Mattys & Scharenborg, 2014; Revill & Spieler, 2012). The increased lexical bias found in older adults (Mattys & Scharenborg, 2014) could be indicative of decreased lexical inhibition, which would result in increased activation among lexical competitors. Indeed, Sommers and Danielson (1999) found that older adults had greater difficulty than younger adults recognizing words with many semantic neighbours, suggesting that older adults had more difficulty inhibiting competitors. We are additionally interested in investigating the individual differences in cognitive factors that affect lexical competition in older and younger adults. Given that word recognition requires listeners to inhibit lexical competitors, we are specifically interested in the role that domain-general inhibitory abilities play in resolving lexical competition.

Current study. Using the visual world paradigm, we investigate the role of inhibition, lexical competition and phonetic sensitivity for word recognition in a group of younger and older adults. Given that older adults have shown weaker encoding of voice onset time (VOT), a temporal acoustic cue to voicing in English, compared to younger adults (Anderson et al., 2012), we manipulate VOT and include a phonological competitor to the target in our display to contrast the effects of lexical competition and phonetic sensitivity. If older adults show increased lexical competition, we expect them to fixate more on the competitor image regardless of the stimuli; however, if older adults display poorer phonetic encoding, we expect them to show increased looks to the competitor image as the stimuli become more ambiguous, indicating that they have more difficulty with uncertainty in the signal.

Method. A group of 23 older adults (ages 60-76, \(M_{\text{age}}=67.8\)) and 19 younger adults (ages 18-33, \(M_{\text{age}}=22.1\)) participated in a visual world paradigm eyetracking task, and completed the Simon task as a measure of inhibition. For the eyetracking task, participants were instructed to click on the image that best matched the word they heard over headphones. Stimuli were six 9-step continua of /b/-/p/minimal pairs (e.g., bear-pear, beach-peach, bees-peas, etc.). Each trial presented a target and a competitor image that corresponded to the minimal pair (note that which image was considered the target depended on the stimulus presented; e.g., a stimulus from the /p/-side of the bear-pear continuum means ‘pear’ was the target image and ‘bear’ was the competitor), as well as two distractor images. Each step of the continua was presented as the target 10 times, for a total of 540 test trials. An equal number of filler trials were presented where the target was one of the
phonologically unrelated distractor images, for a total of 1080 trials. Proportion looks to the target and competitor image were measured.

Additionally, participants completed the Simon task, which is a non-linguistic measure of interference from irrelevant information (Craft & Simon, 1970). In the version we used (Mueller & Piper, 2014), participants were asked to respond with one of the Shift keys on a keyboard, depending on the colour of the circle presented on screen. Crucially, the location of the circle alternates randomly over the course of the task between the left side, center, and right side of the screen. Thus, participants must inhibit responding with the key that matches the location of the circle (i.e., press the left Shift key when the circle is on the left side of the screen), and instead respond with the key that matches the colour of the circle (i.e., press the left shift key when the circle is red, press the right shift key when the circle is blue). The Simon task was scored as the difference in response time between incongruent trials (location of circle conflicts with side of response) and neutral trials (circle is located in the center of the screen). Note that this means participants with poorer inhibition will have higher Simon scores, while those with better inhibition will have lower scores.

**Results.** As in McMurray et al. (2014), participants’ category boundaries were calculated and continuum step was made relative to each participant’s individual categorization boundary (such that the 0 step fell at their category boundary). Proportion looks to the competitor (from 200-2000 ms after the onset of the target word) were computed for each trial, and logit transformed. Only trials where the correct image was ultimately selected were included in the analysis (i.e., the “p” image was clicked when the stimulus was from the /p/-side of the category boundary, and vice versa). Separate mixed-effects regressions were run for the /p/- and /b/-sides of the continuum on the proportion looks to the competitor image, with continuum step, Simon score, and age group as predictors. For both sides of the continuum, we find no evidence for an effect of age group (/p/: $\beta$= -0.31, $t(41)=-1.27$, $p>0.05$, /b/: $\beta$= -0.28, $t(41)=-1.02$, $p>0.05$), nor an interaction between age group and continuum step (/p/: $\beta=0.25$, $t(41)=0.73$, $p>0.05$, /b/: $\beta=-0.2$, $t(41)=-0.62$, $p>0.05$). Figure 1A shows the proportion looks to the competitor image broken down by continuum step and age group. This suggests that older adults do not show increased lexical competition compared to younger adults, nor do they find recognizing increasingly ambiguous stimuli more difficult. However, on the /p/-side of the continuum, we find a significant interaction between continuum step and Simon score ($\beta=1.03$, $t(41)=-2.19$, $p=0.03$), suggesting that participants with poorer inhibition look more to competitors than those with better inhibition, particularly for steps at or near the category boundary. In other words, competitors distract those with poor inhibitory skills as speech becomes increasingly ambiguous (see Figure 1B). This is qualified by a significant three-way interaction between age group, Simon score, and continuum step ($\beta=-2.75$, $t(41)=-2.69$, $p=0.01$), which suggests

**Figure 1:** Proportion looks to the competitor image by (A) relative continuum step and age, and by (B) Simon score and relative continuum step. Note that (B) is for the /p/-side of the continuum only.
that, as the stimuli increases in ambiguity, older adults show an increased effect of inhibition on looks to the competitor compared to younger adults. On the /b/-side of the data, we find a significant interaction between Simon score and age group ($\beta=-1.98$, $t(41)=-2.43$, $p=0.01$), providing further evidence that older adults with poorer inhibition are more distracted by competitors. In sum, we find no evidence that older adults have more difficulty recognizing words as they become increasingly ambiguous compared to younger adults; however, we do find that general inhibitory abilities predict distraction by a competitor, with poorer inhibition predicting more looks to competitor images. Additionally, those with poorer inhibition become increasingly distracted as stimuli become more ambiguous, especially older adults with poorer inhibition.

**Conclusion.** The current study provides evidence that both phonetic sensitivity and lexical competition remain similar in older and younger adults during spoken word recognition. We find that age alone is not detrimental to inhibiting lexical competitors, but general inhibitory skill along with older age is important for how well competitors can be ignored. Given that differences in aging often arise in regards to processing speed (e.g., Ben-David et al., 2011), further analysis including the time course of word recognition in older adults would be of interest to investigate whether any differences in lexical competition arise when timing is taken into account.

**References**


Characteristics of filled pauses across the lifespan
Viktória Horváth ¹, Viola Váradi², Judit Bóna²

¹Research Institute for Linguistics, Hungarian Academy of Sciences; ²Department of Phonetics, Eötvös Loránd University, Hungary

Spontaneous speech is rich in disfluency and one of the most frequent disfluencies is the filled pause. Filled pauses usually indicate speech planning or execution problems, word-finding problems, self-monitoring, and occur when the speaker does not know how to continue speaking. They have a communicative function of helping people manage turn-taking, and mark the speakers’ mental state (Clark 1994; Brennan-Williams 1995; Clark–Fox Tree 2002). There are several factors (the speaker’s age, gender, difficulty of topic domain, length of the utterance, relationship between speakers, etc.) which affect the frequency of filled pauses, one of the most important ones is speaker’s age (Shriberg 1996; Yaruss et al. 1999; Bortfeld et al. 2001).

In Hungarian, there are various forms of filled pauses with a wide range of durations, like ö, m, öh, öm, öhm (Horváth 2010). Analysis of filled pauses have been focused so far mainly on those occurring in young adults’ speech samples, somewhat less attention has been paid to them in typically developing children’s speech and in the elderly’s spontaneous speech production. In addition, there are some inconsistencies in the findings of previous studies.

The aim of the study is to analyse how the use of filled pauses changes across the lifespan, from elementary school age to oldest old age. We analyse the forms, durations, phonetic context, and function of the filled pauses. Our hypotheses are as follows: i) there will be difference in the proportion of the diverse forms of filled pauses between the age groups; ii) the duration of filled pauses will differ across the age groups, iii) children and elderly speakers will produce more silent pauses next to the filled pauses than young adults, and iv) some properties of the age-dependent speech planning processes could be demonstrated by filled pauses.

For this analysis, spontaneous speech recordings of 120 speakers were selected from two Hungarian speech databases: GABI (Hungarian Child Language and Speech Database and Information Repository; Bóna et al. 2014) and BEA (Hungarian Spoken Language Database; Gósy 2012). Speech samples of elementary school children (7-year-olds), adolescents (17-year-olds), young adults (20-25-year-olds), middle-aged adults (45-53-year-old), young-old adults (60-year-olds), and oldest-old (75+) speakers were selected. The method was the same in every case. The subjects were asked to speak on their families, hobbies, free time activities, etc. In every age group there were 20 speakers (10 women and 10 men). They were native Hungarian speakers with normal hearing and without any known mental or speech disorders. From the 30-45-minute-long recordings on average, we analysed 7-8 minutes from each speaker. The occurrence, duration and context-dependency of filled pauses were analysed. The measurements were carried out by Praat.

Preliminary results show that there are age differences in each analysed parameter. Although in every age groups the major part of the occurrences is schwa-like realization, this form is less dominant in children, adolescents than in the other age groups. The most various realizations occur in the two oldest groups. The duration of the schwa-realization depended on the speakers’ age [one-way ANOVA: F(5, 846) = 10.282 p = 0.001, Table 1.]. The older speakers produced shorter schwa-hesitations on average as the school children and the adolescents as well. Schwa-hesitations were realized with the longest durations on average in the adolescents’ narratives (Fig. 1.).
Table 1: The p-values in the Bonferroni post-hoc tests

<table>
<thead>
<tr>
<th>Age-group</th>
<th>7-year-olds</th>
<th>17-year-olds</th>
<th>20-year-olds</th>
<th>50-year-olds</th>
<th>60-year-olds</th>
<th>75-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-year-olds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-year-olds</td>
<td></td>
<td></td>
<td>0.001</td>
<td>0.008</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>20-year-olds</td>
<td></td>
<td></td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-year-olds</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-year-olds</td>
<td>0.001</td>
<td>0.004</td>
<td></td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75-year-olds</td>
<td>0.001</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: The duration of schwa-hesitations (ms)

The majority of children’s and adults’ filled pauses occur with silent pauses, but the ratio of filled pauses without silent pauses grows in the two oldest age groups.

Based on our findings, we can conclude that occurrences, forms, durations of filled pauses show some age-dependent changes in speech production process across the lifespan.

References


This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. K-120234 and the János Bolyai Research Scholarship.
Developments in accent perception between the pre-school and primary school years and the effects of exposure to variation

Ella Jeffries
University of Worcester, University of York, UK
e.jeffries@worc.ac.uk

Background
Perceptual categorisation studies have found that the accuracy with which adults can group speakers into accent/dialect categories varies according to how broadly defined the groupings are, as well as the age of the listeners themselves and their experience of variation relative to their residence in different regions of the country (e.g. Williams et al., 1999; Clopper and Pisoni, 2004). While differences between adolescents and teachers have been found (Williams et al., 1999), the age at which children can use regional accent features in order to group speakers, and how this ability develops, is not clearly understood.

Few studies have investigated children’s awareness of regional accents between infancy and adolescence and the findings are not consistent or conclusive. Floccia et al. (2009) and Wagner et al. (2014) found that children at the age of 5 and 6 were unable to group speakers according to a regional accent distinction. However, Beck (2014) and Evans et al. (2016) found that children performed above chance level in this ability. Furthermore, based on the results of a free classification task, Jones et al. (2017) found evidence for the incremental improvement of dialect categorisation abilities throughout childhood and adolescence, from the age of 4. These inconsistent results highlight the importance of methodological considerations, such as demands on the children’s working memory and the need for control over the linguistic variables used in the experimental stimuli. They also expose the need for further investigation into how this ability develops throughout young childhood and to consider the potential effects of children’s exposure to variation in their linguistic input. These issues are addressed in this paper which reports the results of perceptual experiments run with pre-school and primary school children from York.

Methodology
The present study tested children’s ability to group speakers by phonological variables indexing regional accents. Two separate experiments were carried out. In the first experiment, 20 pre-school children (aged 3-4 years) from York participated in an accent game which involved grouping together cartoon characters based on the way that they spoke. Children were presented with a set of visual stimuli consisting of two cartoon character mothers and five ‘lost babies’. Each character spoke a short sentence containing one vowel variable that distinguishes northern from southern British accents. These stimuli were recorded from the same speaker using different speaker guises. The stimuli were designed with three levels of difficulty, depending on whether the same word and/or phoneme was spoken. The children’s task was to identify which babies belonged to which mother, according to how they spoke. The second experiment was run with 34 primary school children (aged 6-9 years) from York and used a similar experimental design. In this experiment, however, rather than using speaker guises, the stimuli were recorded from different northern/southern British English speakers.

Results and interpretation
In the first experiment, pre-school children scored above chance (at 50%) at grouping speaker guises according to phonological variables indexing northern/southern accent differences. An improvement was found between the 3 and 4-year-olds for all difficulty levels (see Figure 1), supporting previous research demonstrating that ages 3-4 are ‘critical’ for language learning in general, including the learning of variation (e.g. Roberts & Labov, 1995).
In the second experiment, primary school children performed better than the pre-school children (average 77% correct answers) in a harder version of the same task, in which they grouped different speakers according to phonological northern/southern regional variables. An age-related improvement was found for the boys but not the girls in this experiment, indicating the girls had reached a ceiling level of performance by the age of 6.

The results of these two experiments indicate that children progress from the ability to use accent features as grouping criteria for (1) unfamiliar speaker guises to (2) different unfamiliar speakers’ voices. In the second experiment, the older children were able to overlook differences in the pitch and voice quality of individual speakers in order to group them by similar features of their accent.

Binary mixed effects logistic regression models were fit over the results of both experiments. These models revealed that overall, children with more exposure to non-Yorkshire (and therefore non-local) English speaking family and friends performed better (see Figure 2). This indicates that children’s individual exposure to variation in their linguistic input plays a role in their perceptual awareness of these phonological distinctions.

**Figure 1: Performance between the age groups across all difficulty levels in the first experiment**

![Performance between the age groups across all difficulty levels in the first experiment](image)

**Figure 2: Raw data (left) and model prediction (right) showing the effect of children’s exposure to other regional varieties in the second experiment**

![Figure 2](image)
The results from these two experiments highlight the role of age and exposure in children’s developing perceptual awareness of regional accent distinctions. It is proposed that the findings are best explained by interpreting them through an exemplar theoretic account. In such an account, speaker categories develop from the abstraction across social-indexical properties of phonetic variation which accumulate through an individual’s experience with variation in their linguistic input (Foulkes and Hay, 2015).

References
Production and perception of stress in ‘traditional’ and ‘new’ speakers of Breton

Holly Kennard
Faculty of Linguistics, Philology & Phonetics, University of Oxford, UK
holly.kennard@ling-phil.ox.ac.uk

Introduction
This paper examines stress in Breton from a cross-generational perspective. Breton is a Celtic language spoken in western Brittany, and has a somewhat unusual speaker community, consisting of two largely separate groups of speakers: older, ‘traditional’ speakers (usually aged 70+), and younger, ‘new’ speakers (usually aged under 40). This study investigates the impact that a gap in language transmission, and the emergence of new speakers, is having on the language itself, using fieldwork and experimental data to look in particular at how speakers use and perceive stress.

Breton is now considered to be an endangered language, with a large proportion of speakers aged over 75 (Broudic, 2009). Following decades of political and social pressure, the language saw a decline during the 20th century, and normal parent-to-child transmission ceased following the Second World War. Language revitalisation efforts led to the establishment of Breton-medium schooling in the late 1970s, and today there is a growing community of new speakers who had little or no exposure to Breton at home, but acquired the language via immersion schooling or other initiatives. The label ‘new speaker’ has been used in literature on minority languages to describe younger speakers who have acquired the language by means other than intergenerational transmission (see, e.g. Hornsby, 2015; Jaffe, 2015; O’Rourke, Pujolar, & Ramallo, 2015). The Breton spoken by new speakers is termed Neo-Breton and has been identified as differing from that of older, traditional speakers, primarily in terms of its lexicon (avoiding French loanwords in favour of more ‘Celtic’ equivalents (Jones, 1995)), but also syntax, morphology and phonology (Hornsby, 2005).

Breton stress is largely straightforward and uncontroversial to describe, and as such has not been the subject of detailed discussion in the linguistic literature. Stress falls on the penultimate syllable of the word, for example: kádor ‘chair’, baláfenn ‘butterfly’, fourchetézenn ‘fork’ (Desbordes, 1983). There are, unsurprisingly, a number of noted exceptions to this penultimate stress pattern, where the stress is on the final syllable. These include a small set of nouns (e.g. pemóc’h ‘pig’, amánn ‘butter’), certain verbal forms, some place names (e.g. Kastellín), and a number of adverbs and prepositions (including dirák ‘in front of’ and gwechall ‘in the past’). There is also some regional variation: the dialects spoken in the Vanneetais region tend to have final stress (Jackson, 1967). Stress has become relevant to discussions of new speaker varieties. All new speakers are native speakers of French, the dominant language of the community, and the stress pattern of French is quite different: the main stress tends to fall on the final syllable of a phrase group (Di Cristo, 1998). It has been claimed (e.g. Madeg, 2010) that new speakers transfer the final stress pattern of French into their Breton, further differentiating ‘Neo-Breton’ from the traditional Breton of older speakers.

Methodology
To investigate the potential differences in stress patterns between the two generations of speakers, fieldwork was conducted in the southwest of Brittany. Twenty Breton speakers were recruited, of whom ten were older, traditional speakers, and ten were younger, ‘new’ speakers. The older speakers were aged from 57 to 83 (mean age 73.4), and had grown up speaking Breton at home with their parents. They had not received any formal instruction in the language, and few were able to read or write Breton. The younger speakers were aged from 27 to 52 (mean age 38.3) and most had acquired Breton through schooling or similar.
Experiment 1: Production

Looking first at production, speakers were asked to perform an elicitation task where they were shown pictures of objects and named them in Breton. The goal was to elicit 30 disyllabic and 30 trisyllabic words for each speaker. This method avoided the use of translation from French, and therefore possible interference. It was also preferable to using a word-list, which would have been difficult for those older speakers who were not literate in Breton.

Analysis of the production data indicates that far from consistently using a French-influenced final-stress pattern in Breton, as might be expected, many younger speakers use stress in much the same way as the older, traditional speakers do. The picture is one of predominantly penultimate stress, with a small number of exceptions. This is particularly true in disyllabic words (figure 1); however, in trisyllabic words there is more variation: some younger speakers have a greater tendency to use final stress in trisyllables, and there is consequently a greater difference between the generations (figures 2 & 3). Additionally, while the use of word final stress by the older speakers tends to be limited to historical exceptions (such as pemoc'h ‘pig’) and loanwords (such as banan ‘bananas’), younger speakers use final stress across a range of native and non-native vocabulary.

Experiment 2: Perception

Following this, speakers were asked to complete a perception experiment, which was constructed in the same manner as Dupoux et al. (1997)’s seminal work on stress ‘deafness’ in French. Dupoux et al find that French speakers had more difficulty than Spanish speakers in distinguishing between segmentally identical words with differing stress patterns. The experiment consisted of an ABX task, where speakers heard two trisyllabic non-words of the form CVCVCV which differed only with regard to stress, e.g. bopélo and bopeló. They then heard one of the two words again, and had to identify
which this was by pressing one of two buttons, labelled “1” and “2”, on a response box. The expectation was the speakers’ competence in Breton might help them in distinguishing stress patterns, and so they would be more accurate in this task than monolingual French speakers. Additionally, given that younger speakers are generally French-dominant, they might find the task more difficult than older speakers. However, older speakers might find the nature of the task itself more difficult – most of those interviewed were not computer-literate, and the task was challenging.

The results indicate that all speakers are generally accurate at perceiving differences in stress patterns: error rates are lower than in the original experiment run by Dupoux et al (1997). The older speakers did seem to find the task more difficult, however, as their accuracy was lower than that of the younger generation (mean accuracy: senior adults = 81.7%; young adults = 97.2%). This was an expected finding: the older generation found the whole experience challenging, from wearing headphones and listening to nonsense words, to pressing buttons and reacting quickly. This makes it difficult to draw conclusions about whether the older generation are more proficient at distinguishing stress patterns than the younger generation. However, despite their difficulties with the experiment, the reaction times for the older speakers were overall only slightly slower than those of the younger generation.

Considering the generations separately brings out an interesting finding. Among the younger generation of speakers, there is a correlation between increased use of word-final stress in the production task, and lower accuracy in the perception task. It seems that those speakers who find distinguishing stress patterns more challenging tend to use French stress patterns more frequently when they are speaking Breton. There is no such correlation among the older speakers, highlighting the fact that older speakers’ use of final stress is limited to loanwords and established exceptions to penultimate stress, rather than being spread across all lexical items.

Conclusions
Overall, the results show that to claim that Neo-Breton has final stress is an oversimplification. Some new speakers of Breton use final stress more than would be expected, particularly in longer words, and in both native words and loans. This appears to correlate with greater difficulty in distinguishing stress patterns. More generally, it seems that Breton speakers do perceive differences in stress more easily than monolingual French speakers, as expected for a language with variable stress patterns.

References
Auditory-phonetic processing in word recognition across ages
Valéria Krepsz and Mária Gósy
Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
krepszvaleria@gmail.com

Introduction
During language acquisition children have to recognize and remember the sound patterns of words despite their different acoustic manifestations (influenced by talkers, speech rate, contexts, etc.). It is usually assumed that the speaker’s lexicon contains a representation of each word in an idealized form which is matched to heard speech (Swingley & Aslin 2000). Children’s representations of familiar words are reported to be phonetically well-specified already around the age of two (Walley 1993). Their successful word recognition seems to depend upon acquiring lexical and phonological representations with developing matching process that links spoken words to these representations. The intake of information through the auditory system requires an online integration of differing and potentially competing information presented to the two ears. Dichotic listening techniques have been used as a sensitive non-invasive procedure to assess language lateralization under clinical settings and among children with and without learning disabilities (e.g., Hugdahl 2011; Obrzut & Mahoney 2011). However, this technique seems to be a challenging possibility to check the development of auditory-phonetic processing of children (Meyers et al. 2002) since the auditory system is reported to project bilaterally up to the level of the nuclei of the lateral lemniscus (Hugdahl 1999). During a dichotic listening test the participant should listen to different words at the same time including integration of the different pathways to the brain and identification of the bulk of speech sounds forming different words. Auditory-phonetic processing could amend the factors of word frequency and lexical density that are claimed to affect significantly the accuracy and speed of spoken word recognition in children (Krull et al. 2010; Garlock et al. 2011).

The goal of the present study was to collect baseline developmental data on the auditory-phonetic processing of words in a dichotic listening task with the participation of Hungarian-speaking children between the ages of 3 and 10. We hypothesized that children would show (i) gradual increase across ages in the number of dichotically presented words they recalled correctly, and (ii) a more intensive increase in the number of correctly recalled words presented in their right ear than in their left ear.

Methodology
320 right-handed children aged between 3 and 10 years participated in the study. Children were divided into eight age groups; each group included 40 children (half of them were girls in each group). All of them had normal hearing in both ears (screened at 20 dB HL at octave frequencies from 0.25 to 8 kHz) at the time of testing, no known history of delayed onset of language acquisition, of speech and language difficulties (examined prior to enrollment into the study on language production and perception proficiency, as well as handedness using standardized test batteries), and were native monolingual speakers of Hungarian. The children in this study all had a similar socio-economic status and were recruited from various kindergartens and schools in a large city.

A dichotic listening task was used with 15 pairs of frequently occurring disyllabic Hungarian words (e.g., alma/sapka ‘apple/cap’, kesztyű/fésű ‘glove/comb’) that were read by a male voice (word frequency effects in word repetition tasks for preschoolers and elementary-school children were reported to be minimal, see Garlock et al. 2001). Recording of the words was processed according to general demands of dichotic listening test materials. The first part of the test consisted of five pairs of words with the insertion of a silent pause of 600 ms between the pairs. The second part consisted of five times two pairs on each trial.
The words were presented through earphones to both ears of each child at a volume allowing comfortable listening. The participants were asked to repeat what they heard after each trial (non-forced or free-report condition). Individual testing was performed by both of the authors. Participants’ answers were recorded directly on to a computer. Data were scored for each participant as the percentage of correctly recalled words for the right and left ear input. Left ear and right ear scores are the overall number of words correctly repeated in respective single ears. The both ears scores are the total number of items in which both words of the dichotic pair are correctly repeated.

The data were subjected to statistical analyses (GLMM method) using SPSS 19 software. The scores of correctly repeated words were the dependent variables while age, ear (left vs. right), number of errors, and gender were the independent factors. Significance was set at the 95% confidence level.

Results
The number of words repeated correctly in the dichotic listening task showed a significant increase across ages. Summarizing the correct scores of both ears it shows that there is a modest, gradual increase from the age of 3 (mean score: 45.5%) up to the age of 7 (mean score: 68.2%) followed by a steeper increase of about 11% between 7-year-olds and 8-year-olds. No difference was found in the correct scores of word repetition between the ages of 8 and 9 while there is again a modest increase by the age of 10 (mean score: 85.2%). The difference between correct scores both in the left and right ears was also significant (and the interaction of age and ear, as well). In terms of percentage correct, children recalled correctly 72.43% of all words heard in their right ear and 56.98% words heard in their left ear. As expected, more correctly recalled words were found heard in their right ear than in their left ear as the effect of the right ear advantage (Hugdahl 1999).

The mean scores of the words presented in the right ear did not show any significant change across the ages between 3 and 5 years as well as between 8 and 10 years. An increase was found between the ages of 5 and 8 years being steeper between 5-year-olds and 6-years-olds than between 6-year-olds and 8-year-olds (Fig. 1). The correct scores across ages show a different pattern for the words heard in the left ear. Here, the modest, gradual change can be spotted from the age of 3 up to the age of 7. The steeper increase occurs between the ages of 7 and 8 followed by another slight increase up to the age of 10 (Fig. 1).

Figure 1: Correct responses of dichotically presented words across ages

With increasing age, children were able to recall most successfully the second pairs of the four-pair stimuli. In addition, occurrences of erroneously recalled words significantly decreased. As expected, the more similar the segments were in their acoustic structures, the more difficulties children
encountered – with significant differences between kindergarten children and elementary school children. No significant differences were found depending on gender.

**Conclusions**

Older children accurately identified more words than did the younger ones confirming the development of the abilities necessary for word recognition under specific conditions. The number of words correctly recalled by the children in our study was higher than that of CV syllables reported in the literature but the tendencies were similar (e.g., Heiervang et al. 2000; Obrzut & Mahoney 2011; Moncrieff 2011).

Our results make it possible to conclude that children’s auditory-phonetic skills in word recognition seem to show gradual development in terms of the scores of dichotically presented words. Development was similar in both the right and the left ear, considered separately; however, it was heavily influenced by ear preference resulting in a larger increase for the right ear than for the left ear. The findings of right ear preference across ages are in line with those reported in the literature (e.g., Moncrieff 2011). In addition, possible influence of attention and verbal workload could explain the age-specific results including the individual differences in all age groups. Qualitative analysis of the correctly recalled words shed light on the use of identification and differentiation of speech sounds of the words in a condition where both processes are forced to apply at the same time. The successfully recalled words of the second pairs in the four-pair stimuli as a factor of age is assumed to be the consequence of the children’s developing word memory and better integration of the acoustic-phonetic patterns of the words.

The dichotic listening method seems to be a good way to detect the auditory-phonetic abilities of typically developing children including their interactions with memory and attention across ages. Age-specific norms could be used in a better understanding of atypical language development under both educational and clinical circumstances.

**References**


Perception and Processing of Context-dependent Anglicisms across Age Groups*
Dr. Anabella-Gloria Niculescu-Gorpin†, Dr. Monica Vasileanu‡
†,‡Romanian Academy, The Iorgu Iordan – Al. Rossetti Institute of Linguistics, Romania,
anabellaniculescu@hotmail.com

Introduction
Due to the pervasive influence of English, an unprecedented process of language change triggered by direct or mediated language contact can be observed at all language levels of present-day Romanian: lexical, morpho-syntactic, and even orthographic. The main tendency within most of Romanian academia has been to incriminate this phenomenon, in an attempt to stop it or at least slow it down somehow. Thus, the Romanian literature on the subject has described at least some lexical Anglicisms as ‘luxury’ words (Stoichiţoiu, 2006) i.e. borrowings that are not needed since Romanian already has long-established words with similar meanings, which makes the new loans superfluous. Such recommendations, however, have not been observed by the most important ‘language policymakers’, i.e. Romanian native speakers. Our research has focused on the study of this phenomenon, with an emphasis on the perception and processing of Anglicisms by native speakers of Romanian.

The main hypothesis
By means of questionnaires applied to over 150 Romanians living in the country and abroad, Niculescu-Gorpin (2013) observed how native speakers perceived and commented upon current language change phenomena (respondents were asked to mark sentences as correct/possible/incorrect and to specify what they found odd about the items under consideration) and concluded that the level most permeable to Anglicisation was the lexical level.

Based on these findings, and choosing the most frequent lexical Anglicisms coined by Romanian linguists as ‘luxury’, in this paper we have decided to test a further hypothesis that could, if correct, shed light on why Romanian native speakers will use such loan words even though they are advised not to: it seems that the use of at least some ‘luxury’ Anglicisms is context-dependent, i.e. they are in fact needed for appropriate communication as speakers feel that, by using them, they enrich the cognitive effects on their audience and bring about more contextual information.

The experiment: design and methodology
To test our hypothesis, we have designed an E-prime-based experiment containing 15 ‘luxury’ lexical Anglicisms. In choosing the lexical stimuli, we took into account several aspects: (i) most Anglicisms selected are fairly high in frequency and refer to everyday activities and realities, as we wanted our subjects to be familiar with them; (ii) there are at least one or even several, long-established Romanian words that Romanian linguists recommend instead of the loan words, and they are naturally highly familiar to native speakers; these were the words we included in the second, control version of our questionnaire, that was applied to a control group, to allow comparative analysis; (iii) we have chosen stimuli that could be easily associated with images; (iv) we have decided to limit ourselves to only 15 lexical stimuli because a longer list would have made the purpose of the experiment too obvious for the subjects (even in this format, some subjects did ask at the end if we were testing something related to the English influence on present-day Romanian, but could not tell exactly what).

We have built sentences in which we tried to avoid personal markers (such as personal pronouns, possessives, etc.) that would have made the subjects try to relate the content to their everyday lives and personal experiences. Since Romanian is a pro-drop language, we decided to either use a verb in the third person singular/plural or a proper noun for the subject if appropriate.
The experiment runs as shown in Figure 1 below: subjects are shown a fixation mark for 500 ms, then they can read the sentence at their own pace; when they are done, they press the space bar and an inter-stimulus interval black screen appears for 1000 ms; then, they are shown two pictures side by side for 5000 ms – they are required to press the “z” key if the sentence they have just read matches better the image on the left and the “/” key if it matches the image on the right.

Our main hypothesis is that Romanian native speakers will use a particular lexical Anglicism in particular context, say Ro. job, and not the long-established Romanian words or phrases (Ro. serviciu, loc de muncă or slujbă) because they feel that the Anglicism brings into communication some semantic and pragmatic information that the Romanian words would not; for instance, Ro. job will not apply indiscriminately to any type of job, but to a white-collar, office or teaching job. So, in choosing our pictures, we tried to do our best to match these semantic and pragmatic characteristics with appropriate images: one of them would fit all possible contexts in which Romanians would use the long-established word(s) or phrase(s), i.e. what we can call the unmarked context, and the other would best fit the more context-specific, marked case, specific for Anglicisms (see Figure 2). Since the control group got the same two images with the sentence containing the long-established Romanian word, both pictures had to trigger its use. We made sure there was a counterbalancing in the distribution of the pictures (left or right) in both versions of the experiment. Nevertheless, as shown in our interpretation, sometimes it is very difficult to draw a line between the two words and the reasons are manifold.

Subjects were asked to fill in a post-hoc questionnaire containing the following demographics: gender, age, education, knowledge of English and the contexts in which they use English. This analysis discusses the results of our experiment taking into consideration the age variable.
**The analysis. Some considerations**

We have considered the following age groups: (a) 15-18; (b) 19-23; (c) 23-60; (d) 60+; these age groups roughly correspond to categories of occupational status (high-school students, university students, working-age people, and seniors).

We studied each stimulus in two ways: (i) first we analysed the responses of the Anglicism group vs. the control group for each age group taken separately; (ii) then we studied the choices made across all age groups by the Anglicism group and performed a similar analysis for the control group; in all cases, we considered both the images chosen and response times.

Our preliminary results confirm our initial assumption that at least some lexical Anglicisms are context-dependent and are used to enrich communication. For example, for Ro. *shopping*, most respondents chose the image that we expected them to, i.e. one showing a woman shopping for clothes, shoes, etc., and not the image fitting the general context of shopping, for which Romanian uses *cumpărături*. We have also found that the control group did choose the unmarked picture consistently (see Figure 3), which further supports our theory.

*Figure 3: Sample from the experiment - the case of Ro. shopping*

Moreover, our analysis suggests that this preference is higher among young people (especially the 15-18 and 19-23 age groups), with a slight decrease in the third age group, and a more significant one in the fourth. There are other cases like Ro. *job* that, although they are mainly used in a particular context (see the discussion above), have also started to be used in the unmarked context. In such cases there are other factors at play: the age of the Anglicism itself (Ro. *job* has been around for over 40 years, as it had already been recorded in some dictionaries before 1989), the age of the respondents (younger people tend to visualise the concept of *job* as related to their own expectations, i.e. a well-paid position in a company), the subjects’ knowledge of English, etc.

In conclusion, our findings suggest that the native speakers’ perception of lexical Anglicisms tends to change with age, i.e. younger people tend to be more aware of the current language change phenomena triggered by an extensive direct or mediated contact with English, whereas older generations are more resistant to change, on the one hand, and also lack the background knowledge required to adopt such uses, on the other. It thus seems that age and resistance to language change go hand in hand, which is a reflection in language of a more general behavioural pattern.

*This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2480.*

**References**


**Style-shifting and its acoustic correlates in a cohort of 11 year old girls from East London**
Rosie Oxbury & Esther de Leeuw
Department of Linguistics, Queen Mary University of London, UK.
Corresponding author: r.f.oxbury@qmul.ac.uk

It is known that adults are capable of shifting their speech styles according to the context in which they are speaking, and style-shifting receives attention not just in sociolinguistics but also in related fields such as L2 acquisition and phonetics (e.g. Henriksen 2013; Rampton 2013; Sharma 2011). However, research on children’s style-shifting capabilities is thin on the ground (e.g. Reid 1978; Eckert 2008, 2011; Snell 2010). In particular the age range of children immediately preceding adolescence is rarely investigated, yet the variation seen in adolescent speech has been argued to first begin developing among “pre-adolescents” (Eckert 1996, 2011).

This study investigates how a group of 11-year-old girls in East London style-shifted with respect to changes in the English diphthong system associated with Multicultural London English (MLE). An acoustic analysis shed light on the socially complex ways in which these children participate in MLE sound changes. Again, as such, this study presents new data on intrapersonal variation in speech production during a unique and under-studied phase in language development.

MLE is a new dialect emerging in multilingual areas of inner London (Cheshire, Kerswill, Fox & Torgersen 2011). A previous apparent-time, i.e. cross-sectional, analysis of MLE has shown age-related differences in the phonology of this variety. MLE phonology is unusual in showing no adolescent peak, yet other ethnicity- and gender-based differences become apparent between childhood and adolescence (Cheshire et al. 2011). Among 8 year olds, no differences correlating with ethnicity or gender were evidence for the diphthong variables. However, in the 12-13 age bracket, Anglo girls showed “more conservative”, i.e. more Cockney-like, realisations of FACE and PRICE in comparison to their non-Anglo female peers (Cheshire et al. 2011, p.170). There appears to be “increasing gender differentiation with age” (Cheshire et al. 2011, p.170-172), as Anglo women aged 20-35 showed even greater differences from both Non-Anglo women and men of all ethnic groups of the same age in this cross-sectional study.

The most salient feature of MLE is said to be the changes in the diphthongs (Cheshire et al. 2011). on the one hand, there is reversal of Diphthong Shift (Kerswill, Torgersen & Fox 2008), meaning that the first elements of diphthongs in MLE differ from Cockney English and are in some cases closer to Received Pronunciation values for these vowels. At the same time, the diphthongs are becoming more monophthongal, meaning that the change in vowel quality from the first to the second element is less than would be expected in other British dialects. For the present study, we therefore carried out an acoustic analysis of three diphthongs in the speech of five 11 year old girls at school in Hackney. The location was chosen for the sake of comparability with Cheshire et al.’s (2011) results.

To examine whether these children style-shift with respect to MLE phonological innovations, the participants were recorded in situations designed to represent the careful-casual continuum of speech styles (Labov 1972), with the expectation that shifted and monophthongised diphthongs would be more likely to be produced the more casual the situation was. These situations were: reading a wordlist (formal speech); in an interview with one friend and the researcher present (less formal); and during school breaktime (casual speech) (cf. Reid 1978). Four girls classed as “Non-Anglo” (Cheshire et al. 2011) were recorded, and one “Anglo” girl.

The diphthong variables chosen for analysis were those in FACE, PRICE AND GOAT. In Received Pronunciation, these diphthongs are transcribed to be respectively /eɪ/, /aɪ/ and /əʊ/ (Roach 2004) and in Cockney they have been transcribed as respectively /ʌɪ/, /ɒɪ/ and /ʌʊ/ (Wells 1982, p.308). In
MLE, both FACE and PRICE have onsets similar in quality to RP, but are more monophthongal: /eɪ, ɛɪ/ and /æ, aɪ/ respectively (Fox 2015). GOAT in MLE has been described as having a back, close onset, and is also monophthongal (Kerswill et al. 2008), which we tentatively transcribe as /ɤ/ or /o/. We were therefore interested in the F1 and F2 frequencies of the first and second elements of the diphthongs in order to see whether these changed according to style of speech. We predicted that in a less formal setting, more MLE-like pronunciations would occur. To this end the first and second formant frequencies of these variables were measured at 20% and 80% duration points in the diphthong (Di Paolo, Yaeger-Dror & Wassink 2011). The F1/F2 frequency at onset was taken to be that formant’s frequency at the 20% point, and the trajectory of the diphthong was calculated on the basis of the difference in F1 from the 20% to the 80% point and the equivalent difference in F2.

Therefore we predicted that: FACE in playground speech would have a slightly lower F1 at onset (representing a more close pronunciation); PRICE in playground speech should have a slightly higher F2 at onset (more front); and GOAT in playground speech would have a slightly lower F2 at onset (more back). Furthermore, all three were also predicted to be more monophthongal in playground speech, i.e. for the trajectory to be flatter, which was also considered to be the most informative variable.

The findings showed that the children did adapt their speech according to the situation at hand, though not exactly according to a careful-casual stratification of speech styles (Labov 1972). For PRICE, the effect of situation was as predicted, in that the F2 frequency at onset was highest in playground speech for all participants (Table 1), indicating a more forward pronunciation, characteristic of both RP and MLE. Furthermore, the trajectory was shortest in playground speech, suggesting a pronunciation similar to /a/. For FACE and GOAT, style-shifting had an effect, but not entirely as predicted. For most participants the F1 frequency at onset of FACE was highest in playground speech (Table 1), contrary to what was predicted, and indicating a more open pronunciation. This could have been because of the higher intensity of speech in the playground setting (Huber et al. 1999). The trajectory of FACE was shortest in interview speech, and varied within each situation. For GOAT, the children tended to show a higher F2 frequency at onset in playground speech (Table 1), again contrary to what was predicted, and indicating a more forward pronunciation, potentially suggesting GOAT-fronting, a sound change not predicted for the current dataset because it was not previously found among MLE speakers (Cheshire et al. 2011). The trajectory was shortest in interview speech but, as with FACE, showed variability within each situation, as indicated by the IQR (Table 2).

**Table 1: First formant frequencies across situations for the first element of each diphthong**

<table>
<thead>
<tr>
<th>Diphthong first element</th>
<th>FACE</th>
<th>PRICE</th>
<th>GOAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situation</strong></td>
<td>F1 median</td>
<td>F1 IQR</td>
<td>F2 median</td>
</tr>
<tr>
<td>Wordlist</td>
<td>633.3</td>
<td>95.8</td>
<td>1567.0</td>
</tr>
<tr>
<td>Interview</td>
<td>581.0</td>
<td>119.8</td>
<td>1635.5</td>
</tr>
<tr>
<td>Playground</td>
<td>763.5</td>
<td>159.0</td>
<td>1694.6</td>
</tr>
</tbody>
</table>

**Table 2: Length of trajectory across situations for each diphthong**

<table>
<thead>
<tr>
<th>Trajectory length</th>
<th>FACE</th>
<th>PRICE</th>
<th>GOAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situation</strong></td>
<td>Median</td>
<td>IQR</td>
<td>Median</td>
</tr>
<tr>
<td>Wordlist</td>
<td>357.4</td>
<td>285.5</td>
<td>613.7</td>
</tr>
<tr>
<td>Interview</td>
<td>275.1</td>
<td>436.7</td>
<td>379.1</td>
</tr>
<tr>
<td>Playground</td>
<td>336.0</td>
<td>361.4</td>
<td>351.0</td>
</tr>
</tbody>
</table>
These findings reveal intersituational differences in children’s speech production, suggesting that these children adapt their speech to the communicative context. From these results, we have support for previous researchers’ assertions that children are just as stylistically adept as adults (Reid 1978; Eckert 2008, 2011). They also show children’s complicated ongoing acquisition and negotiation of language norms within their community, as suggested by the variance in the data.

References


Native and non-native perception of lexical pitch patterns across the lifespan
Stefanie Ramachers¹, Susanne Brouwer², and Paula Fikkert²
¹Department of German language and culture, Radboud University, Nijmegen,
²Department of Dutch language and culture, Radboud University, Nijmegen
s.ramachers@let.ru.nl

In Limburgian dialects, spoken in the southern province of Limburg in the Netherlands, word-level
pitch patterns can signal lexical differences. For example, haas [haːs] with falling pitch (accent 1)
means ‘hare’, whereas haas with falling-rising pitch (accent 2) means ‘glove’. Limburgian is assumed
to have lexical tone, but it has a lower functional load than tone in for example Mandarin Chinese.
There are only few minimal pairs and there is only a two-way contrast. Nevertheless, any primary
stressed bimoraic syllable carries either accent 1 or accent 2 (Gussenhoven, 2000). Pitch information
is assumed to be connected to a Limburgian speaker’s knowledge of how a word should be
pronounced. However, research on the influence of lexical tone on speech perception in Limburgian
is lacking, which highlights the importance of experimental research on this group of dialects. We
compare performance of native Limburgians to a native Dutch control group on a speech perception
task.

Speakers of Dutch have to learn that pitch is used at the utterance-level to distinguish between
sentence types (e.g., questions and statements) or to signal emotions (i.e., paralinguistic meaning).
Pitch also functions as one of the cues to word stress. Importantly, Dutch is a non-tonal language.
Pitch is not used to contrast word meaning – except for a few rare minimal stress pairs. Moreover,
pitch is only one of several correlated cues to word stress. The fact that pitch is (hardly or) not used
to signal lexical distinctiveness in Dutch might prevent speakers of this language to distinguish words
that differ in pitch only (Schaefer & Darcy, 2014).

However, previous research has shown that non-tone language listeners often show persistent
sensitivity to non-native lexical tones throughout the lifespan in purely perceptual tasks (e.g., Hallé,
Chang, & Best, 2004; Liu & Kager, 2014; So & Best, 2010). Several theories have been proposed
recently to account for these findings. These theories mainly consider the influence of the native
prosodic system (e.g., Schaefer & Darcy, 2014; So & Best, 2010). A pitch movement that signals a
lexical distinction in one language can signal a post-linguistic or paralinguistic distinction in another
language. The challenge for language learners is thus not about ignoring pitch variation, but
assigning it the correct linguistic function. Non-tone language speakers may still attend to pitch,
though perhaps less closely than tone-language speakers.

In the present study, we first compared 6-to-12-month-old Limburgian infants (N=39) to Dutch
infants (N=83) to see whether we could attest language-related differences in the developmental
perception of the Limburgian tone contrast. We tested their discrimination of the contrast using the
hybrid visual habituation procedure (Houston, Horn, Qi, Ting, & Gao, 2007). Infants were habituated
on repetitions of 4 tokens of the pseudoword taag [taːç], carrying either accent 1 or accent 2. Once
the infants were habituated, they were presented 8 trials featuring alternations between 4 old
tokens and 1 new token with the habituated accent (same trials) and 4 trials with alternations
between the 4 old tokens of the habituated accent and 1 new token with the other accent (switch
trials), in pseudo-random order. Successful discrimination is measured by a significant difference in
looking time between same and switch trials.

The results of a mixed ANOVA with Trial Type (same vs. switch) as the within-subjects variable and
Language (Limburgian vs. Dutch) and Age (6 vs. 9 vs. 12 months) as between-subjects variables
yielded a main effect of Trial Type, indicating that both Limburgian and Dutch infants discriminated
the Limburgian tones throughout their first year of life ($F(1,116) = 15.90, p < .001$, partial $\eta^2 = 0.12$,
observed power = 0.977; see Figure 1).
The success of the Dutch infants could be due to several factors, among other things to the degree of similarity to the native prosodic system (Ramachers, Brouwer, & Fikkert, under review). The Dutch infants might have perceived the Limburgian accents as native intonation patterns. To investigate this, we tested adult speakers of Limburgian (N=20) and Dutch (N=19). Testing adults also aids our understanding of whether the infants’ behavior reflects the adult prosodic system.

We expected that Dutch listeners’ sensitivity to the Limburgian accents would be affected by whether they could perceive them as intonation or not. In an AXB categorical discrimination paradigm, we manipulated the position (final vs. non-final position) of the Limburgian accents in our pseudo-word stimuli (cf. Braun & Johnson, 2011). In utterance-final position, rises and falls in Dutch signal questions and statements, respectively. In non-final position, however, pitch patterns do not signal linguistically meaningful information in Dutch. If the Dutch listeners’ performance is driven by assimilation to native prosody, we expected Dutch listeners to discriminate the accents better in trials with monosyllabic pseudo-words, featuring the accents in final position (e.g., taag₁a - taag₁b; comparable to the stimuli in the infant experiment) than in trials featuring disyllabic pseudo-words with the accents in non-final position (e.g., kee₂a - kee₂b). Next to these trials with between-category pitch variation, we also included trials with within-category pitch variation (i.e., where listeners had to discriminate different tokens with the same accent), expecting that Limburgian listeners would perform worse in within-category than between-category discrimination due to influence of their native lexical tone categories.

Mixed ANOVAs revealed differences in performance between Dutch and Limburgian adults. In line with our expectations, overall, Limburgian participants significantly outperformed Dutch participants in between-category variation trials (p < .05). However, this main effect of Language has to be interpreted in light of a significant interaction (F(1,34) = 8.52, p = .006, η²p = .20). Contrary to our expectations, both Limburgian and Dutch participants performed significantly better in disyllables than monosyllables (see Figure 2a). For the Limburgians, this was only significant if they heard an accent 1 word after an accent 2 word, but not vice versa. With respect to the between-category versus within-category variation analysis, we found a significant Language by Trial Type interaction (F(1,37) = 17.74, p < .001, η²p = .30). Dutch participants performed significantly better in within-category trials than in between-category trials (t(17) = -4.65, p < .001, Cohen’s d = -1.10), whereas Limburgians performed equally well in between-category and within-category trials (t(18) = .49, p > .05, Cohen’s d = 0.11; see Figure 2b).
Our results add to the existing body of work on cross-language speech perception. We are the first to have shown that, with respect to pitch processing, adult speakers of Limburgian dialects attend more closely to pitch than speakers of Dutch. This can be attributed to the lexical distinctiveness of pitch in Limburgian. The fact that Dutch listeners still showed sensitivity to different kinds of pitch variations in a purely perceptual task is probably due to their experience with pitch in their native prosodic system.

References
Regional variation in young adult and older speakers of Quebec French: A pilot acoustic study
Josiane Riverin-Coutlée¹ ² & Vincent Arnaud²
¹Département de langues, linguistique et traduction, Université Laval, Canada; ²Département des arts et lettres, Université du Québec à Chicoutimi, Canada
Corresponding author: josiane.riverin-coutlee.1@ulaval.ca

Introduction & Background
Since the 1960s, various authors have described the phonetic differences that characterise the French spoken in Quebec (QF) versus that spoken in the non-meridional, northern part of France (e.g., Gendron, 1966). Salient features of QF such as diphthongisation or affrication of /t d/ before /i y j u/ have been extensively detailed, with little attention paid to regional variation within QF. Some of the few studies that have addressed this subject (e.g. Morin, 1996; Friesner, 2010) have used ALEC (Atlas linguistique de l’Est du Canada, Dulong & Bergeron, 1980), a vast linguistic survey that took place in the 1970s across the province of Quebec and which aimed at collecting the speech of NORMs (Nonmobile, Older, Rural Males; Chambers & Trudgill, 1998). As Reighard (1982) points out, however, using ALEC to investigate phonetic variation has major methodological limitations, notably the fact that it is constituted of transcriptions made in situ by a single investigator mostly focussing on lexical variation.

That said, the reason why NORMs are sometimes preferred over younger speakers in studies of regional variation in QF is that a levelling phenomenon has been assumed to be taking place among the latter, younger group (Dolbec & Ouellon, 1999). Elderly speakers are thus seen as the last keepers of regional variation because their linguistic habits are thought to be relatively unchanged since childhood and because they are unaffected by community changes. This view has been challenged by a growing body of literature revealing that adults do change their speech production over their lifespan. For instance, using the 1971 Sankoff-Cedergren corpus and its 1984 trend and panel follow-up (Sankoff & Cedergren, 1971; Thibault & Vincent, 1990), Sankoff and Blondeau (2007) showed that adult speakers of QF who were followed longitudinally had changed their pronunciation in accordance with a community change, even if in smaller proportions than what was observed across generations. Another reason why regional variation is sometimes perceived as best represented by older speakers is because of the “age-grading” principle (cf. Labov, 1994). Once out of the labour market, older speakers tend to abandon standard norms and return to less standard habits. In contrast, young adults entering the labour market tend to adopt more standard norms and avoid nonstandard variants. Whichever the explanation, young adults and older speakers from the same community are not expected to share the same linguistic habits, and this is true for regional phonetic variants.

Questions & Hypotheses
In a recent study aimed at updating our knowledge of modern-day phonetic variation within QF (Riverin-Coutlée & Arnaud, 2014, in prep.), we found a substantial difference between young adult speakers from two cities in eastern Quebec, Saguenay and Quebec City, in production of the (ɛ#) variable, that is, the /ɛ/ vowel in open syllables in word-final position, e.g., [le] lait (milk) or [epɛ] épais (thick). Ten male and 10 female speakers from each city (n=40), all university students aged 20 to 29 years, took part in a reading task featuring 12 /ɛ/-final words. As shown in Fig. 1, a lowering phenomenon was observed among Saguenay speakers; they produced vowels with a much higher F₁ and lower F₂ than Quebec City speakers. A discriminant analysis revealed that 84% of the tokens could be automatically classified according to the speaker’s city of origin based on these F₁ and F₂ values. These findings were in contrast to previous literature that had reported that (ɛ#) lowering was more commonly produced in informal situations by speakers of lower socioeconomic background throughout the province of Quebec, except in the region where Saguenay is located.
(Gendron, 1966; Deshaies-Lafontaine, 1974; Paradis, 1985; Larochele, 1989; Paradis & Dolbec, 1998). The dramatic discrepancy between this description and our results led us to wonder whether a new phenomenon had emerged.

In this paper, we address this question by comparing the results obtained with young adults to the speech of older speakers, using the apparent-time construct (Labov, 1994). We predict that if (ɛ#) lowering is a new phenomenon, older speakers will not produce it with the same magnitude as young adults, if at all. If no difference is found between the two age groups, then variation is probably stable. We chose to replicate our protocol and avoid already existing recordings in order to circumvent the problems involved in a multiple-source comparison, i.e., highly diverging aims and methods, in addition to some older recordings being of insufficient quality for acoustical analysis.

Methods
In this pilot study, 8 speakers were recorded. They were aged 55 to 68 years and had been retired from work 3 months to 10 years at the time of recording, except for one female speaker who had been sporadically working from home for the last 24 years. Half were born and raised in Saguenay and the other half in Quebec City, with an equal number of men and women in each city. Participants were recorded at home and were instructed to read aloud sentences shown on a laptop screen, after which the experimenter pretended she misunderstood the last word of the sentence. The speaker would then repeat it once, as in the following example:

Speaker: Paul a un beau gilet. Paul has a nice shirt.
Experimenter: Un beau? A nice?
Speaker: Gilet. Shirt.

Only the repeated, isolated target words were analyzed. Each speaker produced 12 different words ending with /ɛ/, for a total of 96 tokens. These were manually segmented using Praat (Boersma & Weenink, 2017). Formant detection settings were optimally adjusted for every token, and then $F_1$ and $F_2$ were measured at 50% duration. Data were not normalized.

Results & Discussion

As displayed in Fig. 2, Saguenay speakers produced tokens with a higher $F_1$ and lower $F_2$ than Quebec City speakers. Comparing Fig. 1 (young adults) with Fig. 2 (older speakers) reveals that young adults and older speakers had similar patterns of production. An exploratory linear discriminant analysis, a supervised classification procedure that allocates tokens to categories (here, Saguenay and Quebec City) based on predictors ($F_1$, $F_2$), indicated that 92% of tokens were automatically classified correctly, highlighting that Saguenay and Quebec City productions are acoustically distinct.

Overall, although there are some limitations (e.g., small number of older speakers, participants not screened for socioeconomic background), a marked regional difference was found among older and young adult speakers, despite the levelling phenomenon that has been assumed to be taking place in QF. In addition, (ɛ#) lowering does not seem subject to age grading, which we interpret as evidence of stable regional variation. One possibility is that our 55 to 68 year old participants were too young to exhibit the patterns described in earlier literature. However, an alternative explanation for the differences between our results and previous studies might be that (ɛ#) lowering, far from being relatively new, is declining in modern-day QF (Morin, 2002). This hypothesis could be tested in a further study investigating younger strata of the population. In conclusion, considering more than one age group in research about regional variation in QF thus allows for a better understanding of the current state and dynamics of phonetic variation and change.

References


— (in prep.). Regional variation in Quebec French: On contemporary (ɛ#) lowering.


---

**Figure 1:** Tokens produced by young male (M) and female (F) speakers from Saguenay (SG) and Quebec City (QC)

**Figure 2:** Tokens produced by older male (M) and female (F) speakers from Saguenay (SG) and Quebec City (QC)
Are ejectives on the increase in English? Evidence from a single speaker

Adrian P. Simpson
Institute of German Linguistics, University of Jena, Germany
adrian.simpson@uni-jena.de

Recent studies have begun to systematically document the occurrence of ejectives in different varieties of English (Ogden 2009; Gordeeva & Scobbie 2013; McCarthy & Stuart-Smith 2013; Simpson 2014). This growing body of research could also be taken as an indication that ejectives are becoming more prevalent in English (Simpson 2014). However, it is equally possible that due to changes in analytical expectations or observational techniques researchers have become more aware of a phenomenon that was always there. Having said that, Simpson (2014) notes that it is unlikely that the widespread occurrence of such a marked phonetic feature in English would have escaped the attention of an acute observer like Catford, who mentions the occurrence of ejectives in English forty years ago only in passing: “in English they occasionally occur as the realization of final [p, t, k] in pathological speech, and in some northern English dialects” (Catford 1977: 70).

This study presents empirical evidence that ejectives are indeed on the increase for one variety of English, and, perhaps most interestingly, in the speech of one speaker. Data are drawn from the lab speech of a single speaker of Suffolk English gathered during two different recordings made approximately twenty years apart.

Method

The first recording, made in 1993, was designed to collect electroglottographic data on glottal dissimilation in Suffolk English, following an impressionistic phonetic description of patterns observed in spontaneous speech (Simpson 1992). The elicitation material comprised multiple tokens of simple sentences, such as “I’ll look at it at work” read individually from a computer monitor. Acoustic and EGG signals were recorded simultaneously in a sound-treated booth. The main aim of the data collection, which was to look at glottal activity during words, such as “at”, “it”, “put”, proved unsuccessful. However, an interesting by-product of the study was the high number of ejectives that the speaker produced for sentence-final /t k/, something that had not been observed in her spontaneous speech.

The second recording from 2014, was made as part of a pilot study in preparation for a project examining the distribution and production mechanisms of ejectives in English, German and Georgian. Building on the prevalence of ejectives that had been found 20 years earlier, it was speculated that laboratory elicitation techniques would again provoke the subject to produce ejectives. By contrast, to the elicitation materials of the 1993 recording, where the placement of sentence-final plosives in the elicitation material had been by chance, the sentences used in the second recording contained a more systematic distribution of sentence-final /p t k/. The recording was carried out in a quiet room at the subject’s home and sentences were again read individually from a computer terminal.

Results

The first recording contained 53 sentence-final /t k/ tokens (38 /k/, 15 /t/). In total, 32% of these were realised as ejectives, with only three of the 15 /t/ tokens being produced as ejectives. The second recording contained 50 sentence-final fortis plosives (10 /p/, 20 /t/, 20 /k/). In total, 58% of these were realised as ejectives ([p’] = 70%; [t’] = 65%; [k’] = 45%). This represents a 20% increase in ejective realisation for /k/ and more than a threefold increase in ejective realisation for /t/ compared to the first recording. These results are summarised graphically in Figure 1. Besides a significant increase in ejective realisation, it is interesting to note that this speaker produces more non-dorsal ejectives than she does dorsal ones, contrary to what we might have expected from other languages and other varieties of English (Ladefoged 2005; McCarthy 2011).
Discussion
Despite the two recordings not being directly comparable due to a mismatch of sentence-final /p t k/ in the two sets of elicitation materials, the results do strongly suggest an increase in the prevalence of ejectives in this speaker’s production over a twenty-year period. Although the findings are from a single speaker, they show yet again that sociophonetic features can develop and change over time within an individual, something shown most strikingly by changes in the Queen’s vowel space (Harrington et al. 2000). What still remains unclear is why ejectives should be so prevalent in this subject’s English in the first place. They may be a way of enhancing perceptual salience of articulatory place (Ogden 2009) in a linguistic activity demanding hyperspeech, such as reading. Alternatively, they could be an epiphenomenon of plosives preceding pauses accompanied by glottal closure (Local & Kelly 1986) in naturally occurring talk that has been carried over into read speech. Further analysis of read and spontaneous talk from the speaker discussed here together with other speakers from the same speech community should help us to shed more light on the distribution and functional aspects of ejectives in this variety of English.

*Figure 1:* Differences in ejective realisation of utterance-final fortis plosives from 1993 to 2014. (Note: no utterance-final bilabials in 1993 recording).
References
Sociophonetic perspectives on first language acquisition

Paul Foulkes
Department of Language & Linguistic Science, University of York, UK
paul.foulkes@york.ac.uk

*language never exists apart from the social fact, for it is a semiological phenomenon. Its social nature is one of its inner characteristics* (Saussure 1916 [1974: 77])

Linguistic scholars have always recognised that language varies systematically (i.e. non-randomly) across speakers, regions, and contexts of use. Comments on linguistic variation date back at least to Panini (ca. 600 BC; Chambers 2002a). Studies of phonological development have nevertheless generally treated most forms of variation simply as an inconvenience that complicates the task of acquisition. A popular textbook summarises phonological acquisition in rather typical terms:

*A fundamental problem an infant faces in acquiring the phonology ... is determining what constitutes a sound in the target language. That is, since sounds may be very similar but not identical phonetically, the infant must determine which of these similar phones comprise the same phonetic category (and later, the same phoneme) ... perceiving speech as phonetic categories is crucial for reducing the information in the signal, for programming the production of speech, and for discovering the units used for rules of language.* (Fromkin 2000: 658-9)

While Fromkin et al clearly acknowledge variation in the speech signal, by implication that variation is phonetically limited. The child’s task in acquisition thus resembles circling the relevant symbols for the language on an IPA chart, and adding annotations for positional allophones. The acquisition task also appears to involve *eliminating* phonetic variation in order to ‘reduce information’. Chambers (2002b) provides a preliminary model of how this might be achieved, which he refers to as the **sociolinguistic monitor**, to explain how children born to immigrants seem not to acquire features of their parents’ L2, but instead conform to the sociolinguistic patterns of their peers.

At some level this picture of phonological acquisition is probably realistic. The linguistic system must permit us to understand ‘the same word’ in different voices and in different contexts. But several issues arise when we consider the tasks of speaking and listening in more detail:

1. **variation can be phonetically unpredictable**
   - consider, for example, the range of /r/ variants used by speakers of English (at least [β o ɹ ɾ r ɥ ɰ ʁ w], ranging from labial to uvular in place, trill to approximant in manner);
   - acoustic patterns typical for an adult might represent completely different phonological units compared with the acoustic patterns used by a child.

2. **variation is highly complex, usually statistical rather than absolute, and results from many different sources simultaneously.** These include variation shaped by aspects of linguistic structure, such as positional allophones, but also a huge range of biological, learned (social) and external factors. Collectively these can be termed **indexical or sociophonetic factors** (since phonetic patterns index some social characteristic of the speaker or speech):
   - regional & social background (e.g. age, class, gender, communities of practice...);
   - speech style (e.g. degree of formality, etc);
   - pragmatic intent;
   - conversational structure (see e.g. Ogden 2012);
   - characteristics of the individual voice, including short-term effects of health & emotion;
   - external phenomena (e.g. transmission medium, environmental setting...).
(3) variation is ubiquitous
   • every speech event is unique.

(4) variation is meaningful
   • we tailor our speech to signal indexical properties under our control;
   • we interpret indexical properties to identify people we are talking to, interpret their
     background, understand their stance towards us, and negotiate spoken interaction;
   • perceptual processing is affected by systematic variation: we understand words and
     sounds faster if spoken in a familiar voice (Nygaard et al 1994) or a voice with indexical
     features congruent to the context (Walker & Hay 2011).

Learning to speak and to listen therefore involve much more than extracting a finite set of abstract
symbols and their rules of combination from a phonetically varying input. A competent language
user needs also to learn to control variation in her own speech, and to interpret a vast and ever-
changing range of phonetic variation when encountering it. These are not restricted tasks, faced only
once in a while. In ordinary activities people are constantly handling both referential and indexical
information in the speech signal. Some tasks require a constant shifting of focus, with linguistic and
indexical information both present and both important, but to different degrees at a particular time.
In a telephone call, for example, the opening exchange demands close attention to indexical
information to identify the talker, while the rest of the call involves repeated shifts in importance for
the abstract and indexical as propositional information is shared by the participants while
conversational turns are negotiated via appropriate phonetic cues. Note also that people learn and
respond to new sources of indexical information throughout their lives, facing unique speech events
and encountering new indexical categories, such as newly-met individual talkers. From the
perspective of acquisition, ‘reducing information’, as Fromkin proposes, might be necessary or useful
to identify a word and its referential meaning, but it is damaging if a key task in communication is,
for example, to identify the talker or to judge an interlocutor’s attitude. The phonological system,
then, if we consider that in its broadest sense of ‘knowledge about sounds’, must facilitate the
exchange and interpretation of phonetic variation as one of its key functions. The structure, primes
and processes in the phonological system need to be suited to the transmission of both referential
and indexical information. Yet few phonological models offer much insight into how this is achieved.

In this talk I outline a number of developmental studies, in both production and perception, where
the indexical strand is the focus. These studies illustrate the range of indexical information
presented to children, and show that children learn indexical patterns from the outset of life,
alongside and sometimes earlier than aspects of ‘pure’ abstract phonology. The studies show that:

• input to children is not monolithic, but embeds multiple sources of indexical variation. Speech
to boys contains more vernacular variants than speech to girls (Foulkes et al 2005). Standard
variants are used more in instruction and discipline, while local dialectal variants
dominate play, routine, and intimacy (Smith et al 2007, Nardy et al 2013).
• children start to develop sociolinguistic differences parallel to those found for adults by
around age 3;0 (Roberts 1997, Foulkes & Docherty 2006)
• phonetically transparent, frequent and non-arbitrary patterns (e.g. related to speaker sex or
adults versus children) are learned early, while more opaque, rare or arbitrary features (e.g.
regional accents) take more time (Foulkes 2010, Jeffries 2016 & this meeting)
• marginal, non-mainstream features apparent in the ambient community are not necessarily
ignored or filtered out of memory. Khattab (2009, 2013) shows that children of immigrants
use non-native features to converge in conversation with their parents (thus contra
Chambers 2002b), showing learning of highly marked phonetic forms and sophisticated
knowledge of how to deploy them for a communicative goal.
An adequate model of phonology to cater for indexical as well as referential information remains a distant goal. However, some promise is offered by exemplar models of language learning (Pierrehumbert 2003, Foulkes & Hay 2015). These models incorporate stochastic processing as a means to abstract both high level linguistic elements and also social categories from the ambient language. I explore some predictions made by exemplar theory and also outline some challenges.

In summary, I argue that a full understanding of the nature of phonological development and the qualities of the mature phonological system require a revised view of what it means to be a competent speaker-listener. A phonology is not just an abstract set of symbols and rules of combination. It also encapsulates a highly complex array of indexical information. Indexical variation is not a nuisance to be eliminated but a design feature of language.

References
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.

Bronwen G. Evans¹, Lidia Martín-Alvarez¹, Elizabeth Wonnacott²

¹Dept of Speech, Hearing and Phonetic Sciences, University College London, UK.
²Dept of Language & Cognition, University College London, UK.
bronwen.evans@ucl.ac.uk

Introduction

It is well-established that adult second-language learners benefit from High Variability Phonetic Training (HVPT; e.g., Logan et al., 1991). However, such training likely demands a significant amount of phonological awareness and attentional resources which may present difficulties for young children. Indeed, although in some studies children have been shown to improve more than adults after training (Giannakopoulou et al., 2013), HVPT appears to benefit older rather than younger children, suggesting that the ability to use HVPT might improve with age even though plasticity for speech likely declines (Shinohara & Iverson, 2013). The present study further investigates the potential benefits of high (i.e., HVPT) vs. low variability training for perception of the English /i/-/ɪ/ contrast in children aged 9-12yrs and adults. Additionally, the study investigates whether or not learning in the perceptual domain transfers to production and if this is affected by training paradigm (i.e., HVPT vs. LVPT).

Method

Sixty-three native monolingual Spanish participants (44 children, 19 adults) completed 5 sessions of high (HVPT) or low variability (LVPT; single talker) training. All had learned or were learning English at school but none had experience of living in a native English-speaking country.

Training was a 2-AFC minimal pair, picture identification task (e.g., ship-sheep) with feedback. On each training trial participants saw two pictures representing the two words in the minimal pair (e.g., a sheep and a ship) and heard one of the two words (e.g., ship). There were 40 different words (i.e., 20 minimal pairs) in total. The task was to choose the correct picture. Participants gave their response by clicking with a mouse on the picture of the word that they thought they had heard. If they chose the correct picture, then a smiley face appeared on the screen along with a coin, which participants collected for each correct answer over the course of the training. Participants then heard the word again with the matching image simultaneously displayed on the screen. If participants chose the wrong picture, then a sad face appeared on the screen and no coin was added. Participants then heard the word with the correct picture displayed simultaneously on the screen. Each training session comprised 320 trials, presented in a random order. In the HVPT condition, stimuli were produced by 4 talkers (40 words x 4 speakers x 2 repetitions), whilst in the LVPT condition they were produced by a single talker (40 words x 1 speaker x 8 repetitions).

To assess potential improvement, participants completed a category discrimination task (words and non-words, new talkers) and a word repetition task before and after training. In the category discrimination task, participants heard three words/non-words on each trial, spoken by two different talkers who varied randomly throughout; 2 words/non-words were the same and one was different (i.e. wheel/will/wheel for words; frit/frit/freet for non-words). The position of the 'odd' stimulus varied randomly throughout. Participants made their response by clicking on a number on the screen corresponding to the position of the 'odd one out', e.g., for wheel/will/wheel they needed to click on ‘2’ and for frit/frit/freet, ‘3’. They received no feedback and were not able to replay the stimuli. All participants completed a short practice session (3 trials) with words produced by 2 talkers from the training sessions. For production, participants recorded the same 40 words used in the category discrimination task in a random order. Participants were instructed to listen to the word and wait for the following 2-sec beep (interstimulus interval 0.5 secs) before repeating the word...
they had heard. To assess potential improvements in production, for each participant, a subset of 5 minimal pairs (i.e., 10 words) from the pre- and post-test recordings was extracted. Four native monolingual British English speakers identified the words (total: 380 trials) in a 2-AFC task. Listeners completed two blocks of testing, one for adults and another for children, with the order of presentation counterbalanced across participants and the order of presentation within each block randomized.

**Results & Discussion**

Preliminary analysis of the data demonstrated that for the pre-test category discrimination task and training session 1, adults outperformed children. Consequently, separate analyses were conducted for children and adults for both of these tasks. Separate repeated measures ANOVA analyses for adults and children demonstrated that all participants improved across training sessions, but LVPT-children improved more than HVPT-children. Interestingly, unlike in previous studies, adults showed no HVPT advantage. However, improvements in category discrimination were affected by both age and training condition. Children, but not adults improved in word-based category discrimination, and only children in the HVPT condition improved in non-word discrimination (Fig. 1). For adults, although there was a numerical improvement from pre- to post-test in the HVPT condition, there was no significant main effect of time or interaction with training condition; adults in both HVPT and LVPT showed no improvement from pre- to post-test in word and non-word discrimination. This could have been due to the relatively small number of adult participants or because adults started at a higher level of performance and had reached an upper limit of learning (see also Iverson & Evans, 2009). In contrast to perception, LVPT but not HVPT-children improved in their production of the /i/-/ɪ/ contrast. Likewise, there was a tendency for adults in the LVPT but not HVPT condition to improve in their production (Fig. 2). Future work is planned to investigate whether learning in production and perception operate differently, and whether this differs in children vs. adults (cf. Alshangiti & Evans, 2015).

Overall, the results suggest that children, unlike adults, may derive some benefit from training on an isolated contrast (see also Nishi & Kewley-Port, 2007). Additionally, although variability appears to be crucial for the generalization of perceptual learning, LVPT may be more beneficial for the acquisition of new articulatory targets.

**References**


Figure 1: Boxplots to show children and adults’ performance (proportion correct) at pre- (blue boxes) and post-test (green boxes) for the category discrimination task for words (left panel) and non-words (right panel). Participants who completed training in the HVPT condition are in the top row and those who were in the LVPT condition are in the bottom row.

Figure 2: Boxplot to show potential improvements in vowel intelligibility from pre- (blue boxes) to post-test (green boxes) for children and adults in the HVPT (top row) and LVPT (bottom row) conditions.
Processing grammatical gender in Dutch: Evidence from eye movements
Susanne Brouwer¹, Simone Sprenger², and Sharon Unsworth¹
¹Radboud University, Nijmegen, ²University of Groningen, Groningen, The Netherlands; s.brouwer@let.ru.nl

Studying how children process language can help us better understand children’s linguistic development. In the past fifteen years, children’s production data has shifted to new techniques allowing us to access children’s online processing (Snedeker & Huang, 2015). However, the link between production and online processing remains poorly understood. The aim of the present study is therefore to use both production and online processing techniques to better understand the acquisition of grammatical gender in Dutch-speaking children.

In many languages across the world, nouns have grammatical gender (Corbett, 1991). Dutch has a two-way gender system and makes a distinction between common and neuter gender. In this language, grammatical gender is marked on a number of agreeing elements accompanying the noun or referring to it; these include definite and demonstrative articles, relative pronouns, and adjectives (see Blom, Polišenská, & Weerman, 2008, for an overview). The focus in the present study is on the acquisition of definite articles. Common nouns are preceded by the definite article de, as in de schoen ‘the shoe’, whereas neuter nouns are preceded by the definite article het, as in het huis ‘the house’.

Previous research has demonstrated that grammatical gender in Dutch is typically acquired late by Dutch-speaking children (e.g., Blom et al., 2008). More specifically, Dutch-speaking children continue to make production errors with neuter gender until at least age six, overgeneralizing common gender forms of the article (de ‘the’) to neuter nouns (*de huis instead of het huis ‘the house’; e.g. Van der Velde, 2003). As most of this work used production data only, it is possible that children’s knowledge of Dutch gender has been underestimated. We therefore examined whether Dutch-speaking children (and adult controls) are able to use gender marking in the article preceding the object label to (a) anticipate the upcoming object label or to (b) facilitate the processing of that label as it is presented, using eye-tracking. In addition, we investigated whether children’s eye gaze behavior was related to their production of gender-marking on articles.

We tested 4- to 7-year old Dutch children (N=49) and adults (N=19). We chose this age range for the children because previous research has shown that during this period, children are in a transition stage, moving from non-targetlike to targetlike production of the (neuter) definite article (e.g. Blom et al., 2008). Accordingly, the children were divided into a non-targetlike (N=23) and a targetlike group (N=26) on the basis of their scores for neuter nouns on an elicited production task (see Unsworth et al., 2014, for details; following Blom et al., 2008). In the eye-tracking task, participants were presented with auditory sentences (e.g. Zie je het gele huis? Vind je ’m mooi? ‘Do you see the yellow house? Do you like it?’) and visual displays with two objects, representing nouns of either the same (uninformative, het huis ‘the house’ vs. het bed ‘the bed’) or different gender (informative, het huis ‘the house’ vs. de schoen ‘the shoe’). In our analyses we looked at whether participants could use gender-marking anticipatorily (i.e., before the onset of the noun) and facilitatively (i.e., from noun onset).

Our predictions are as follows. In informative contexts, we expect adults and targetlike-producing children to successfully use the available gender cue during online sentence comprehension. We predict that they use gender information as soon as possible, i.e., anticipatorily, even before the onset of the noun. The non-targetlike producers may behave in one of two ways. If comprehension does indeed precede production, the eye gaze behavior of the non-targetlike producers should pattern like the targetlike producers. In other words, although non-targetlike children are not able to...
produce targetlike gender-marked articles, they may be able to use them during online sentence comprehension. If non-targetlike producers indeed show better performance in comprehension than production, this would also indicate that previous findings showing a comparatively late development of Dutch gender may in part reflect an over-reliance on production data. However, it is also possible that the children in the non-targetlike group may not yet have acquired gender, and are thus also not able to use gender as a cue during online sentence comprehension.

Figure 1 illustrates our results. The data of the Dutch adults showed, as predicted, that they were able to anticipate the upcoming noun on the basis of the article (see Fig. 1A). This result is in line with previous work on gender processing with native-speaker adults in other languages (e.g., Lew-Williams & Fernald, 2007). The behavior of the targetlike producing children, who had high accuracy scores on the neuter nouns in the production task, patterned like the adults’ (see Fig. 1B). Before the onset of the noun, this group also showed proportionally more looks to the target noun in the different-gender condition, where the article was informative, than in the same-gender condition, where the article was uninformative.

Figure 1: Time course of gaze allocation to targets in the same and different-gender condition for (A) adults and children, divided into (B) targetlike and (C) non-targetlike production groups. The vertical dotted line represents the end of the time window analysis.

The non-targetlike producing children, whose scores on the neuter nouns of the production task were poor, behaved differently from the adults and the targetlike production group (see Fig. 1C). These children were not able to use the gender cue anticipatorily, that is, they did not make use of gender-marking before the onset of the noun. However, they were able to use the gender cue in a
facilitative way. That is, after noun onset, this group benefitted from the informativeness of the gender cue. How to interpret this discrepancy? If one considers facilitation as a successful way of using gender during processing, one could argue that the previous production data on the monolingual acquisition of Dutch gender may have underestimated the knowledge of these children. In other words, as for many other aspects of language, targetlike comprehension precedes targetlike production abilities. However, one could also argue that these children’s online comprehension abilities are not yet completely targetlike due to the lack of anticipation skills in this group. In either case, our results extend the previous offline comprehension data (Unsworth & Hulk, 2010) by showing that children as young as 4 years old are able to make (at least some) use of gender-marking. Moreover, our results indicate that learning to use gender-marking during comprehension may not necessarily take as long as previous research, based on production data only, has suggested.

In conclusion, this study has shown that Dutch adults and 4- to 7-year old children are able to use gender-marking during online sentence comprehension. Importantly, children use gender facilitatively before they can successfully produce it. However, once production is at targetlike level, they can also use gender-marking anticipatorily. Targetlike production thus seems to function as a trigger for online comprehension to be successful.

References
High or Low? Comparing high- and low-variability phonetic training in adult and child second language learners.

Elizabeth Wonnacott¹, Anastasia Giannakopoulou², Helen Brown³, Meghan Clayards⁴

¹ Language and Cognition, University College London,
² School of Psychology, University of Bedfordshire, UK
³ Department of Psychology, University of Warwick, UK
⁴ Department of Linguistics and School of Communications Sciences and Disorders, McGill, Canada
e.wonnacott@ucl.ac.uk

Speech-sound contrasts that have multiple phonetic cues can be particularly difficult for foreign-language learners. Research has shown that adult learners can be trained to discriminate non-native speech contrasts using phonetic training, generally involving minimal pair identification. Seminal work by Logan, Lively and Pisoni (1991, 1993) suggested that high variability input – i.e. where the contrast is exemplified across multiple tokens that are spoken by multiple talkers – is critical for generalization. The high variability phonetic training methodology has since become a standard in the field and has been successfully employed in numerous studies with adult learners (e.g. Iverson & Evans, 2009).

In contrast, only a handful of published studies have explored these effects in child learners. Each of these has compared children and adults given high-variability training, but have produced mixed results with respect to age. Wang and Kuhl (2003) tested 6-14 year old and adult novice learners on Mandarin tones, finding an advantage for adults at pre-test but no difference in learning rate between the two age groups; Heeren and Schouten (2008, 2010) trained novice 12 year old and adult native Dutch speakers to discriminate a Finnish length contrast, and again found an advantage for adults at pre-test, but similar improvement in both age-groups; Shinona & Iverson (2013) compared learning of the English /r/-/l/ contrast in native Japanese children (8-12 and 6-8 years), adolescents (15-18 years) and adults (25-59 years); adolescents and older children improved more between pre- and post-test than either 6-8 year olds or adults; Giannakopoulou, Uther and Ylinen (2013a) trained the English vowel contrast /i/ versus /ɪ/ (e.g. bean-bin) with child (7-8 years) and adult (20-30 years) native Greek learners of L2 English: Greek adults showed stronger performance at pre-test, however children showed greater improvements than adults at post-test. However, none of these studies directly compared the effectiveness of high versus low talker variability in children.

The current work compares the effect of variability – specifically variability in talkers– on Greek speaking 8 year old children and adults trained on the /i/-/ɪ/ speech sound contrast (e.g. ship vs. sheep; following Giannakopoulou et al 2013a). Critically, this computer-based training study contrasted high and low variability phonetic training.

Method

Native Greek speaking 8-year-olds (N=52), and adults (N=41) completed 10 training sessions in which they were exposed to 20 minimal word pairs (e.g. ship vs. sheep). On each training trial participants saw two pictures representing the minimal pair (e.g., a sheep and a ship) and heard one of the two words (e.g., ship). The task was to choose the correct picture, with trial-by-trial feedback provided. In the low variability condition audio stimuli were produced by a single talker whereas in the high variability condition they were produced by four different talkers (with total exposure matched across conditions). To control for prior knowledge of the /i/-/ɪ/ vowel contrast participants completed a pre-test in which we tested their ability to discriminate between the two vowels using a 3 interval oddity task. This discrimination task was repeated after the final training session to allow us to determine the amount of improvement (relative to the pre-test) for each participant. Test items were identical across conditions and included both trained and untrained words produced by
both a trained and an untrained talker. The untrained words and talkers are a crucial test of generalization of learning.

**Results**

In training (Figure 1), both age groups showed improvements across the sessions in both conditions. However, there was greater improvement in the low variability condition than in the high variability condition, although for children this only emerged in the second half of training. Higher accuracy in the low-variability condition may be due to the fact that repeated exposure to the same items produced by the same talker allowed participants to attune to idiosyncratic cues associated with that talker (Clopper & Pisoni, 2004).

In the discrimination task (Figure 2), we did not see the predicted benefit of high-variability input for either age group. Instead, children showed an effect in the reverse direction – i.e. reliably greater improvements in discrimination following low-variability input, even for untrained generalization items, although the result was qualified by (accidental) differences between the high- and low-variability groups at pre-test in this age group. Adults showed a numeric advantage for high-variability but were inconsistent with respect to voice and word novelty. Adults were however close to ceiling in this task.

In contrast to the findings of Giannakopoulou et al. (2013a), we did not see greater improvements for child over adult learners in either training or discrimination.

**Discussion.**

This research adds to the handful of studies demonstrating that, like adults, child learners can improve their discrimination of a non-native phonetic contrast via computerized training. There was no evidence of a benefit of training with multiple talkers, even for generalization. The results also do not support the findings of greater plasticity in child learning found in a previous paper (Giannakopoulou et al., 2013a). We discuss these results in terms of various differences between training and test tasks used in the current work compared with previous literature.

*Figure 1: (a) Adult (left) and (b) child (right) performance during training (error bars show standard error).*
Figure 2: (a) Adult and (b) child discrimination data. Mean increase in percent correct responses from pre to post-test (error bars show standard error).

References


Author Index

Akeroyd, Michael 41
Arnaud, Vincent 93
Auszmann, Anita 13

Barry, Johanna 41
Bastiaanse, Roelien 25
Baum, Shari 44
Başkent, Deniz 25
Brekelmans, Gwen 15
Brouwer, Susanne 90, 105
Brown, Helen 108
Bóna, Judit 4, 72

Cabrera, Laurianne 53
Casaponsa, Aina 41
Clayards, Meghan 44, 69, 108
Colby, Sarah 44, 69

D’Imperio, Mariapaola 54
Davis, Chris 58
de Leeuw, Esther 37, 87
Dellwo, Volker 7
Delvaux, Véronique 64
Dr. Gyarmathy, Dorottya 18

Eilola, Tiina M. 37
Esteve-Gilbert, Nuria 54
Evans, Bronwen 102
Evans, Bronwen G. 15

Farah, Chadi 54
Fikkert, Paula 90
Fougeron, Cécile 64
Foulkes, Paul 99

Gendrot, Cedric 64
Giannakopoulou, Anastasia 108
Giroud, Nathalie 7
Gordon-Salant, Sandra 40
Gősy, Mária 10, 81

Hagmueller, Martin 66
Hazar, Valerie 1, 58
He, Lei 7
Hermes, Anne 61
Horváth, Viktória 72
Iverson, Paul 21
Jeffries, Ella 75
Kennard, Holly 78
Kim, Jeesun 58
Krepsz, Valéria 10, 81
Laganaro, Marina 64
Lorenzi, Christian 53
Mair, Katharine 21
Martin-Alvarez, Lidia 102
Mattys, Sven 47
McCarthy, Kathleen 21
Meyer, Martin 7
Mora, Joan C. 22
Ménard, Lucie 64
Nagels, Leanne 25
Neuberger, Tilda 28
Niculescu-Gorpin, Anabella-Gloria 84

Oxbury, Rosie 87
Pailiereau, Nikola 31
Palmer, Shekeila 47
Pearce, Marcus T. 37
Pellegrino, Elisa 7
Poulton, Victoria 69
Ramachers, Stefanie 90
Redford, Melissa 34
Riverin-Coutlée, Josiane 93
Roman, Stephane 54
Rosen, Stuart 53
Simpson, Adrian P. 96
Skoruppa, Katrin 21
Sommers, Mitchell 57
Sprenger, Simone 105
Tenderini, Miriam S. 37
Terband, Hayo 50
Tuomainen, Outi 1, 58
Unsworth, Sharon 105

Vakula, Tímea 4
Vasileanu, Monica 84
Vianeva, Viktoria 41
Váradi, Viola 72

Wagner, Anita 25
Wonnacott, Elizabeth 15, 102, 108

Yamaguchi, Naomi 31
Keyword Index

Acoustic analysis 87
Acoustic change complex 21
Adverse listening conditions 58
Affective priming paradigm 37
Age 84
Age-related differences in learning 102
Aging 7, 44, 47, 58, 61, 69
Amplitude modulations 53
Apparent time 93
Articulation 61
Articulatory Phonology 61
Attention control 22
Auditory development 53
Automatic Speech Recognition 66

Breton 78

Childhood 53
Children 13, 54, 75
Closure duration 28
Coarticulation 34
Cochlear Implant (CI) users 25
Cognitive contributions 41
Cognitive control 22
Cross-sectional database 64

Dichotic test 81
Diphthongs 87
Disfluent repetition 4
Duration 72

Early cochlear implantation 54
Ejective 96
Eye-tracking 22, 25, 105

Filled pauses 72
Focus 54
Formants 31
French 31
Fricatives 21
Function 4
Function words 22

Geminates 28
Gestural coordination patterns 61
Grammatical gender 105

Healthy 61
Hungarian 28

Individual differences 22, 44
Individual variation 25
Inhibitory control 22

L1 acquisition 31
Language acquisition 81
Language change 84
Language contact 84
Language cues 41
Language development 105
Language Model 66
Late bilinguals 37
Lexical Anglicisms 84
Lexical competition 25, 69
Lifespan 4, 72
Limburgian dialects 90
Listening in noise 41
Longitudinal change 96
Longitudinal database 64

Maturational differences in learning 15, 108
Multicultural London English 87

N400 37
Native and non-native discrimination 90
Native speakers’ perception and processing 84
New Speakers 78

Online sentence processing 105

Partner-directed gaze 58
Perception 75
Perceptual learning 44
Perceptual Reorganization 90
Phonetic context 72
Phonetic training 15, 102, 108
Phonology 13, 78
Pre-adolescent language 87
Production 31
Prosody 34, 54
Pupilometry 25
Quebec French 93

Regional accent 75
Regional variation 93
Rhythm 22

Second language acquisition 15, 108
Second language learning 102
Segmental durational characteristics 7
Silent pause 18
Sociophonetic 96
Software 31
Speech intelligibility 53
Speech perception 28
Speech perception development 21
Speech production 54
Speech rhythm 7
Speech Segmentation 47
Speech sound processing 81
Speech Technology 66
Spoken French 64
Spoken word recognition 69
Spontaneous speech 18
Statistical Leaning 47
Stems 10
Stop consonant 28
Stress 78
Style-shifting 87
Suffixes 10
Syllable structure 61

Talker variability 15
Task effects 41
Temporal patterns 4, 18
Tone languages 90

Variability 31
Voice Acoustics 66
Vowel reduction 34
Vowels 13, 21, 31, 93

Word durations 10
Working Memory 47