Measuring children's sensitivity to phonological detail using eye tracking and pupillometry

Katalin Tamási

Submitted to the
Faculty of Human Sciences of the
University of Potsdam

2016

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-395954
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-395954

The work reported in this dissertation has been conducted under the auspices of the Erasmus Mundus Joint International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) of the Universities of Potsdam (DE), Newcastle (UK), Groningen (NL), Trento (IT) and Macquarie University, Sydney (AU).

Supervisors

Prof. dr. Barbara Höhle	(University of Potsdam)
Dr. Cristina McKean	(Newcastle University)
Prof. dr. Adamantios I. Gafos	(University of Potsdam)

Acknowledgments

Research, the social and human sciences kind in particular, cannot happen in a vacuum. It builds on and grows out of the generosity and graciousness of all the people involved. This work is no exception.

I first and foremost thank the participants and their families by generously giving their time and effort. Without them, this research would not have been possible.

I would like to express my gratitude to the three most extraordinary supervisors: Barbara Höhle, Cristina McKean, and Adamantios Gafos. As their expertize lies in different areas, I was able to learn about the unique skill sets needed for language acquisition research and experimental phonology and phonetics. They were always there for me no matter whether I needed guidance, feedback, encouragement, or references. All three of them generously gave their careful and constructive feedback on everything I have presented and written during the doctoral program and patiently read the dissertation from cover to cover. For these I will always be thankful.

I am much indebted to Barbara Höhle for being a constant during these three years. I thank her for having invited me to become part of her vibrant and thriving research program. The BabyLAB headed by her is one of the most smoothly run labs I know of; it was a real privilege to be a part of such a team. During the course of almost three years at the University of Potsdam, I have appreciated Barbara as a mentor. Working one floor below her office, I often had the opportunity to pick her brain about anything research-related. Despite her numerous research and community roles, she always made herself available and, whenever necessary, cleared away any potential hurdles with utmost efficiency.

Many thanks to Cristina McKean for her infectious enthusiasm and

tireless feedback throughout these three years. She is living proof that the supervising relationship can flourish irrespective of long distances and time zone differences. During my mobility period at Newcastle University, she made me feel most welcome by going out of her way to help with everything. Her (literal and figurative) door was always open.

I want to thank Adamantios Gafos for regularly finding the time in his incredibly busy schedule to meet with me, either in person or via Skype. A lightning-quick analyzer of ideas and results, he managed to come up with questions that made me think even deeper about how the arguments fit together. He challenged me to consider ramifications of the results by situating them in the large scheme of things.

This research could not have been completed without the help of the University of Potsdam BabyLAB personnel: Tom Fritzsche, Katja Schneller, Carolin Jäkel, Steffi Meister, Vanessa Löffler, Elisabeth Metz, and Lennard Gottmann. My many thanks to them for conducting participant recruitment and testing. I would like to thank Tom also for being an amazing collaborator, his technical support in getting the experiments off the ground, and his freely shared expertize in everything related to pupillometry and eye tracking.

The research presented in the dissertation was financially supported by the grant EMJD 520101-1-2011-1-DE-ERA. I hereby gratefully acknowledge all the support provided by the Erasmus Mundus Joint International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) of the Universities of Potsdam (DE), Newcastle (UK), Groningen (NL), Trento (IT) and Macquarie University, Sydney (AU). The directors and other researchers involved – most notably Barbara Höhle, David Howard, Roelien Bastiaanse, Lyndsey Nickels, Gabriele Miceli, and Ria de Bleser – have established an outstanding doctoral program. It was a privilege to get to know such wonderful people, with their researchers' hat on and off. I very much appreciate their continuous support in reviewing progress reports, keeping up with developments, welcoming us to summer and winter schools, and writing references when needed.

It has been a wonderful opportunity to travel to four of the IDEALAB consortium members, which really opened up the world to me. The summer schools at the University of Potsdam and winter schools at Macquarie University in Sydney and at the University of Trento were all

great experiences with inspiring discussions in and outside of the classroom. With regard to my host institution in Fall 2015, Newcastle University, many acknowledgments are in order. For all the generous invitations to attend meetings and seminars, the opportunity to present my work-in-progress projects, and being charming hosts in general, I would like to thank Cristina McKean, Ghada Khattab, David Howard, James Law, Jalal al-Tamimi, Tom King, Nick Riches, and many other researchers in and around the School of Education, Communication and Language Sciences.

My special gratitude goes to the Phonetics Lab and the Language Acquisition Colloquium at the University of Potsdam, and the Phonology and Phonetics Group and the Child Language Seminar at Newcastle University for their discussions and feedback on earlier drafts of the works presented in this dissertation. I thank Ghada Khattab, Natalie Boll-Avetisyan, Tom Fritzsche, Aude Noiray, and Silvana Poltrock in particular for giving valuable advice and feedback on my presentations. I name a few, but thank you all. I am furthermore grateful for Jalal al-Tamimi and Stavroula Sotiropoulou for cross-checking my production data annotations.

With the generous approval of the IDEALAB consortium, I was able to apply for an internship at the National University of Singapore. I also would like to thank Leher Singh for accepting my request to spend the summer of 2016 in her lab at the Infant and Child Language Centre at NUS. It has been a valuable learning experience.

I am immensely grateful for the funding provided for conference travels. Without the help of the Potsdam Graduate School, the *Kommission für Forschung und wissenschaftlichen Nachwuchs*, the Newcastle University conference fund, the National Science Foundation, and of course IDEALAB, I might not have been able to attend the international conferences that are so indispensable for professional development. My thanks also go out to the audiences of the 22nd and 23rd Manchester Phonology Meeting, the 33rd and 34th European Workshop on Cognitive Neuropsychology, the International Child Phonology Conference 2015, the Child Language Symposium 2015, the Attentive Listener in the Visual Word Conference, the Child Language & Eyetracking: Analyses & Rationale Workshop, and the 15th Laboratory Phonology Conference.

Whenever I had a question related to bureaucracy, I knew whom to turn to: Thank you Anja Papke, Helena Trompelt, and Ulla Behr, for working your administrative magic!

My dear IDEALAB- and Zebra-building-mates in Potsdam, Newcastle and elsewhere, Leigh, Conny, Laura, Adria, Vania, Tina, Miren, Sana, Michela, Farnoosh, Sean, Ana, Oksana, Nenad, Bernard, Assunta, Seckin, Srdjan, Kathi, Stavroula, Hui-Ching, Alexa, Weng, Stepan, Nele, Maja, Tanner, and many others, we have been through so much together. I thank you all for forming this wonderful and close-knit community – for renting and sharing accommodation everywhere we went, making office life relaxed and productive at the same time, and, most importantly, all the creative fun!

Thank you to my friends around the world (you know who you are) who haven't given up on me and continue to be there for me even in this busy and intense period of my life.

Drága Anyukám, Laci, Ferencz mama és papa, Tamási mama és papa, Zsuzsa, Dávid és minden családtagom, akiknek a nevét túl hosszú lenne most mind felsorolnom. Nagyon szerencsés vagyok, hogy ti itt vagytok nekem. Köszönöm nektek, hogy a fizikai távolság ellenére is mindenben támogattok és segítetek, amiben csak tudtok.

Zsombor, I cannot express how grateful I am for your patience, encouragements, all forms of support and endless confidence in my capabilities. Thank you for making me take all those much-needed breaks and diverting my attention to life beyond the PhD. You manage to be everything I could ever ask for and more. Now that I will have more free time and we seem to have the space for it, let's get that 'proper boardgaming table' finally!

Singapore, 30 October 2016

Contents

List of Figures

List of Tables

Chapter 1

General introduction

Even though I cannot remember having a shape sorter as a small child, I was told that I enjoyed playing with it thoroughly. And indeed, I can reconstruct the satisfaction and fulfillment that accompanied finding the hole that goes with a specific shape. I think this game is an excellent representation of the phenomenon that each shape has its own hole to match with, and so ideally there exists a one-to-one correspondence between shape and hole – a concept that every child mastering the shape sorter game learns.

Now imagine that a particularly devious parent tricks the infant by providing her with shapes that almost but not quite fit the holes in her shape sorter. Will the infant still be able to squeeze those shapes through, by accommodating small deviations? In the shape-sorter world, it may not be possible to pass anything through unless it is the exact match (allowing for minute variations due to differences in factory procedure, materials, wear-and-tear, etc.).

Naturally, there are aspects of spoken word recognition that a simple shape sorter analogy cannot possibly capture. Critically, *in lieu* of being solid objects, spoken words unfold over time. As such, phonemes within words partially overlap and their realization is context-dependent (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Therefore, it is no trivial feat to access the intended lexical representation even when provided with the ideal speech input (we shall discuss ideal and less-than-ideal inputs shortly). During word recognition, the speech signal is transmitted

to the auditory cortex, from which we extract individual words, and eventually map the signal to its respective lexical representation (Pulvermuller & Fadiga, 2010; Saffran, Werker, & Werner, 2006). Most current speech recognition models agree that this process involves three, potentially interrelated stages (identified by Frauenfelder & Tyler, 1987, summarized by Dahan & Magnuson, 2006).

During the first stage of word recognition, initial contact between the speech signal and possible lexical representations takes place. The speech signal activates lexical representations that are closest in corresponding to the signal. Distance between the signal and lexical representation can be conceptualized in several ways, contingent on the model of preference. It can be calculated in terms of phonemes (ORIGINAL COHORT model: Marslen-Wilson & Welsh, 1978; NEIGHBORHOOD ACTIVATION model: Luce, 1986; Luce & Pisoni, 1998), acoustics (REVISITED COHORT model: Marslen-Wilson, 1987; MINERVA2: Goldinger, 1998; DISTRIBUTED COHORT model: Gaskell & Marslen-Wilson, 2002), or acoustic / phonetic features (TRACE: McClelland & Elman, 1986; McMurray, Tanenhaus, & Aslin, 2002). Furthermore, different types of similarity are regarded to be crucial for activation by different models. Some emphasize the importance of word-initial similarity (COHORT models: Gaskell & Marslen-Wilson, 2002; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), some the similarity between lexical neighbors (NEIGHBORHOOD ACTIVATION model: Luce, 1986; Luce & Pisoni, 1998), and others integrate and extend the two by considering onset- and rhyme-overlapping similarity between any two words (TRACE: McClelland & Elman, 1986). For example, upon hearing the input '[t]hat's a baby', the following lexical representations may receive a boost of activation: the exact match <baby>, the word-initial cohort candidate <bay>, the lexical neighbor <maybe>, etc.).

The second stage in word recognition is selection, during which the activated candidates are evaluated with regard to the linguistic context. Most word recognition models conceptualize this stage as a competition among the candidates (AUTONOMOUS SEARCH model: Forster, 1989, LOCALIST ACTIVATION models: Blount & MacKay, 1991; Gaskell & Marslen-Wilson, 2002; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Morton, 1969). For instance, the sentential context, specifically the preceding

article in '[t]hat's a [...]' restricts the possible set of candidates to nouns. In our example, the lexical representations <baby> and <bay> stay in competition and <maybe> drops out due to the sentential constraint on word category.

The third and final stage in word recognition is integration. This involves the – not necessarily linguistic – environment in which the speech input is given, such as the visual context (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). If successful, the most likely candidate according to the hearer's assumptions remains active. Suppose our example speech input was about a picture representing a young and small human being. In this case, the visual context further restricts the candidate set, so that a single lexical representation survives. After the dropout of lexical representations incompatible with the visual information such as <bay>, the speech input '[t]hat's a baby' successfully integrates with the lexical representation <baby>.

The question that I will pursue in this dissertation is how infant word recognition works with ideal vs. non-ideal speech input. We consider what happens when the infant is provided with a speech signal that is a close, but not an exact match of a word (e.g., *vaby* instead of *baby*). How well can infants tolerate small changes made to the word and integrate the speech signal with the appropriate lexical representation? Would those manipulated word forms activate the intended lexical representation as readily as the correctly pronounced word, or would they require more cognitive effort to do so? Studying those questions may bring us closer to infer how infant lexical representations are structured.

1.1 Specificity and detail in lexical representations

We have seen above that spoken word recognition is a complex, multi-stage process. It entails interfacing the speech signal with possible candidates for lexical representation (<baby>, <bay>, <maybe>), selecting the most appropriate candidate <baby> out of all the plausible lexical representations by using linguistic and non-linguistic contextual information and integrating the speech signal therewith (Dahan & Magnuson, 2006). The question is how specific lexical representations need to be to enable successful word recognition.

On the one hand, lexical representations need to be specific enough to be differentiated from other representations. That is, the hearer needs to be able to arrive at a single lexical representation to integrate with the heard input. On the other hand, lexical representations need to be abstract enough to accommodate inter- and intra-speaker variation. That is, the hearer needs to set aside information not directly relevant to word recognition: categorical factors such as speaker identity, gender, and accent and continuous factors such as speech rate and emotional state.

Even though the word recognition process takes place with minimal effort in adults given sufficient information and context, it works less well in young language learners: Children seem to require clearer, less ambiguous input for the communicational intent to come across. The gating task is a method that allows the researcher to measure what proportion of the speech signal is necessary to recognize a given word. During a gating task, participants are played increasingly greater portions of the word and are instructed to identify it as accurately and quickly as possible (e.g., Grosjean, 1980). In these tasks, preschool-aged children[1] require more acoustic information than adults to decide which word they are presented with (Garlock, Walley, & Metsala, 2001; Walley, 1988). Also, children's speech recognition is less assisted by sentential context (Nittrouer & Boothroyd, 1990) and more hampered by background noise than adults' (e.g., Klatte, Lachmann, & Meis, 2010).

In what follows, we summarize current research regarding the degree of specificity and detailedness of early lexical representations. Then we discuss supporting and contradictory experimental evidence and the methodologies employed therein. Finally, we attempt to adjudicate between them based on the available findings and motivate our research.

1.1.1 Representational immaturity vs. cognitive limitations? In search of the reasons behind poor performance

It is possible that the need for less ambiguous input in children's word recognition stems from representational immaturity. Under this assumption, early lexical representations may be less specific than adult ones

[1] A note on terminology: Throughout the dissertation, by *preschool-aged*, I mean 3-to-6-year-olds and by *infant*, I mean children below the age of 3 years.

such that they cannot be readily matched with the speech input. Accordingly, some acquisition models assume that lexical representations begin as 'holistic' and gradually become more specific. Young language learners may only store holistic properties about words that can be made more unique if motivated by lexical pressures. For this reason, I refer to those models *en masse* as holistic (Charles-Luce & Luce, 1990, 1995; Fikkert, 1995, 2010; Jusczyk, 1992, 1993, 1997; Metsala & Walley, 1998; Treiman, 1983; Vihman, 2010; Walley, 1988, 1993; Walley, Smith, & Jusczyk, 1986; Waterson, 1971).

Holistic accounts come in several varieties and focus on different aspects of language development. For example, infants' mental lexicons are proposed to operate with general auditory analyzers that do not preserve phonetic information (WRAPSA model: Jusczyk, 1992, 1993) or to use 'global perceptual processes' (Charles-Luce & Luce, 1995, p. 733). In Radical Templatic Phonology (Croft & Vihman, 2003; Vihman, 2010; Vihman, Nakai, DePaolis, & Hallé, 2004), words are thought to provide the perceptual units of recognition, from which language-specific phonotactic templates, i.e., patterns complete with syllabic and metrical information, are abstracted out as part of the learning process. Underspecification theory conceptualizes early words to only contain place feature specification for the vowel, from which it can spread to neighboring consonants (Fikkert, 1995, 2010).

There is an alternative explanation for children's poor performance in word recognition tasks. Even though children may possess highly specified representations early on, they may be less proficient speech decoders than adults owing to developmental differences in attention, memory capacity, executive function, and experience. Following Zesiger, Lozeron, Lévy, and Frauenfelder (2011), we call this stance the early specificity hypothesis.

Early specificity models posit that fine-grained phonetic detail is present but not always accessible for speech processing (Munson, Edwards, & Beckman, 2011; Werker & Curtin, 2005). Instead, the abstract phonological level of representations emerges later as a result of vocabulary growth and/or language use. Various lexical restructuring and emergence accounts (Bybee, 2003; Ferguson & Farwell, 1975; Mehler, Dupoux, & Segui, 1990; Menn, 1983; Munson et al., 2011; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Pierrehumbert, 2002; Storkel, 2002; Studdert-Kennedy,

1986; Swan & Goswami, 1997; Waterson, 1971; Werker & Curtin, 2005) are more compatible with this view than the holistic one, though the boundaries between the two are not always clear-cut.

Holistic and early specificity hypotheses agree that abstractness in lexical representations is emergent. That is, the common feature of these accounts is that they both assume that lexical representations incrementally acquire specificity and detail throughout language development, from the beginning of lexical acquisition to the proficient user stage (until school-age and beyond: Garlock et al., 2001; Metsala, 1997, 1999; Metsala, Stavrinos, & Walley, 2009; Metsala & Walley, 1998; Munson et al., 2011; Storkel, 2009; Walley, 1988, 1993; Walley, Metsala, & Garlock, 2003; Werker & Curtin, 2005; Ziegler & Goswami, 2005). Change in the mental lexicon is proposed to be instigated by factors such as vocabulary size (Charles-Luce & Luce, 1990), neighborhood density (Garlock et al., 2001), phonotactic probability (Storkel, 2009), word (type and token) frequency (Goodman, Dale, & Li, 2008), word familiarity (Metsala, 1999), and age of acquisition (Garlock et al., 2001). At the start of the word learning process, children's mental lexicon contains a few, easily distinguishable words. The acquisition of vocabulary items may necessitate more efficient ways to differentiate words from each other. As more similar-sounding words are acquired, the pressure to discriminate between them may drive the encoding of an increasing amount of detail (Charles-Luce & Luce, 1990, 1995; Hollich, Jusczyk, & Luce, 2002; Hoover, Storkel, & Rice, 2012; Jusczyk, 1993; Metsala, 1997; Pierrehumbert, 2002; Stokes, 2010, 2013; Storkel, 2002, 2009; Walley, 1993; Werker & Curtin, 2005). In particular, dense lexical neighborhood, high phonotactic probability (De Cara & Goswami, 2003; Garlock et al., 2001; Stokes, 2010, 2013; Storkel, 2002, 2009), and familiarity with and/or repeated exposure to words (Barton, Miller, & Macken, 1980; Goodman et al., 2008; Metsala, 1999) have been proposed to facilitate the emergence of sub-lexical, i.e., syllabic and, subsequently, phonemic structure in the mental lexicon.

To reiterate at which point holistic and early specificity hypotheses diverge, it is whether the specificity and/or detail is present from the early stages of language acquisition. Holistic models assume no specificity and detail to be preserved in infant lexical representations. Early specificity models, on the other hand, assume specificity and/or detail to

be present from the initial stages of language acquisition, though representations may not be efficiently structured to be readily accessible for the purposes of word recognition and meta-linguistic awareness (Munson et al., 2011; Werker & Curtin, 2005). As such, task demands and high cognitive load may mask the presence of specificity and detail (Werker & Curtin, 2005).

Experimental evidence supporting holistic accounts mostly comes from studies with 'offline' methodologies. Offline tasks assess the participants' overt and explicit responses given to the experimental stimulus, such as identity judgment – e.g., 'does *baby* and *vaby* sound the same?' Expected response: 'yes' or 'no' – or speech production – e.g., 'can you name this picture?' Expected response: '(this is a) baby'. Note that, as such, the outcome of these tasks is contingent not only on linguistic knowledge, but also on metalinguistic awareness as well as the maturational level of motor (articulatory, fine-motor) systems and cognitive processes (memory, attention).

Offline studies on lexical processing include observing language production and other overt behavior in several ways. This may include spontaneous speech, that is, speech recorded during conversations and/or play time (Ferguson, 1986; Ferguson & Farwell, 1975; Waterson, 1971). Further, elicited speech involves speech that the child is specifically instructed to make as part of a task requirement (Nittrouer et al. 1989, Pitrat, Logan, Cockell & Gutteridge, 1995, cited by Garlock et al. 2001). Metalingustic tasks may require the children to identify phonemes and syllables from the speech stream (e.g., 'do you hear a *t* in the word *stay*'?) or manipulate phonemes and syllables by addition or deletion (e.g., 'can you add *s* to the beginning of the word *pit*'?) (Eilers & Oller, 1976; Metsala, 1997; Metsala et al., 2009; Treiman & Baron, 1983). Other metalinguistic tasks include gating tasks (discussed in the previous section), (non)word repetition tasks (e.g., 'can you say this word after me: *gezik*'?), (Garlock et al., 2001; Walley, 1988), (dis-)similarity judgments (e.g., 'do the words *blif* and *belif* sound the same?') (Berent, Harder, & Lennertz, 2011; De Cara & Goswami, 2003; Pertz & Bever, 1975), and word plays (e.g., 'let's add *ez* to the beginnings of words!') (Treiman & Baron, 1983; Treiman & Breaux, 1982). Accordingly, offline studies assess the child's lexical knowledge by the consciously generated, mostly speech production output that the child

provides (Metsala, 1997; Metsala et al., 2009; Treiman & Baron, 1983).

Most of these studies found the speech production of preschool-aged children to undergo various voicing and nasal assimilations, consonant harmonies (Ferguson & Farwell, 1975; Menn, 1983), and excessive coarticulation (Nittrouer et al., 1989). None of these patterns are exhibited in typical adult speech production. Moreover, children failed to detect small – phonemic or featural – changes (Eilers & Oller, 1976) and failed to manipulate words on the phonemic level, being only able to consciously access the syllabic structure, if that at all (Garlock et al., 2001; Treiman & Baron, 1983). These findings are interpreted such that children may recognize and categorize on the basis of overall prosodic or acoustic shape, but not that of phonetic or phonological features (Vihman et al., 2004). Children's lexical representations may furthermore be more diffuse than adults', using the word or the syllable as organizing units. Thus, additional level of detail may be emergent (Garlock et al., 2001).

A curious exception, in favor of the early specificity hypothesis, comes from children's spontaneously occurring 'slip of the tongue'-type speech errors. These errors seem to operate on the phonemic level (*big dog* < *dig dog*, not **dog dog*: Gerken, 1993, cited by Gerken, Murphy, & Aslin, 1995; Stemberger, 1989) and as such support the existence of phonemically specified early lexical representations. This shows that it is possible to garner support for the existence of a phonemic layer in early lexical representations by looking at the production output.

Lack of granularity in lexical representations may not be the only reason behind the inability to perform well in most offline tasks. It is plausible to assume that high cognitive load associated with those tasks contributed to poor performance, a point that has been already remarked by Gerken et al. (1995) and Fernald, McRoberts, and Swingley (2001), among others. This is the reason why infant language researchers are keen on developing paradigms that present the least amount of cognitive demand. Children may give suboptimal responses due to developmental lag in memory, attention, and general information-processing skills, irrespective of their linguistic abilities. It stands to reason that a task wherein the role of such factors are minimized is more likely to uncover infants' linguistic abilities.

Offline studies, therefore, need to be complemented by studies which do not demand learned behavioral responses from children for three main

reasons. First, language production and related overt behavior is subject to the maturation of motor coordination systems, which continue to develop well into school-age (Kiparsky & Menn, 1977; Smit, 1993; Thelen, 1996). Thus, studies that assess children's production are confounded by the immature state of the motor system. Second, preschool-aged children who are unable to reliably identify and manipulate phonemes in words (Charles-Luce & Luce, 1990) might still be able to possess tacit phonemic knowledge as it develops prior to explicit, propositional knowledge (Ellis, 2008). Third, it is notoriously difficult to elicit complex overt responses from infants and preschool-aged children, thus limiting the age range one can investigate. For these reasons, the findings of this line of research need to be enriched with other methodologies that are devised to look at covert correlates of – possibly younger – children's lexical knowledge.

Online tasks, in contrast to offline ones, focus on performance during the 'here-and-now' of (language) processing and assess responses that are largely outside of the participant's conscious control such as eye movement and pupil dilation. For a discussion on the offline / online methodology spectrum, see Hewlett (1990). The following section summarizes the results on infant lexical knowledge gained from online studies.

1.1.2 Detecting mispronunciation: specificity in lexical representations

Probing perceptual abilities is extensively used in infant language research. It has been widely recognized that infants, especially in the first half year of their life, are sensitive to phonetic detail, be it native or non-native (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk & Aslin, 1995; Kuhl, 1993; Werker & Lalonde, 1988; Werker & Tees, 1984; Werker, Yeung, & Yoshida, 2012). Moreover, infants are able to form phonetic categories by setting aside not directly relevant acoustic differences (Hochmann & Papeo, 2014). Such knowledge may be mistakenly interpreted as an index of the ability to detect abstract, phonological features. Phonetic discrimination and generalization skills, albeit impressive, do not invite inferences on the structure of lexical representations, especially as these skills emerge before lexical knowledge. Therefore, it is still unclear precisely what details are stored in the early mental lexicon (Fikkert, 2010; Pater, Stager, & Werker, 2004; Stager & Werker, 1997; Walley, 1993).

Due to its size, the adult lexicon contains many phonological neighbors, i.e., words that differ by a single feature or phoneme – e.g., *cat* and *pat* (Luce & Pisoni, 1998). Depending on the number of their phonological neighbors, words can be situated in dense or sparse lexical neighborhoods. Therefore, the ability to detect a small yet contrastive change between words, especially between words in dense lexical neighborhoods, is crucial to building up a mature lexicon. Even though affected by neighborhood density (Luce & Pisoni, 1998), adults in real-life conversations have little to no trouble in differentiating phonological neighbors, hardly any confusion ensues despite the acoustic, perceptual, and featural resemblance. The question arises whether this knowledge is evident in children. Can they distinguish closely resembling words, or, rather, do they merge them into one category? The confusion of similarly sounding words could indicate that early lexical representations are not specific enough and thus do not encode fine-grained detail. Contrastively, sensitivity to small differences between words would suggest that infant lexical representations are specific.

Several paradigms have been devised with the goal to test the lexical knowledge of differently aged infant populations. One way to address whether infants represent detail is to take infants' ability to detect mispronunciations of words as a measure of specificity and detailedness of early words (for a review, see Altvater-Mackensen & Mani, 2013). The name-based categorization paradigm (Nazzi, Floccia, Moquet, & Butler, 2009; Nazzi & New, 2007) and the intermodal preferential looking paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Golinkoff, Ma, Song, & Hirsh-Pasek, 2013) were adapted to investigate infants' lexical knowledge, along with paradigms originally used to measure perceptual skills such as the head turn preference (Hallé & de Boysson-Bardies, 1996; Jusczyk & Aslin, 1995) and switch paradigms (Stager & Werker, 1997).

Two types of tasks have been employed with these paradigms. Word learning tasks in which infants learn to associate novel labels with previously unknown objects have been used with all the above experimental paradigms. They are more widely used than word recognition tasks in which infants are presented with already known labels and their mispronounced forms (popular with intermodal preferential looking paradigms). Both word learning and word recognition tasks have their pros and cons.

The main advantage of word learning tasks over word recognition tasks is that lexical knowledge is not as much a constraining factor, thus, the stimuli set can be better controlled for phonetic and phonological characteristics. On the other hand, apart from the difference in cognitive load, there is some indication that words acquired at the experimental session might not be as precisely represented as previously known words (Ballem & Plunkett, 2005; Werker & Curtin, 2005) and so performance in word learning tasks may not provide a reliable proxy of children's lexical knowledge. Furthermore, infants may find tasks that call for their real-world knowledge, that is, word recognition tasks with existing object–label associations, more engaging than word learning tasks. The paradigms are presented below in a loosely decreasing order of cognitive demand.

One of the word learning tasks is name-based categorization. In name-based categorization tasks, participants undergo a presentation and a categorization phase. The presentation phase consists of infants learning label-object associations with three differently looking objects. Two objects receive the same label (e.g., *nuk*), and the third a slightly different label (e.g., *muk*). In the categorization phase, the infant is asked which object 'belongs to' one of the objects labeled *nuk*, effectively testing whether the infant managed to group the two similarly labeled objects into one category. Infants succeed in the task if they select the similarly named object (e.g., the one labeled as *nuk*) and fail if they select the differently named one (e.g., the one labeled as *muk*). Twenty-to-thirty-month-olds are shown to succeed in name-based categorization tasks (Nazzi et al., 2009; Nazzi & New, 2007).

In switch or interactive word learning paradigms, infants are tested on their ability to learn object–label associations with two different objects. After learning novel object–label associations in the familiarization phase, infants may be presented with either a novel association (e.g., labeling an object *muk* after it was consistently labeled *nuk*) or an established one (e.g., labeling an object *nuk* throughout) and were tested on their ability to detect whether the association is novel or not. Seventeen-to-twenty-four-month-olds succeed with such paradigms, an age group slightly younger than that in name-based categorization tasks (Curtin, Fennell, & Escudero, 2009; Dietrich, Swingley, & Werker, 2007; Eilers & Oller, 1976; Havy & Nazzi, 2009; Mani, Coleman, & Plunkett, 2008; Werker, Fennell, Cor-

coran, & Stager, 2002).

In the intermodal preferential looking paradigm, the familiarization phase is similar to those of switch paradigms: it involves infants being exposed to object–label associations with novel labels. In the testing phase, infants are expected to match the label with the object it was paired with in the familiarization phase (as evidenced by longer looking times towards that object) and are furthermore expected not to match a mispronounced label with the object (as evidenced by no difference in looking times or by longer looking times towards the other object). A further discussion on the methodology can be found in Section 1.2. It has been found that 14-to-17-months-olds succeed with this paradigm, again younger than with the previous paradigms (Ballem & Plunkett, 2005; Fennell & Waxman, 2010; Yeung, Chen, & Werker, 2013; Yoshida, Fennell, Swingley, & Werker, 2009).

In a head turn preference paradigm (Hallé & de Boysson-Bardies, 1996; Jusczyk & Aslin, 1995), word learning and word recognition tasks are typically set up as follows. Participants are presented with lists of correct vs. mispronounced (or familiar vs. novel) words through loudspeakers on either side while their listening preferences – as measured by the duration of the head turns towards the active loudspeaker – are recorded. Infants are first trained with practice items to learn the contingency between the side of the head turn and the condition, then familiarized with certain items and finally tested with both previously presented or not presented items. The mispronunciation is detected in case the duration of the gaze orientations are different in response to the correct vs. mispronounced (or familiar vs. novel) items. Longer head turn towards the novel items than to the familiar items is interpreted as novelty preference, the opposite as familiarity preference.

Studies based on head turn preference paradigms yielded mixed results regarding infant sensitivity to fine-grained detail. Hallé and de Boysson-Bardies (1996) found that 11-month-old French-reared infants were not sensitive to most types of the mispronunciations (voicing and manner change, word-medial change), a result interpreted as evidence for holistic lexical representations. However, using the same procedure with infants of British-English and Dutch backgrounds, 11-month-olds were able to detect onset, but not word-medial and offset changes (Swingley, 2005;

Vihman et al., 2004). By 14-months of age, infants are able to overcome the difficulties in detecting offset changes (Swingley, 2009). As in head turn preference paradigms only auditory stimuli are presented (no referential cues are provided) and the stimuli are played in a loop, it is not clear whether infants treat those lists of strings as words, i.e., whether they attempt to match them to lexical representations or react to acoustic / phonetic differences within the stimuli. For these reasons and since task characteristics related to cognitive and attentional load determines how well infants can perform the task (Werker & Curtin, 2005; Yoshida et al., 2009), head turn preference paradigms have been proposed to be less sensitive as tools of mispronunciation detection than other methods (Zesiger et al., 2011).

Intermodal preferential looking paradigms have been employed in word recognition studies as well. With this paradigm, 12-to-19-months-olds can detect change made to already known words with a variety of contrasts (Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002). Specifically regarding word-recognition contexts, children as young as 12-month-old are able to detect mispronunciations involving place of articulation (e.g., *bin-din*) (Fennell & Werker, 2003; Fikkert, 2010; Jusczyk & Aslin, 1995; Pater et al., 2004; Zesiger et al., 2011). By 19 months of age, children are shown to pick up on a range of phonological contrasts including changes in voicing (e.g., *dog-tog*), manner of articulation (e.g., *swing-twing*) (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Ren & Morgan, 2011; Swingley, 2003, 2005; Swingley & Aslin, 2000, 2002; Vihman et al., 2004; White & Morgan, 2008; White, Morgan, & Wier, 2005) as well as height and backness in vowels (e.g., *bed-bid, brush-brash*) (de Boysson-Bardies, Hallé, Sagart, & Durand, 1989; Mani, Mills, & Plunkett, 2012; Mani & Plunkett, 2011b). Although mostly demonstrated with onset manipulations, sensitivity is not restricted to the word-initial position. Infants are able to detect mispronunciations in word-medial and final positions as well, be it a vocalic or consonantal change (Mani et al., 2012; Mani & Plunkett, 2011a; Ren & Morgan, 2011; Swingley, 2009, 2016).

Overall, word learning and word recognition studies assessing infants' lexical knowledge suggest that lexical representations are more specific than previously shown by offline tasks. That is, online tasks show lexical representations to highly specified even for infants. The observed

trend is that the age at which infants succeed in detecting a segmental change is contingent on the cognitive demands posed by the experimental paradigm. We have seen that the cognitive load associated with a task is loosely proportional to the age at which the task can be used to detect mispronunciation skills (consistent with PRIMIR's Werker & Curtin, 2005 assumptions). Children below the age of three years are demonstrably able to detect the distinction between correct pronunciation of familiar words and a whole range of phonetic changes introduced to those words. This shows that similarly to adult lexical representations, early words are highly specified, i.e., sufficiently specified so as to more readily allow establishing a match with the correct than with the mispronounced version of the label (*baby* ∼ <baby> vs. *vaby* ∼ <baby>). As such, mispronunciation detection studies converge on the finding that infant lexical representations are specific.[2]

1.1.3 Detecting *degrees of* mispronunciation: Sub-phonemic detail in lexical representations?

In the previous section, we saw that infants are able to discriminate correctly produced and featurally manipulated word forms and concluded that such an ability indicated specificity of early lexical representations. Adults, however, seem to possess an even finer-grained sensitivity that goes beyond what has been demonstrated for infants to date. Their performance is affected by the degree of featural similarity between the correct and manipulated word. That is, adults react differentially to small vs. large degrees of featural manipulation of the input, e.g., *baby* and *vaby* vs. *baby* and *shaby*. Note that the contrast in question lies between the

[2]There is an another online paradigm not based on mispronunciation detection whose findings support this conclusion. It probes infants' lexical knowledge by silent or cross-modal priming (Mani & Plunkett, 2010a, 2011b). In these tasks, 18-month-old and 24-month-old infants were first presented with a prime referent (e.g., a cat) and then two other familiar referents, a target and a distractor picture side by side (e.g., those of a cup and a shoe), one of which phonologically related to the prime referent (the labels *cat* and *cup* share their onset). Finally, either the target or the distractor picture is named. Even though the prime referent is only silently presented during the task, infants looked significantly more to the primed target picture than the unprimed distractor picture when labeled. This result is taken as evidence that individual phonemes from an implicitly named label are automatically extracted already at 18 months of age.

differential response given to small vs. large degrees of mispronunciation, as opposed to the contrast between correct and mispronounced labels, e.g., *baby* vs. *vaby*.

Adult gradient sensitivity to various phonological and phonetic feature changes has been documented using several different paradigms: auditory lexical decision (Milberg, Blumstein, & Dworetzky, 1988), phoneme monitoring (Connine, Titone, Deelman, & Blasko, 1997), intra-modal and cross-modal priming (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Marslen-Wilson, Moss, & van Halen, 1996), and preferential looking paradigms (McMurray et al., 2002; Mitterer, 2011; Reinisch, Jesse, & McQueen, 2010; Salverda, Dahan, & McQueen, 2003; White, Yee, Blumstein, & Morgan, 2013) (though for equivocal findings see Cole, Jakimik, & Cooper, 1978 and Ernestus & Mak, 2004). Such gradient sensitivity suggests that mature lexical representations are not only specific, but also fine-grained enough to specify the degree of overlap with phonological neighbors and other minimally different nonwords. This is only possible if lexical representations contain sub-phonemic detail. Note that with the term *sub-phonemic*, I aim to remain agnostic as to the nature of the stored information, i.e., whether it entails acoustic / phonetic or abstract / phonological features. Mani and Plunkett (2011a) used the term *sub-segmental* for similar reasons. See Section 5.2 for a discussion on the topic.

Moreover, this fine-grained sensitivity has been also found to extend to preschool-aged children (Creel, 2012; Gerken et al., 1995). A question arises whether, given the right conditions, infants are able to exhibit this sensitivity. Would they be able to appreciate the degree of featural similarity between closely resembling words? Such a finding would suggest infants to be able to encode information at the sub-phonemic level in their lexical representations.

After reviewing mispronunciation detection studies on gradient sensitivity, we found the available evidence inconclusive. Based on the available research, it is not well established whether infants are able to detect differing degrees of mispronunciation. On the one hand, such ability has been demonstrated with intermodal preferential looking paradigms (degree of phonological overlap positively predicted the proportion of target looks: Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008;

White et al., 2005. On the other hand, lack of sensitivity to degree of mismatch has also been found using similar paradigms (Bailey & Plunkett, 2002; Swingley & Aslin, 2002). This discrepancy will be detailed below and further revisited in Chapters 2 and 3.

Table 1.1 below summarizes the literature on children's scope of sensitivity in word recognition.[3] In the remainder of this section, we discuss the results along with methodological limitations and advancements thereof introduced by each piece of research. We start with the lexical decision task developed by Gerken et al. (1995) (see section 1 in Table 1.1). In this task, preschool-aged children were required to indicate via button press if the label they were presented with was a word or not. Even though Gerken et al. (1995)'s results were consistent with gradient sensitivity (that is, sensitivity to the difference between small and large degrees of mispronunciation), the authors employed a very limited stimulus set –only one word per child was manipulated in multiple ways – so as to minimize task demands, which leaves open the question of generalizability. Further, due to the nature of the task (i.e., pressing a button on either side depending on the correctness status of the heard word), recruiting children below the age of 3 years would not have been feasible.

Working with the intermodal preferential looking paradigm enabled Bailey and Plunkett (2002) and Swingley and Aslin (2002) to assess sensitivity in younger populations (14-24 month-olds) (see sections 1 and 2 in Table 1.1). Although these studies demonstrated differential response to correct vs. mispronounced words, they did not detect a difference between small (one-feature) and large (two- or three-feature) deviations. The overall procedure and results of Zesiger et al. (2011) are very similar to those of Swingley and Aslin (2002): with the manipulation of one or two features and using two familiar images, both 12- and 17-month-old infants were found sensitive to mispronunciation, but not to the degree thereof.

[3]Creel (2012) is not included in the summary table because of its slightly different aim: the investigation of accented speech recognition with a preschool-aged group: 3-5 year-olds. However, its overall findings are comparable to those that found gradient sensitivity in infants.

Table 1.1: Summary table of studies probing for gradient sensitivity.[1]

Study	**Gerken et al. (1995)**	**Bailey and Plunkett (2002)**
Paradigm	lexical decision task	IPLP
Age (months)	36–48	18, 24
Manipulation	Corr / 1F / 2F	Corr / 1F / 2F
Position in word	whole word	onset
Stimulus set	restricted	predominantly labial-initial
Stimulus creation	not balanced	PoA / MoA / V
D familiarity	N/A	yes
Preset ISI	not reported	not reported, probably yes
Fixation point	N/A	none
Trial structure	only auditory, button press	5.5 s, naming at 2250 ms
Exclusion crit.	<90% practice trials correct	>1500 ms of looks
Analyses	RT, accuracy	LLT
Results	Corr | 1F | 2F	Corr | 1F, 2F
Interpretation	gradient sensitivity	no gradient sensitivity

Study	**Swingley and Aslin (2002)**	**White and Morgan (2008)**
Paradigm	IPLP	IPLP
Age (months)	14–15	19
Manipulation	Corr / 1F / 2,3F	Corr / 1F / 2F / 3F / Nov
Position in word	onset (5/6)	onset
Stimulus set	restricted	predominantly labial-initial
Stimulus creation	not balanced	PoA / MoA / V, type counterb.
D familiarity	yes	no
Preset ISI	not reported, probably yes	no
Fixation point	only before first trial	before trial & naming
Trial structure	9 s, naming at 3 s	13 s, naming at 4 s
Exclusion crit.	367–2000 ms naming	naming score, 2 images fixated, 0-3 s naming
Analyses	PTL, Latency to *T*	PTL
Results	Corr | 1F, 2/3F	Corr | 1F | 2F, 3F | Nov
Interpretation	no gradient sensitivity	gradient sensitivity

Table 1.1 – continued from previous page

Study	**Ren and Morgan (2011)**	**Mani and Plunkett (2011a)**
Paradigm	IPLP	IPLP
Age (months)	19	18, 24
Manipulation	Corr / 1F / 2F / 3F / Nov	Corr / 1F / 2F / 3F / Nov
Position in word	coda	onset
Stimulus set	not reported	featurally balanced
Stimulus creation	not reported	±high / ±back / ±round
D familiarity	no	no
Preset ISI	no	yes
Fixation point	before trial & naming	none
Trial structure	13 s, naming at 4 s	5 s, naming at 2500 ms
Exclusion crit.	none	familiar words, 2 images fixated
Analyses	PTL	PTL, LLT
Results	only linear trend reported	Corr | 1F | 2F,3F Corr | Nov
Interpretation	gradient sensitivity	gradient sensitivity (24, not 18)

[1]Only relevant experiments of the respective studies are included; i.e., Gerken et al. (1995): experiments 2 and 3, White and Morgan (2008) experiment 1, Ren and Morgan (2011) experiment 1. White et al. (2005) is not considered separately as in many ways it is a precursor to the first experiment of White and Morgan (2008). Abbreviations: IPLP = intermodal preferential looking paradigm, Corr = correct word form, 1-3F = one-three feature change, Nov = novel word form, PoA = place of articulation, MoA = manner of articulation, V = voicing, counterb. = counterbalanced, *D* = distractor, N/A = not applicable, ISI = inter-stimulus interval, RT = response time, LLT = longest took towards target, PTL = proportion of target looks, *T* = target

White and Morgan (2008) were able to demonstrate infant gradient sensitivity. They took several measures to improve on previous methodologies that probably contributed to their successful detection of sensitivity to the degree of mispronunciation in infants (see Section 2 in Table 1.1). They introduced an unfamiliar distractor picture, manipulated features more systematically by employing a fuller range of featural contrasts, counterbalancing for feature type, and controlling for word position; to yield stronger effects they used a fixation point and experimenter fine-tuning of the inter-stimulus interval; and to produce more robust results they employed more stringent exclusion criteria than previous studies.

The exclusion criteria and other procedural changes imposed by White and Morgan (2008) merit further discussion because many subsequent studies including our intermodal preferential looking study (presented in Chapter 3) adopted most of them. According to White and Morgan (2008), what made the difference that enabled them to detect gradient sensitivity was introducing an unfamiliar picture as distractor. They suggest that using an unfamiliar picture may have encouraged the infants to consider mispronunciations of the target label as novel word forms and thus more suitable for a picture whose label is unknown (via the mutual exclusivity mechanism[4]).

Trials that were not associated with increased looks towards the target picture when presented with the correct target label were excluded. White and Morgan (2008) argued that the validity of the intermodal preferential looking paradigm is predicated on the assumptions that (1) children are familiar with the correct form of the target label, (2) they can recognize the target image, (3) and they can attach the target label to the target image, evidenced by their looking preference patterns – hence the exclusion of nonconforming trials. To make an informed decision on which picture is a more suitable match to the heard input, children needed to fixate on both images in the pre-naming phase. Trials which did not adhere to this requirement were excluded for the above reason. To maximize the chances of detecting an effect, restricting the analysis to an either *ad hoc*

[4]Mutual exclusivity is an assumption that an object can be labeled in exactly one way and vice versa, one label refers to exactly one object. It is a word learning mechanism that allows the child to map unknown labels to unfamiliar objects *in lieu* of objects whose name they are already familiar with (Halberda, 2003; Hirsh-Pasek, Golinkoff, & Hollich, 2000).

or *post hoc* time window is a popular exclusion criterion among studies that analyze data with gross averaging techniques (for a review, see Luche, Durrant, Poltrock, & Floccia, 2015).

The introduction of the fixation point in between trials and before the naming phase is a standard practice in EEG (electroencephalographic) and MEG (magnetoencephalographic) research as it encourages the participant to make use of the whole visual field and decreases the chance of perseverance in looking direction. This is an especially important consideration for infant research. Luche et al. (2015) showed (re-)centering to be advantageous in intermodal preferential looking studies as well.

Using a similar methodology, Ren and Morgan (2011) and Mani and Plunkett (2011a) extended the findings of White and Morgan (2008) by showing that gradient sensitivity can be demonstrated when degree of mispronunciation is manipulated in the word-final coda position and in vowels, respectively (c.f., section 3 in Table 1.1). These findings show that encoding sub-phonemic information seems to take place throughout the whole word and is not restricted to the prominent position of the word onset. The general finding is also concordant with those obtained by simulations of infant performance (Mayor & Plunkett, 2014) using the TRACE model of speech recognition (McClelland & Elman, 1986), indicating gradient sensitivity to sub-phonemic detail (although note that gradedness emerged only when inhibition within lexical competitors and / or phonemes is suppressed). Mani and Plunkett (2011a) moreover went on to demonstrate that, for vowel mispronunciations, acoustic distance tends to be a better predictor of children's looking preference than pure featural distance. This point will be elaborated on in Section 5.2. The studies of Tamási, McKean, Gafos, Fritzsche, and Höhle (in press) and Tamási, McKean, Gafos, and Höhle (2016a) will be presented in Chapters 2 and 3, respectively (section 4 in Table 1.1).

Taken together, we conclude that it is necessary to complement offline studies with studies that minimize cognitive and task demands. Findings collected from online, mispronunciation detection and priming studies are at odds with holistic models that infants' and preschool-aged children's lexical representations do not specify information below the syllabic level. Instead, these findings are consistent with the early specificity hypothesis that not only adult, but also preschool-aged and infant lexical representa-

tions are specific.

1.1.4 Detecting mispronunciation *in clusters*: Specificity and detailedness of clusters in lexical representations

Recall that the main research question of the dissertation is whether infants are able to recognize ideal (i.e., correctly pronounced) vs. non-ideal labels (i.e., speech input whose pronunciation is manipulated on some level). We have seen that infants are able to differentiate the correct form from a close, but not exact match (e.g., *vaby* instead of *baby*). We interpreted sensitivity to the difference between the correct and incorrect form to indicate that integration with the appropriate lexical representation (in this case, <baby>) suffers as a result of the manipulation. More cognitive effort is required to interface with the lexical representation when a mispronounced label is provided as opposed to the correctly pronounced label.

So far, we have focused on research that manipulated words on or below the phonemic level (e.g., *baby* and *daby* differ by a single place feature). This section introduces a research avenue that goes beyond the phonemic level by manipulating words that contain consonant clusters, i.e., adjacent consonants (e.g., *s* and *t* in *stone*). In the remainder of this section, we briefly review research on cluster acquisition, motivate our research, and sketch the main research questions. These points will be revisited in more detail in Chapter 4.

Children at the initial stages of language acquisition do not attempt to produce clusters at all, but postpone production until they are approximately two years of age (Lleó & Prinz, 1996). When they do attempt production, clusters are found to be prone to deletion and simplification errors (e.g., *play* → *pay*) (Barton et al., 1980; Dyson & Paden, 1983; Fox & Dodd, 1999; Lleó & Prinz, 1996; McLeod, Doorn, & Reed, 2001; Smit, 1993; Stemberger & Treiman, 1986; Watson & Scukanec, 1997). Moreover, children have trouble manipulating clusters as evidenced by their problems in non-word repetition (Gathercole, Willis, Emslie, & Baddeley, 1991), spelling (Treiman, 1991; Treiman & Cassar, 1996) and breaking up of words containing clusters (Treiman, 1983). By the time children learn to consciously manipulate the internal structure of words and become literate – at least children acquiring alphabetic writing systems –, they are forced

to break up the clusters into their individual phoneme elements (De Cara & Goswami, 2003). What is not clear is whether such word manipulations and learning to read helps children to develop lexical representations with adult-like detail? Or, rather, are infants' lexical representations containing clusters already structured similarly to adults'?

The inability to produce and manipulate clusters correctly may suggest that early words only contain a single slot for a cluster that branches out as the lexicon grows when more differentiation is needed. This assumption is compatible with holistic accounts (Fikkert, 2010; Vihman, 2010). Early specificity models (Munson et al., 2011; Werker & Curtin, 2005), on the other hand, hypothesize that lexical representations are specific. In the case of consonant clusters, this would mean that each consonant is represented in detail. In line with the arguments reconstructed in the previous sections, the holistic and early specificity accounts differ in their assumptions about whether detail beyond the phonemic structure is encoded in early lexical representations.

To date, no studies using online methods have considered how consonant clusters are represented in the – early and mature – mental lexicon. Therefore, it is timely to study what information may be stored about consonant clusters in lexical representations. We address this question with two complementary approaches, looking at both perceptual and production skills of infants using different types of clusters in Chapter 4. Since there exists no adult research that could be used as a comparate, adult participants were also recruited to participate in the study.

The manipulation in our online study involves inserting or epenthesizing a vowel in between the consonants (e.g., *stone* → *sətone*). We propose to manipulate two types of clusters. Homorganic clusters are produced at approximately the same place of articulation (e.g., the consonants *s* and *t* in the word *stone*), while heterorganic clusters are produced at different places (e.g., *s* and *w* in the word *swine*). We study how specific lexical representations that contain these two types of clusters are by asking whether those lexical representations are specified and detailed enough to be differentiated from their epenthesized forms.

Thus we plan to assess how epenthesis within homorganic and heterorganic consonant clusters may affect word recognition. Would infants be able to activate the appropriate lexical representation upon presentation of an epenthesized as well as the correct word form? Or rather, could infants still recognize manipulated word forms such as *sətone* as the intended word *stone*, albeit with more cognitive effort? If infants are not able to differentiate between words containing correct vs. epenthesized cluster forms, that would indicate that they represent clusters holistically, without the detail required to be differentiated from the epenthesized version. If however, infants are able to give a differential response to correct vs. epenthesized forms, that would suggest more cognitive effort was required to activate and interface with the corresponding lexical representation. This in turn would speak to the existence of highly specific cluster representations. A related possibility is that homorganic and heterorganic clusters behave differently in lexical representations. Homorganic clusters may form a more cohesive unit than heterorganic clusters by virtue of their place features. In this case, epenthesis in homorganic clusters may induce a larger pupillary response than epenthesis in heterorganic clusters in comparison to their correctly produced counterparts. This result would be consistent with the existence of structural differences between the two cluster types in lexical representations.

1.2 Methodological background

1.2.1 Why use pupillometry and eye tracking?

> *Direct description of the child's actual verbal output is no more likely to provide an account of the real underlying competence than in the case of adult language [...] Obviously one can find out about competence only by studying performance, but [it] must be carried out in devious and clever ways...*
>
> –N. Chomsky (1964, p. 36, cited by Golinkoff et al., 2013)

Pupillometry is a method highly suitable for assessing the performance of young children, being based on an involuntary psycho-sensory reflex, i.e., pupil dilation (Laeng, Sirois, & Gredebäck, 2012; Loewenfeld, 1993; Nieuwenhuis, Geus, & Aston-Jones, 2011). In pupillometry, instead of recording the pattern of gaze fixations, the eye-tracking equipment is used to measure change in pupil size over time. Increased pupil dilation has been found to be an index of (short-term) working memory load and hence task difficulty, as shown by digit span tasks (Kahneman & Beatty, 1966), mental rotation tasks (Just, Carpenter, & Miyake, 2003), mathematical calculations (Hess & Polt, 1960), and tasks manipulating attentional allocation (Karatekin, 2004).

As activity in the Locus Coeruleus - Norepinephrine (LC-NE) system is one of the key modulators of task performance, efficiency, and attentional allocation, it is also implicated in the stimulus-evoked dilation of the pupil: The more neuronal activity the LC-NE system exhibits, the larger task-evoked pupillary response is obtained (Aston-Jones & Cohen, 2005; Laeng et al., 2012; S. Marshall, 2002; Nieuwenhuis et al., 2011). In parallel to adult research, greater pupillary response in young children has been interpreted to be a proxy of surprise, novelty, and cognitive effort (Hepach & Westermann, 2013, 2016; Jackson & Sirois, 2009; Karatekin, 2007; Sirois & Brisson, 2014). More recently, pupillometry has been found to be a viable tool in child language research, being sensitive to detecting acoustic (dis-)similarity (Hochmann & Papeo, 2014), semantic mismatch (Kuipers & Thierry, 2011), and – most important for the current project

– mispronunciations (Fritzsche & Höhle, 2015).

We highlight three aspects that makes pupillometry an especially appealing tool in language development research.

1. Pupillometry is minimally demanding. The processing of the experimental stimuli, i.e., watching while listening, does not necessitate an overt and explicit behavioral response. In our studies, processes unrelated to the investigated phenomenon – recognition of distractor pictures, memory requirements, evaluation and decision processes – are greatly reduced. Since task demands affect engagement and performance, removing or minimizing these confounds set up ideal conditions to succeed in the task. In addition, as less is expected from children, they are less likely to become fussy during the experiment, which may prevent data loss.

2. The continuous nature of the pupillary response may provide an alternative to investigate children's reaction to degrees of mispronunciation than a pseudo-categorical response employed by preferential looking paradigms, i.e., looking at either the target or the distractor image (M. M. Bradley, Miccoli, Escrig, & Lang, 2008; Kahneman, Tursky, Shapiro, & Crider, 1969; Klingner, 2010b).

3. Pupillometry is inexpensive and easy to learn. Although electrophysiological (de Haan, 2007) and brain-imaging techniques (Hebden, 2003; Peterson & Ment, 2001) avoid the shortcomings that looking time paradigms introduce, such techniques require specialized equipment and expertise thereof. The eye-tracking equipment needed for pupillometry is already widely available in the child language research community and the technical competencies required for pupillometry can be readily acquired by those already familiar with the equipment.

We hypothesized that pupil dilation may directly reflect the costs induced by processing a mispronounced word and therefore provide a fine-grained insight into the effect of the degree of mismatch between the correct and the mispronounced form. The pupillary response has been exploited as a dependent measure in single-picture pupillometry studies (Fritzsche & Höhle, 2015; Tamási et al., in press; Tamási, McKean, Gafos, & Höhle,

2016b). In single-picture pupillometry, children are presented with a picture whose label they are familiar with (e.g., a picture of a baby) and a correctly pronounced or mispronounced label (e.g., *baby* or *shaby*). During the task, participants' pupil dilations are monitored.

Moreover, the pupillary response as a dependent measure has been used first in an intermodal preferential looking paradigm by our study (Tamási et al., 2016a, presented in Chapter 3). The intermodal preferential looking paradigm is a well-known and popular application of the eye-tracking machine in language development research. During the task, children are presented with real-word objects or pictures on screen (infants and preschool-aged children: typically two objects / pictures, older children and adults: four objects or pictures, sometimes written words, c.f., Allopenna, Magnuson, & Tanenhaus, 1998), accompanied by an auditory label that matches one of the pictures. The picture related to the label is named the target, the other pictures are considered distractors. The associated looking behavior is monitored and then collected by the eye tracker to be analyzed later. Longer looking time to one as opposed to the other pictures is interpreted as preference (Golinkoff et al., 2013; Oakes, 2011; Zesiger et al., 2011) and more specifically, an attempt to establish a semantic link between the heard label and the picture (e.g., Swingley & Aslin, 2000).

The intermodal preferential looking paradigm is a standard procedure with established guidelines that have been developed in the last 30 or so years (see Golinkoff et al., 2013 for an extensive review). The popularity of intermodal preferential looking paradigm may be due to its convenience, accessibility, and relatively inexpensive hardware and software requirements of the eye tracker compared to other neuropsychological measures such as EEG (electroencephalography) and fMRI (functional magnetic resonance imaging). Several of the advantages detailed with regards to single-picture pupillometry apply to preferential looking paradigms as well. Like single-picture pupillometry, preferential looking paradigms do not require infants to explicitly respond to instructions or perform any overt action, which again allows infants to reveal their language abilities before the mastery of speech production. Apart from infants, clinical or otherwise atypical populations (children living with autism, hearing impairment, pervasive language disorders such as specific language impairment,

or motor deficits such as cerebral palsy) may benefit from the 'hands-free' approach of online methodologies such as single-picture pupillometry and the intermodal preferential looking paradigm (Houston, Stewart, Moberly, Hollich, & Miyamoto, 2012; Naigles & Tovar, 2011). Some studies praise the potential of these online methodologies to become part of the diagnostic arsenal of several disorders (Friend & Keplinger, 2008; Houston et al., 2012). Given the above considerations, we believe the online methodologies pupillometry and eye tracking eminently qualify as the 'devious and clever' means envisioned by Chomsky, (1964, p. 36), empowering the researcher to study the hidden layers of language development.

1.2.2 Data analysis

Due to preliminary tests on the pupil dilation data of the single-picture pupillometry study reported in Chapter 4, the primary analysis of choice in this dissertation is linear mixed effects modeling. Visual inspection of the data distribution and statistical tests indicated that the data violated both the assumptions of normality (Kolgomorov-Smirnov test, $D = 0.997$, $p < .001$) and that of homogeneity of variance (Fligner-Killeen test, $\chi^2(3) = 574.234$, $p < .001$). For this reason, linear mixed effects models with random intercepts and slopes were employed using the `lmer` function in the `lme4 R` package (Bates, Maechler, Bolker, & Walker, 2014). Estimates were chosen to optimize the log-likelihood criterion. Apart from handling unbalanced data sets and categorical variables better than generalized linear models such as ANOVA, linear mixed effects models allow the inclusion of multiple random effects in the same model (Baayen, Davidson, & Bates, 2008; Bates, 2005).

A statistical model that includes both fixed and random effects is called a mixed effects model (Bates, 2005). Repeatable predictors, normally manipulated by the experimenter (e.g., the number of feature changes in our experiment), are encoded as *fixed effects*. With fixed effects, we test whether the differences within the predefined contrasts are significant (e.g., whether the correctly pronounced items are treated differently than mispronounced items). Non-repeatable predictors (i.e., participants and items) are incorporated into the same model as *random effects*. Random effects estimate the unique effect of an individual participant or item, disregarding other random and fixed effects. By accounting for random

effects in the model, we are better able to assess the effect of our experimental manipulation. Moreover, in accordance with the suggestion of Baayen (2008, Chapter 7.1) and Barr, Levy, Scheepers, and Tily (2013), random slopes can be specified for each random effect: Each predictor that is dependent on participants (e.g., vocabulary size) may receive a within-item random slope and each predictor that is dependent on items (e.g., neighborhood density) may receive a within-subject random slope provided that the model does not become overfitted.

As the role of lexical and sub-lexical variables such as neighborhood density, phonotactic probability, vocabulary size, and word frequency is well-documented in spoken word recognition (Charles-Luce & Luce, 1990, 1995; Demuth & McCullough, 2009; Garlock et al., 2001; Goodman et al., 2008; C. C. Levelt, Schiller, & Levelt, 2000; Luce, 1986; Luce & Large, 2001; Luce & Pisoni, 1998; Mattys, Jusczyk, Luce, & Morgan, 1999; Metsala, 1997; Stokes, 2010, 2013; Storkel, 2002, 2009), we introduce those into our statistical models as control variables. Although they are of no primary concern to our research, we wish to control for their effects nevertheless. This way, we attempt to disentangle their effects on the outcome from that of the experimental manipulation. Section 5.2 pulls together and interprets the effects of lexical and sub-lexical factors in our studies, and situates them in the literature.

Post-hoc cluster-based time-course analyses provide a data-driven approach to explore when significant differences emerge across any two factor levels (Maris & Oostenveld, 2007). By investigating the latency and the duration of the interval of the contrast between two levels, it becomes possible to determine how differently they were handled by the participants. Cluster-based time course analyses have originally been proposed to analyze electro- and magnetoencephalographic data, but can be extended to any continuous dependent measure including looking preference and pupil dilation (e.g., Dink & Ferguson, 2016; Luche et al., 2015). The brief description of the algorithm to test contrast significance is as follows. First, individual paired sample t-tests across the two factor levels find the significant ($p < .05$) t-values across the whole time frame. Second, clusters (e.g., contiguous significant t-values) are identified, for which a cluster-level t-value is calculated as the sum of all single sample t-values within the cluster. Third, the significance of cluster-level t-values are assessed

by generating Monte Carlo distributions ($N = 2000$) thereof and determining the probability of their occurrence given the distribution. Those clusters whose t statistic exceed the threshold (Bonferroni-corrected for multiple comparisons depending on the number of levels) are then tabulated for each contrast. Post-hoc time-course analyses are introduced in more detail in Chapter 2. Exploratory approaches that involve peak and latency of the smooth pupillary curve as well as the peak and latency of the velocity of the pupil change are described in the Appendix.

1.2.3 Population

Throughout the studies to be presented, 30-month-old children were recruited. The rationale for testing this specific age group and not younger children came from methodological considerations. The lexicon of children below the age of 30 months severely limits the featural makeup of eligible words as regardless of the specific language context, those lexicons tend to contain primarily labial-initial words (Vihman & Croft, 2007). The 30-month-old lexicon, in contrast, licensed the creation of a more diverse and balanced stimulus set than would be possible for a younger age group. More specifically, this allowed a featurally balanced consonantal set, additionally cross-balanced for feature type and change in Chapters 2 and 3 (voice changes: voiced-to-voiceless, voiceless-to-voiced; manner changes: stop-to-fricative, fricative-to-stop; place changes: inward [labial-to-coronal, labial-to-dorsal, coronal-to-dorsal], outward [vice versa], c.f., Tables 2.1 and Table 3.1). Furthermore, the 30-month-old lexicon also enabled us to cross-balance for type of onset manner in homorganic and heterorganic clusters in Chapter 4 (c.f., Table 4.1).

1.3 Issues addressed in the dissertation

The question this dissertation aims to investigate is when and how details in lexical representations arise over the course of language acquisition. Ultimately, we attempt to determine what infants know about words and track the emergence of such knowledge with predominantly online methods. In what follows, we present three studies borne out of this research program. The first two studies delve into detailedness within the phoneme in early words using single-picture pupillometry (Tamási et al., in press, c.f., Chapter 2) and preferential looking paradigm alongside pupillometry (Tamási et al., 2016a, c.f., Chapter 3) and the third study examines level of detail beyond the phoneme using single-picture pupillometry and speech production analysis (Tamási et al., 2016b, c.f., Chapter 4).

Thus two overarching issues are addressed in the dissertation: In Chapters 2 and 3, we examine whether there is evidence for sub-phonemic detail in infant lexical representations. In Chapter 4, we move on to investigate cluster processing in infants and adults, with the ultimate aim to study how clusters are represented in the infant and adult lexicon.

Part I

Looking within the phoneme

Chapter 2

Pupillometry registers toddlers' sensitivity to degrees of mispronunciation[1]

[1]A version of this chapter is being published as: Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (in press). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology.* doi: 10.1016/j.jecp.2016.07.014

Abstract

This study introduces a method suited for investigating toddlers' ability to detect mispronunciations in lexical representations: pupillometry. Previous research has established that the magnitude of pupil dilation reflects differing levels of cognitive effort. Building on those findings, we use pupil dilation to study the level of detail encoded in lexical representations with 30-month-old children whose lexicons allow for a featurally balanced stimulus set. In each trial, we present a picture followed by a corresponding auditory label. By systematically manipulating the number of feature changes in the onset of the label (e.g., *baby* $\sim$ *daby* $\sim$ *faby* $\sim$ *shaby*), we test whether featural distance predicts the degree of pupil dilation. Our findings support the existence of a relationship between featural distance and pupil dilation. First, mispronounced words are associated with a larger degree of dilation than correct forms. Second, words that deviate more from the correct form are related to a larger dilation than words that deviate less. This pattern indicates that toddlers are sensitive to the degree of mispronunciation and, as such, it corroborates previous work that found word recognition modulated by sub-phonemic detail and by the degree of mismatch. We thus establish that pupillometry provides a viable alternative to paradigms that require overt behavioral response in increasing our understanding of the development of lexical representations.
Keywords: *Phonological development; featural distance; lexical representations; mispronunciation detection; pupillometry; eye tracking.*

2.1 Introduction

The nature of lexical representations stored by children as they develop a mental lexicon is a widely studied aspect of language acquisition. An unresolved issue is the level of detail encoded in early lexical representations: whether they are holistic and undifferentiated or, rather, adult-like in their detailedness. Recent findings show that children's lexical processing is modulated by featural manipulations made to words (e.g., *dog-tog*), suggesting that early words must be sufficiently specified so as to enable establishing a match to a given label (Fikkert, 2010; Swingley & Aslin, 2000; Yoshida et al., 2009). However, the precise degree of this specificity in early words requires further investigation. Whilst some studies have found evidence for children's ability to detect differing degrees of mismatch (Mani et al., 2012; Ren & Morgan, 2011; White & Morgan, 2008), others have not (Bailey & Plunkett, 2002; Swingley & Aslin, 2002). This paper seeks to determine children's sensitivity to degree of mispronunciation with a tool which minimizes task demands: pupillometry.

Numerous studies attempting to uncover the nature of early lexical knowledge have probed infants' perceptual abilities. It has been widely recognized that infants are excellent discriminators of phonetic detail – be it native or non-native – in their first months of life and are able to form phonetic categories (Curtin & Archer, 2015). Furthermore, infants are able to categorize consonant-vowel sequences by disregarding irrelevant acoustic differences (Eimas et al., 1971; Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). Such skills have been sometimes interpreted as an index of the infants' ability to detect phonological features. However, these discrimination skills are not necessarily revealed during word processing as children may not distinguish newly learned words from phonological neighbors (Stager & Werker, 1997). These discrepancies between discrimination and word recognition have raised the question: What details are stored in the developing lexicon? To address this issue, studies may take infants' ability to detect *mispronunciations of words* as a measure of specificity and detailedness of early words (e.g., Swingley & Aslin, 2002).

In such studies, 17-19-month-olds have demonstrated sensitivity to a range of contrasts effected through featural changes including voicing (e.g., *dog-tog*), manner of articulation (e.g., *swing-twing*) (Swingley & Aslin, 2002) as well as height and backness in vowels (e.g., *bed-bid, brush-brash*)

(Mani et al., 2012). Moreover, children as young as 14 months old have the ability to detect mispronunciations involving place of articulation (e.g., *bin-din*) (Swingley & Aslin, 2000).

Given this body of research in mispronunciation detection it can be concluded that infants are able to detect the difference between correct and featurally manipulated word forms. A natural step forward is to ask how far their lexical knowledge extends: Are infants sensitive to the degree of mispronunciation (i.e., to the degree of featural distance between the correct and incorrect forms)? Such gradient sensitivity would suggest that lexical representations are not only specific, but also fine-grained enough to encode the degree of overlap with other minimally different words. This is only possible if early words contain sub-phonemic detail.

So far, only a handful of studies, all using the preferential looking paradigm, have considered the question of whether children younger than three years are sensitive to differing degrees of mismatch between a target word and its mispronounced variant. These studies have obtained mixed results: some demonstrate sensitivity to degree of mismatch (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008) while others do not (Bailey & Plunkett, 2002; Swingley & Aslin, 2002).

By including a greater range of contrasts than previous studies and, crucially, unfamiliar distractors in a preferential looking paradigm, White and Morgan (2008) demonstrated sensitivity to mismatch in 19-month-olds. This study manipulated the number of consonantal feature changes introduced to target word onsets. The results indicated a gradual decline in the proportion of children's target looking time as the number of feature changes increased (corrected for looking times in the salience phase). For example, children's looking time towards the picture of 'keys' was greater than towards an unfamiliar object (e.g., an abacus) when presented with the correct label *keys*. With labels exhibiting a one-feature change (*teys*), infants still preferred to look towards the target, but to a lesser extent than in the correct condition. When two-feature changes were introduced (*deys*), infants exhibited non-significant target preference and with three-feature changes (*zeys*) a non-significant distractor preference (the two- and the three-feature-change conditions overlapped). These findings suggest that children were able to retrieve the appropriate lexical representation and consequently, match the label with its corresponding picture when

presented with the correct label. Similarly, the children were able to do so when the onset differed by one feature, though less successfully than with the correct label (as evidenced by the drop in the proportion of looking times towards the target). Further, children did not appear to establish a link with the item exhibiting two- and three-feature changes and either the target or the distractor. Ren and Morgan (2011) replicated these findings by manipulating coda consonants. Also, a similar graded sensitivity in looking time has been observed when manipulating vocalic rather than consonantal featural distance (Mani & Plunkett, 2011a). Twenty-four-month-olds (but not 18-month-olds) showed sensitivity to the degree of mismatch such that correctly pronounced labels and one-feature deviations resulted in target preference, while larger two- and three-feature deviations yielded a weak distractor preference.

We highlight here an important shared methodological characteristic of the studies described above, which provides the impetus for our study: the use of a preferential looking paradigm with two pictured objects presented simultaneously in each trial. As White and Morgan (2008) point out, the presence of two potential referents for the auditorily presented (mispronounced) word form requires a process that determines whether the presented label is a new word that may be mapped to the distractor object or whether it can still be mapped to the target object. Therefore, the looking patterns obtained in a preferential looking paradigm are not only affected by the featural distance between the correct label and its mispronunciation but also by properties of the distractor object (e.g., the familiarity of the object or whether the child knows the distractor label).

In the current study, degree of pupil dilation in response to a single picture was measured, avoiding the need to present a distractor. Thus by eliminating the competition between two potential referents of the label, a potentially more sensitive measure of the effect of featural distance may be obtained. Pupillometry is a method highly suitable for assessing the performance of young children, being based on an involuntary psycho-sensory reflex, i.e., pupil dilation (Loewenfeld, 1993). In pupillometry, instead of recording the pattern of gaze fixations, the eye tracking equipment is used to measure change in pupil size over time. Increased pupil dilation in young children has been interpreted to be a proxy of surprise, novelty, and cognitive effort (for a review, see Karatekin, 2007). More recently, pupil-

lometry has been found to be a viable tool in language research. It was shown to be sensitive to detecting acoustic (dis-)similarity (Hochmann & Papeo, 2014), semantic mismatch (Kuipers & Thierry, 2011), and – most important for the current study – mispronunciations (Fritzsche & Höhle, 2015).

A number of aspects of pupillometry make it an especially appealing tool for language acquisition research. First, pupillometry is minimally demanding. The passive processing of the experimental stimuli (i.e., watching while listening) does not necessitate an overt behavioral response. In our study, processes unrelated to the investigated phenomenon – recognition of distractors, memory requirements, evaluation and decision processes – are greatly reduced. Second, pupil dilation is a continuous response, which may provide a more appropriate way to investigate children's reaction to degrees of mispronunciation than a pseudo-categorical response employed by preferential looking paradigms (i.e., looking at either the target or the distractor image) (Klingner, 2010b). Third, pupillometry is inexpensive and easy to learn. Although electro-physiological and brain-imaging techniques generally avoid the shortcomings that looking time paradigms introduce, such techniques require specialized equipment and expertise thereof. The eye tracking equipment needed for pupillometry is already widely used in the child language research community and the technical competencies required for pupillometry can be readily acquired by those already familiar with the equipment. Due to these properties of pupillometry, it may be the case that pupil dilation more directly reflects the costs induced by processing a mispronounced word than a measure of looking time within a preferential looking methodology, and therefore may provide a more fine-grained insight into the effect of the degree of mismatch between the correct and the mispronounced form.

Furthermore, in contrast to previous studies, our study employs a featurally balanced consonantal set, additionally cross-balanced for feature type and change. Words with a diverse set of initial consonants were selected, then systematically manipulated not just by the number, but also by the type and direction of feature changes (see Table 2.1). Therefore, we chose to investigate 30-month-old children whose lexicons allow for the creation of a more diverse and balanced stimulus set than would be possible for younger children (early lexicons tending to contain predominantly

labial-initial words, c.f., Vihman & Croft, 2007).

Table 2.1: Stimulus list, organized by condition, noted with IPA (Correct = correctly pronounced onset, Δ1F = one-feature change, Δ2F, two-feature change, Δ3F = three-feature change).

Word (*English*)	**Correct**	**Δ1F**	**Δ2F**	**Δ3F**
Baby (*baby*)	b	d	f	ʃ
Bett (*bed*)	b	p	k	ʃ
Boot (*boat*)	b	d	z	ʃ
Buch (*book*)	b	v	f	ʃ
Decke (*blanket*)	d	t	v	f
Dusche (*shower*)	d	t	p	f
Fahne (*flag*)	f	v	t	d
Fisch (*fish*)	f	p	z	g
Fuß(*foot*)	f	p	b	g
Kaffee (*coffee*)	k	t	ʃ	v
Kamm (*comb*)	k	p	f	v
Käse (*cheese*)	k	g	b	v
Pony (*pony*)	p	t	v	z
Schaf (*sheep*)	ʃ	t	d	g
Schere (*scissors*)	ʃ	t	d	g
Teddy (*Teddy bear*)	t	p	b	v
Tisch (*table*)	t	d	b	v
Sofa (*sofa*)	z	v	b	p
Sonne (*sun*)	z	d	f	p
Suppe (*soup*)	z	d	t	k

The current study attempts to determine whether pupillometry can be used to obtain a gradient measure of lexico-phonological knowledge. Specifically, we test for the following effects. *Effect of mispronunciation*:

Table 2.2: List of fillers.

Word (*English*)
Adler (*eagle*)
Birne (*pear*)
Ente (*duck*)
Finger (*ibid.*)
Fuchs (*fox*)
Hemd (*shirt*)
Herz (*heart*)
Hund (*dog*)
Korb (*basket*)
Lampe (*lamp*)
Mantel (*coat*)
Mond (*moon*)
Mund (*mouth*)
Pilz (*mushroom*)
Pinsel (*brush*)
Schachtel (*box*)
Torte (*cake*)
Weste (*vest*)
Wolke (*cloud*)
Zebra (*ibid.*)

The degree of pupil dilation is larger in the mispronounced conditions than in the correct condition. This may indicate that mispronounced labels are harder to match and process along with the activated representation than correctly produced labels (Fritzsche & Höhle, 2015). *Effect of featural distance*: If, in addition to the effect of mispronunciation, the degree of pupil dilation is predicted by the number of featural changes made, this result

would provide evidence that the degree of mispronunciation modulates lexical processing.

2.2 Method

2.2.1 Participants

Forty-eight 30-month-old monolingual German children (26 girls) were recruited (*M = 30, SD = 0.56*) from the BabyLAB Participant Pool at the University of Potsdam. Caregivers reported no developmental and sensory disabilities. We assessed the children's vocabulary knowledge and familiarity with the experimental items using the parental report measure FRAKIS (i.e., the German adaptation of the MacArthur-Bates CDI, c.f., Szagun, Schramm, & Stumper, 2009). Participants were reported to be familiar with the majority of (correct) experimental items ($M = 79.9\%$, *SD = 16.9*). The children's reported average vocabulary ($M = 410$; $SD = 112$) aligned closely with FRAKIS norms for 30-month-old German-speaking children ($M = 439$, Szagun et al., 2009). Five children were excluded from the analyses due to providing insufficient data (see Results).

2.2.2 Stimuli

In order to identify words likely to be known by toddlers, 20 easily depictable words with CVC or CVCV syllable structure and word-initial stress were selected from FRAKIS (Szagun et al., 2009). Word onsets were manipulated to create four conditions: 20 correct (unchanged) items (e.g., *Schaf*, [ʃaːf], 'sheep'); 20 items with one feature change (e.g., [taːf], manner of articulation change); 20 items with two (e.g., [daːf], manner of articulation and voicing change); and 20 items with three (e.g., [gaːf], manner of articulation, voicing, and place of articulation change). Mispronunciations resulted in non-words for the children.[2] Type (i.e., voice, manner, place) and direction of feature change were counterbalanced. From each word, three mispronunciations were created by either changing one, two, or three features. These phonologically related items (e.g., [ʃaːf],

[2]Two real words produced by the manipulation (*Kuppe*, 'knoll', and *Wisch*, 'note') are unlikely to form part of the children's lexicon. Re-analyses with the exclusion of those two items yielded the same significant contrasts as in the original analyses.

'sheep', [taːf], [daːf], [gaːf]) formed an item family. Forty additional easily depictable words from FRAKIS were included: 20 filler items that were always produced correctly, and 20 items related to another study (20 items with onset clusters, to be reported in Chapter 4. Altogether, participants were presented with 35 correctly and 25 incorrectly pronounced items in each version of the experiment. The experimental stimuli are listed in Table 2.1 and the fillers are included in Table 2.1. Easily recognizable color drawings depicting a referent of the original word were converted to a similar size (approximately 200 x 200 pixels displayed in a 300 x 300 pixel area). Four versions of the task were created, each item family occurring once in each version with the four conditions counterbalanced across the four versions; children never saw the same picture or heard the same label more than once.

We controlled for luminance in various ways. First, all depicted objects were visually adjusted to be of equal size and placed in front of a white 300 x 300 pixel background. This image filled 7% of the 1280 x 1024 pixel screen. Luminance measurements in Adobe Photoshop CS6 for each pixel of each image (values range from 0 to 255) showed that the images had a mean luminance of 232 (*range*: 200 – 249, *SD*: 13.4). More than half of the pixels were white (i.e., the background color) in all images such that the median luminance was 255 for each one of them. Individual luminance values for each stimulus are provided in Table 3. Second, the rest of the screen was uniformly set to gray (RGB value: 179, 179, 179) for all stimuli, which resulted in the majority of the screen (93%) having an identical luminance value throughout the trials. Third, the eye tracking calibration period (30 seconds) provided ample time for the participants' eyes to adjust to the ambient light. Furthermore, an adaptation period before the measurement was part of each trial. Prior to the critical word presentation images were shown in silence for 1000 ms. The last 100 ms were used as baseline for adjusting the following data points. Pupil dilation latencies to light are reported to vary between 150 and 400 ms for control participants (c.f., p. 435 in Holmqvist et al., 2011). Considering the luminance of the screen (image + background), the difference between the brightest and darkest image as measured by Photoshop lies at 3.36, which amounts to about 1.3% of the range of possible values (i.e., completely black to completely white). Ambient luminance in the testing room was

not constant across participants although it was within each participant. Natural light was blocked during the test and a fluorescent lamp provided light which was dimmed to a comfortable level for the participant. Apart from controlling the visual stimuli in the way outlined above, no other corrections were performed.

Table 2.3: Luminance values of experimental items.

Picture	Mean	SD	Median
Schaf	244.72	34.16	255
Sofa	199.97	79.84	255
Sonne	231.64	37.02	255
Suppe	244.3	29.68	255
Tisch	228.38	57.20	255
Boot	226.22	58.66	255
Teddy	215.59	73.15	255
Dusche	232.36	53.51	255
Decke	231.93	37.03	255
Fuß	246.53	25.00	255
Fisch	241.42	39.33	255
Fahne	239.11	50.06	255
Bett	225.55	59.40	255
Pony	214.86	74.98	255
Kaffee	231.85	54.76	255
Buch	209.28	78.58	255
Baby	234.5	53.17	255
Schere	248.41	29.14	255
Kamm	245.47	38.76	255
Käse	248.86	16.43	255

2.2.3 Procedure

Children were told that they were to watch a short movie, during which they should sit still and as a reward they could choose a book afterwards. After obtaining assent from the children and written informed consent from the caregiver, children were seated in their caregiver's lap and positioned such that their eyes were approximately 60 cm from the computer screen. Their pupil sizes were monitored by a Tobii 1750 corneal reflection eye tracker (temporal resolution: 50 Hz, spatial accuracy: .5' to 1', recovery time after track loss: 100 ms). All visual stimuli were shown centrally on a 17" (1280 x 1024) TFT screen with a size of 300 x 300 pixels forming a horizontal and vertical viewing angle of 7.4°. The experiment started following the calibration period (five screen positions, ∼30 seconds).

In each trial, a picture was presented and remained on screen for the duration of the trial (four seconds). One second after the picture appeared, the corresponding (correctly or incorrectly produced) auditory label was played. The critical window of analysis was the three-second interval following the onset of the auditory stimulus. The experiment encompassed 12 blocks, each containing five trials (altogether 12 x 5 = 60 trials, of which 20 fillers and 20 unrelated). Before each block, an 'attention-getter' was presented (a short silent movie clip of animated cartoon characters and animals). The attention-getters were played in a loop until the experimenter pressed a key to start the next block. On average, the experiment lasted 15 minutes.

After the experiment, caregivers were asked to complete a questionnaire in order to estimate the child's vocabulary size and their familiarity with the experimental words. The questionnaire comprised the 600 FRAKIS items (Szagun et al., 2009), plus 12 additional items relevant to an experiment not reported here. On average, the questionnaire took 20 minutes to complete.

2.3 Results

We transformed the Tobii output (T1750, ClearView) files to matrices to be analyzed by R (version 3.1.0, R Core Team, 2014). The pupil data consisted of the estimated absolute mean diameter in mm for each data point (approximately every 20 ms) over the period of a trial. Sudden

brief changes in pupil diameter (more than 0.05 mm in 20 ms) that are considered to be artefacts produced by the eye tracker were excluded from further analyses (see Appendix for more information). Missing points were linearly interpolated if the interval missing was not more than 400 ms (the maximum duration of typical blinks, Beatty & Lucero-Wagoner, 2000). Afterwards, left and right pupil size values were averaged following Fritzsche and Höhle (2015). The overall correlation between left and right pupil size was high for all participants ($M = .95$, $SD = 0.03$).

In order to ensure that the words used in the experiment were part of the child's lexical inventory, only those trials that included words (and their mispronunciations) reported to be known in the parental questionnaire were considered in the analysis of each individual child's data. Successful trials were defined as those containing pupil measures from at least half the length of the trial. Based on this criterion, the proportion of successful trials was tabulated for each participant. Those participants who did not reach a threshold of 50% of successful trials (following Fritzsche & Höhle, 2015) were excluded from further analyses (5 participants). The mean number of successful trials was 17.19 out of 20 ($SD = 1.89$) in the experimental trials and 17.38 out of 20 ($SD = 1.94$) in the filler trials. The mean number of successful trials per experimental condition was 4.30 out of 5 ($SD = .10$).

Since variations of the pictures' luminance values, the ambient light, and individual differences affect pupil size (Beatty & Lucero-Wagoner, 2000), mean pupil dilation was calculated on a trial-wise basis, i.e., each trial served as its own baseline. This was possible because in the first second of each trial, when the picture was presented in silence, participants' eyes adjusted for that particular luminance (Beatty & Lucero-Wagoner, 2000). Specifically, we corrected for inter-subject and inter-trial variation by subtracting a silent baseline value (i.e., a mean value of a 100 ms interval before the onset of the auditory label). For this reason, we did not collect individual stimulus and ambient luminance values. Trials with no data points in the baseline interval were excluded from further analyses (1.7% of trials). Manipulating the duration of the baseline interval (20 ms and 500 ms) did not significantly affect the results.

We employed linear mixed effects models with random intercepts and slopes using the `lmer` function (estimates were chosen to optimize the

log-likelihood criterion) in the `lme4 R` package (Bates et al., 2014). The linear mixed effects models were built so that their random structure was maximally specified (Barr, Levy, Scheepers, & Tily, 2013; Jaeger, Graff, Croft, & Pontillo, 2011). Each intercept and slope fitted by the model was adjusted by the effect of condition and neighborhood density nested in participants and by vocabulary size nested in items. Due to the possibility of overfitting and hence producing convergence errors, the model could only be computed when vocabulary size in the random structure was dichotomized. Since the Helmert-coded levels of `featural distance` were collinear (as they should be, being nested within each other), the correlation term in the random effect structure in `featural distance` was removed (Jaeger et al., 2011).

The most parsimonious model was chosen through comparisons using Likelihood Ratio Tests (Pinheiro & Bates, 2000) via the `anova` function from the `stats` package (R Core Team, 2014). `Featural distance`, a within-subject factor with four levels was entered as a fixed effect into the model: correctly pronounced as well as mispronounced with one-, two- and three-feature change. We also included potentially confounding (sub-)lexical factors as control variables (word frequency, neighborhood density, and transitional probability, taken from the Clearpond database: Marian, Bartolotti, Chabal, & Shook, 2012), and children's vocabulary size estimated from the parental questionnaire. Participants and items were entered as random effects into the model. Mean change in pupil diameter (i.e., the mean value extracted from each three-second window of analysis, starting from presentation of the critical stimulus) was used as the outcome measure. The most parsimonious model contained `featural distance`, `vocabulary size` and `neighborhood density` as fixed effects.

Mean pupil dilation in each condition is presented in Figure 2.1 (bar plot) and Figure 2.2 (time-course plot). Visual inspection suggested that correctly pronounced words were generally associated with smaller pupil size change than mispronounced words. Also for mispronounced words, the one-feature change condition was associated with a smaller degree of pupil dilation than the two- and three-feature change conditions. There seemed to be no difference between the conditions with two and three feature changes. Statistical analysis using the mixed effects model described above confirmed these observations. In the model, `featural distance` signifi-

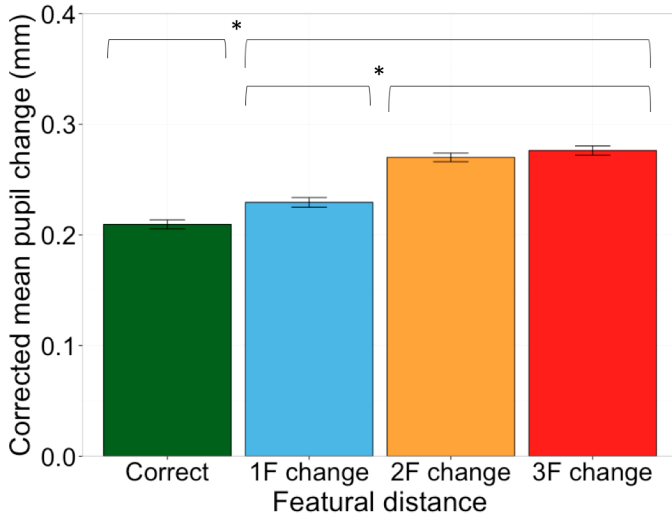

Figure 2.1: Mean pupil size change in response to differing degrees of mispronunciation. Significant contrasts ($t > 1.96$) between the correct vs. the mispronounced items and between the one- vs. two- and three-feature changes are marked with asterisks. Error bars represent the standard error built around the mean.

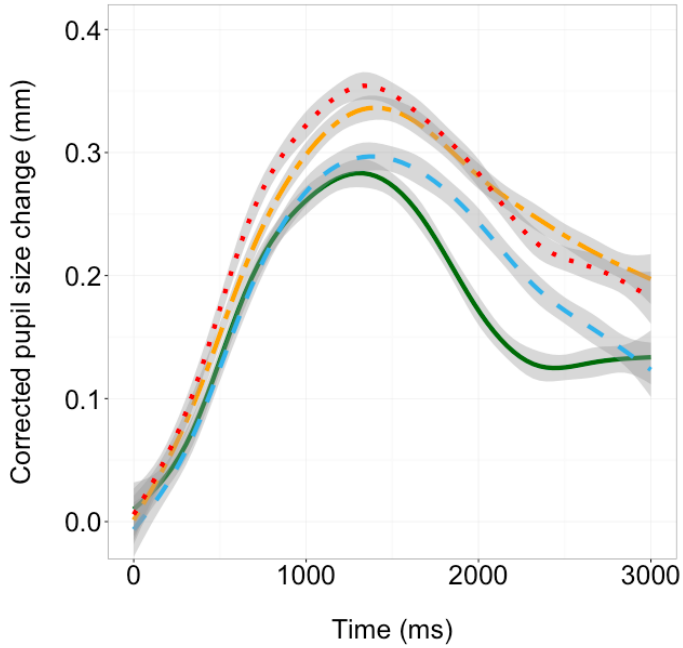

Figure 2.2: Mean pupil size change over time in response to differing degrees of mispronunciation (correct: solid green, 1F change: dashed blue, 2F change: dotdash orange, 3F change: dotted red). A 95% confidence interval was built around the fitted values, shown with gray shading.

cantly contributed to model fit ($\chi^2(3) = 10.19$, $p < .017$). We subsequently found the correct vs. mispronounced contrast and the one-feature change vs. two- and three-feature change contrast to be significant ($\beta_1 = 0.041$, $SE = 0.019$, $t = 2.21$, $\beta_2 = 0.044$, $SE = 0.020$, $t = 2.39$). The contrast between the two- and three-feature change conditions was not significant ($\beta_3 = 0.005$, $SE = 0.023$, $t = 0.21$).

Apart from analyzing the mean pupil size change, we also assessed peak dilation. The raw data points were fitted with the `smooth.spline` function from the `stats R` package (R Core Team, 2014). Further details on how peak values were calculated can be found in the Appendix. Analyses parallel to those performed with mean dilation determined that mean and peak dilation measures were in agreement by showing the same tendencies ($\beta_1 = 0.036$, $SE = 0.021$, $t = 1.73$, $\beta_2 = 0.041$, $SE = 0.022$, $t = 1.87$, $\beta_3 = 0.021$, $SE = 0.029$, $t = 0.72$).

Time-course analyses (post-hoc cluster-based permutation tests: Maris & Oostenveld, 2007) explored when significant differences emerged between each condition pair (using the `eyetrackingR` package: Dink & Ferguson, 2016). First, individual paired sample t-tests found the significant ($p < .05$) t-values across the whole time frame. Second, clusters (e.g., contiguous significant t-values) were identified, for which a cluster-level t-value was calculated as the sum of all single sample t-values within the cluster. Third, the significance of cluster-level t-values were assessed by generating Monte Carlo distributions ($N = 2000$) thereof and determining the probability of their occurrence given the distribution. Those clusters whose t statistic exceeded the threshold ($t = 2.64$, Bonferroni-corrected for multiple comparisons) were then tabulated for each contrast. With this method, using the `time_cluster_data` function, significant contrasts were identified between all conditions expect in the two-feature change vs. three-feature change contrast (c.f., Table 2.4). Concerning the contrasts between correct and two-feature change as well as correct and three-feature change, marginally significant time intervals were obtained in addition to the significant time intervals (c.f., Table 2.4). Comparable results (i.e., significant contrasts across all condition pairs except the two-feature change vs. three-feature change) were obtained when the function `time_cluster_data` was supplied with a formula containing a linear mixed effects model.

Table 2.4: Significant contrasts across conditions in time-course analyses. Interval = time interval in the naming phase, $\sum t$ = cluster-level t value, p = p value associated with cluster-level t, Corr. = correctly pronounced familiar label, Δ1F = one-feature change, Δ2F, two-feature change, Δ3F = three-feature change introduced to the onset.

Contrasts	Interval (ms)	$\sum t$	p
Corr. vs. Δ1F	1700–2300	10.78	*
Corr. vs. Δ2F	1200–1300	1.53	†
	1500–2900	35.94	**
Corr. vs. Δ3F	400–700	1.53	†
	1700–2400	12.48	*
Δ1F vs. Δ2F	2100–2900	16.21	**
Δ1F vs. Δ3F	300–500	3.07	*
Δ2F vs. Δ3F	–	–	n.s.

†: $p < .1$, *: $p < .05$, **: $p < .01$, n.s.: not significant

To determine whether filler items were treated similarly to the correctly pronounced experimental items in the study, a separate analysis was carried out on a restricted data set containing only those two levels of condition. The condition (correct vs. filler) variable was then sum-coded and used as a fixed effect in a linear mixed effects model, which was otherwise identical to the ones run in previous analyses. As expected, no reliable difference was detected between pupillary responses given to filler items and those of correct experimental items indicated by the non-significant likelihood ratio test statistic comparing the null model and the model containing the fixed effect ($\chi^2 = 0.028$, $p = .87$).

2.4 Discussion

Our findings indicate that pupillometry is a viable method in lexical representation research as it can capture a differential response for correctly pronounced and mispronounced items through the measurement of the degree of pupil dilation change in toddlers. First, the significant differences in mean pupil dilation between correct and mispronounced items replicate previous findings (Fritzsche & Höhle, 2015) and provide further evidence that the processing of mispronounced words leads to greater pupil dilation compared to correctly produced words. Second, the results indicate that the degree of mispronunciation also influenced the pupillary response, that is, conditions that involved more than one feature change were associated with larger pupil dilation than those with only one feature change. This result indicates that children's lexical processing is modulated by the degree of mismatch. The fewer features are shared between the correct and the mispronounced form, the harder it is to match the stimulus with the lexical representation, which is reflected by larger degrees of pupil dilation. It is possible that the effect found here relates to general surprise caused by hearing a sound sequence which was not expected as a target label. However, we would argue that a form-related explanation is more plausible, that is that the pupillary response is an indicator of cognitive effort to establish a link between stimulus and lexical representation. This argument is supported by the findings of Fritzsche and Höhle (2015), a single-picture study with children of the same age, whereby pupillary responses given to correct and semantically unrelated labels were comparable.

Results obtained from the time-course analyses are consistent with the ones obtained by linear mixed effects models. Significant differences between all levels of `featural distance` were found except the two-feature change vs. three-feature change. The correct vs. one-feature change contrast emerged in a relatively late time window in comparison to other larger contrasts of featural distance. This result suggest that when processing a one-feature change difference, the amount of cognitive resources that are recruited to activate the corresponding lexical representation is initially quite similar to the amount of resources needed to process a correct pronunciation. A similar argument can be said about the other one-feature change difference contrast, the one- vs. two-feature change contrast emerging relatively late.

However, the results do not support complete graded sensitivity as the three-feature change condition yielded a pupil size change that was not significantly larger than that of the two-feature change. This latter result is fully consistent with previous findings from preferential looking studies (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008). Recall that these studies demonstrated graded sensitivity, and no significant difference between two- and three-feature deviations. It is possible that maximal pupil dilation was achieved with two-feature deviation and an additional feature change could no longer be reflected through a larger degree of dilation. From a physiological standpoint, however, larger degrees of pupil dilation have been achieved than those reported here using other non-linguistic tasks with similar age groups. This suggests that a ceiling effect is an unlikely explanation for the findings (for an overview on task-evoked pupillary response see Beatty & Lucero-Wagoner, 2000).

Another possibility for the lack of difference between the two- and three-feature-change conditions is that individual feature changes may interact with one another and/or that the effects of featural combinations may not be linearly additive (i.e. the difference between the two- and three-feature changes may not be proportional to the difference between the one- and two-feature changes). Further investigation is needed to explore the unique effect of type and direction of feature changes in consonants.

In accordance with past studies (Ren & Morgan, 2011; White & Morgan, 2008), we defined degree of mismatch between correct and deviant form in terms of phonological features. However, sounds that differ in terms of phonological features also differ acoustically. It is unclear to what extent the degree of acoustic difference between the correct and the incorrect form correlates with the degree of featural distance. It is possible that (as argued by Mani & Plunkett, 2011a for vowels) the pattern of graded sensitivity found here may relate more to acoustic rather than phonological properties. Further research is required to address the question of whether mismatch detection is based on physical acoustic or phonological distance. These alternatives can be tested by quantifying acoustic distance and assessing whether acoustic and/or phonological distance significantly and independently account for the results (for an analysis with vowels, see Mani & Plunkett, 2011a).

Our study is the first demonstration of children's sensitivity to the degree of mispronunciation in a passive listening task using pupillometry. Our results demonstrate that young children are sensitive to the contrast between small (one-) and large (two- and three-) feature changes. These results corroborate previous research that found toddlers' word recognition to be modulated by featural changes (Fikkert, 2010; Swingley & Aslin, 2000; Yoshida et al., 2009) as well as by degree of mismatch caused by such changes (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008), suggesting that early lexical representations encode subphonemic detail.

In summary, this study demonstrates that pupillometry can be used as a tool for mispronunciation detection with 30-month-old children, providing a minimally demanding alternative to other extensively applied paradigms. It therefore proves to be a readily available, low cost and reliable method with which to conduct speech processing research with infants and young children. As such, pupillometry holds promise to accelerate the rate of new discovery in this important field.

Chapter 3

Children's sensitivity to sub-phonemic detail: Enriching the preferential looking paradigm with pupillometry

Abstract

This paper offers a novel approach to the study of lexical development by complementing the intermodal preferential looking paradigm with a measure automatically collected using such a paradigm: pupil dilation. Given that the magnitude of pupil dilation reflects cognitive effort in all age groups including infants, pupillometry is an ideal method to assess early lexical knowledge. Based on past findings, it was predicted that both children's looking behavior and pupillary response were sensitive to phonological overlap. We manipulated degree of mismatch by introducing featural changes to the label of the target referent and adding a novel label that was related to the distractor referent. Both measures, children's looking behavior and pupillary response, were sensitive to phonological overlap, corroborating previous studies that found gradient response in one or the other measure. Moreover, time-course analyses have shown for the first time that large featural change was associated with oscillating looking behavior (shifting between target and distractor preference). Time-course analyses of looking behavior also detected complete gradient sensitivity to degree of mismatch. These findings further support the notion that early words are represented in great detail, containing subphonemic information.
Keywords: *lexical development; featural distance; mispronunciation detection; eye tracking; pupillometry.*

3.1 Introduction

During language acquisition, infants are required to identify the building blocks of the ambient language by segmenting the auditory input and categorizing the resulting sounds into discrete groups of phonemes. This process of phoneme categorization is intricately intertwined with both word learning and word recognition processes. For instance, the ability to detect a small yet contrastive change is critical to building up an adult-like lexicon that contains minimal pairs, i.e., words that differ by a single feature or phoneme (e.g., *tap* and *cap*).

One approach to test children's lexical knowledge is by presenting them with correctly pronounced and featurally manipulated (that is, mispronounced) meaningless word forms (e.g., *tap* and *dap*). Studies showed that

by 24 months of age, children develop the skill to reliably differentiate the correct from the mispronounced word form. Differential response to correct vs. mispronounced labels has been achieved with a variety of methods appropriate for testing young children (i.e., target preference with correct target labels in intermodal preferential looking paradigms: Arias-Trejo & Plunkett, 2010; Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Durrant, Delle Luche, Cattani, & Floccia, 2015; Höhle, van de Vijver, & Weissenborn, 2006; Mani et al., 2008; Mani & Plunkett, 2007, 2010a, 2010b, 2011a, 2011b; Ramon-Casas, Swingley, Sebastián-Gallés, & Bosch, 2009; Ren & Morgan, 2011; Swingley, 2003; Swingley & Aslin, 2000, 2002; White & Morgan, 2008; White et al., 2005; dishabituation to mispronounced labels in head-turn preference or habituation paradigms: Fennell & Werker, 2003; Fikkert, 2010; Swingley, 2005; Vihman & Croft, 2007; Werker, Fennell, Corcoran, & Stager, 2002; Yoshida, Fennell, Swingley, & Werker, 2009; difference in event-related brain potential [ERP] signature given to correct vs. mispronounced labels in a single-picture paradigm: Mani, Mills, & Plunkett, 2012; greater pupil dilation in response to mispronunciation in single-picture pupillometry paradigms: Fritzsche & Höhle, 2015; Tamási, McKean, Gafos, Fritzsche, & Höhle, in press; Tamási, McKean, Gafos, & Höhle, 2016b). These findings suggest that early lexical representations are highly specific, that is, sufficiently specified so as to enable establishing a match with the correct, but not with the mispronounced label.

Whether infants are sensitive to differing degrees of mismatch – phonological overlap as measured by featural distance – between the heard label and the lexical entry is not so well established given previous findings. On the one hand, such ability has been demonstrated with intermodal preferential looking paradigms. In those studies, degree of phonological overlap positively predicted the proportion of target looks (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005). Infants' gradient sensitivity to mispronunciation would indicate that early lexical representations are not only specific, but also detailed and fine-grained (i.e., encoding the degree of overlap with other words on a sub-phonemic level). On the other hand, lack of sensitivity to degree of mismatch has also been found. That is, looking behavior was not linked to the degree of phonological overlap (Bailey & Plunkett, 2002; Swingley

& Aslin, 2002). White and Morgan (2008) attributed this null result to using familiar distractor pictures with labels known to children (therefore, in any given trial, both presented images had known labels). Accordingly, recent studies that did detect gradient sensitivity used distractor pictures depicting an item unfamiliar to infants (e.g., a French horn), which – via mutual exclusivity, a word-learning mechanism (Halberda, 2003) – enabled matching the mispronounced label with the distractor and thus allowed children's gradient sensitivity surface (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005).

As seen above, the most popular paradigm to assess children's lexical knowledge is intermodal preferential looking, typically conducted with an eye tracker (for a recent overview, see Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). Even though paradigms involving eye tracking yield a valuable body of data, only a fraction thereof are routinely considered and analyzed in language studies. The most widely reported measure is the overall proportion of target looking time (Golinkoff et al., 2013).

This study extends the methodology of intermodal preferential looking paradigms by complementing it with a measure automatically collected via the eye tracker: pupil dilation. As an early psycho-sensory reflex, greater degree of pupil dilation in children has been linked to cognitive effort, violation of expectation, and novelty (Karatekin, 2007), making it an appealing tool for probing infant knowledge. Recently, pupillometry has proven to be a promising method in infant language research as it has detected children's sensitivity to acoustic (dis-)similarity (Hochmann & Papeo, 2014), semantic incongruity (Kuipers & Thierry, 2011, 2013), and – most crucial for the current study – featural manipulations resulting in mispronunciations (Fritzsche & Höhle, 2015; Tamási et al., in press, 2016b).

Using single-picture pupillometry paradigms – presenting a single visual stimulus per trial –, thirty-month-old children have been shown to give a differential pupillary response to correctly pronounced labels vs. their mispronunciations: the general finding being that mispronounced labels were associated with larger degrees of pupil dilation than correct labels (Fritzsche & Höhle, 2015; Tamási et al., 2016b). This asymmetry was interpreted such that more cognitive effort was needed to establish the link between the mispronounced label and the picture (in order to re-

construct the correct phonological form and map onto the corresponding lexical representation) than doing so with the correct label. Such finding and interpretation are consistent with those of earlier studies that demonstrated the specificity of lexical representations with other methodologies (c.f., second paragraph of this section).

Using a similar paradigm of single-picture pupillometry, children from the same age group were shown to be sensitive to the degree of mispronunciation based on their pupil dilation patterns (Tamási et al., in press, presented in Chapter 2). The degree of mismatch between the correct and mispronounced form – manipulated by changing the number of feature changes – positively predicted the degree of pupil dilation (i.e., the more feature changes were introduced to the label, the greater pupil dilation resulted). This finding is again in line with intermodal preferential looking studies that demonstrated gradient sensitivity to the degree of mismatch and thus indicating early lexical representations to be fine-grained (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005).

In the current study, the objective was to test whether looking time and pupil dilation measures can be aligned such that both are sensitive to degree of mismatch. We hypothesized that the amount of cognitive effort required to activate the corresponding lexical representation is affected by the degree of mismatch in the label. Low cognitive effort is expected to manifest in easy and fast identification of the target image and relatively low pupil dilation. Increasing the amount of cognitive resources needed to solve the task by manipulating degree of mismatch is predicted to make target identification harder and slower and the pupillary response more pronounced.

Following past intermodal preferential looking studies, the degree of mispronunciation was manipulated by featural distance (0–3 feature changes to the correct label and a semantically and phonologically unrelated label, e.g., [b]*aby*, correct / [d]*aby*, Δ1F / [f]*aby*, Δ2F / [ʃ]*aby*, Δ3F / *sushi*, novel). While children were presented with familiar target and novel distractor referents and the auditory label, their looks and pupillary responses were monitored. First, children's increased looks towards the distractor image as a function of degree of mispronunciation would show their tendency to associate the label with the novel distractor instead of

the familiar target (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005). Second, extrapolating from the findings of past studies using single-picture pupillometry paradigms, mispronunciation was expected to increase the effort of recognizing the heard label and integrating it with the target image and the corresponding lexical entry, resulting in larger degrees of pupil dilation (Fritzsche & Höhle, 2015; Tamási et al., in press, 2016b). Critically for the current study, we expect degree of phonological overlap to be a predictor of both looking behavior (given the findings from intermodal preferential looking paradigms: Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005) as well as that of pupillary response (based on the findings of a single-picture pupillometry study: Tamási et al., in press, presented in Chapter 2).

3.2 Method

3.2.1 Participants

Fifty-nine thirty-month-old children (M = 30 months 7 days, SD = 16 days, 32 boys), all monolingual speakers of German, were recruited from the BabyLAB Participant Pool at the University of Potsdam. Caregivers reported no sensory and developmental disorders. Children's vocabulary knowledge and familiarity with the experimental items was assessed using the vocabulary list FRAKIS (i.e., the German adaptation of the MacArthur-Bates Communicative Development Inventory: Szagun, Schramm, & Stumper, 2009) and the vocabulary list including novel items. The children's reported average vocabulary (M = 451.1; SD = 91.9) aligned closely with FRAKIS norms of German-speaking children of the same age (M = 439, Szagun et al., 2009).

3.2.2 Stimuli

A total of 20 experimental words over 20 trials, either mono- or disyllabic, (and 20 other words related to another study to be reported in Chapter 4) were selected and produced by a German native speaker in an enthusiastic, child-directed manner (listed in Table 3.1). 15 of the experimental words

were familiar labels (taken from the German adaptation of the MacArthur-Bates CDI: Szagun et al., 2009) and 5 were unfamiliar (whose referents are presumably not known to thirty-month-old children).

Degree of mispronunciation in familiar word onsets was manipulated so as to create four conditions. In each version of the task, six correct (unchanged) items (e.g., *Bett*, [bɛt], 'bed'); three items with one-feature change (e.g., [pɛt], voicing change); three items with two-feature change (e.g., [kɛt], voicing and place of articulation change); and three items with three-feature change (e.g., [ʃɛt], voicing, place of articulation, and manner of articulation change). The proportion of correctly vs. incorrectly pronounced labels were chosen to be as balanced as possible, see the Procedure section. Each manipulation constituted voicing, place of articulation, or manner of articulation change made to the label onset (counterbalanced in one- and two-feature change conditions). Direction of feature change (voiceless vs. voiced, labial vs. coronal vs. dorsal, stop vs. fricative) was also counterbalanced. Mispronunciations resulted in non-words for the children.[1] Novel word onsets were always presented unchanged.

Easily recognizable color drawings depicting the referents of the experimental items were selected and converted to a similar size (approximately 200 x 200 pixels displayed in a 300 x 300 pixel area). The areas of interest included the 400 x 400 pixel frame around each picture. Additional pictures, 15 novel and 5 familiar pictured referents, were chosen. These pictures were paired with labeled pictures and thus they themselves were never labeled. This resulted in altogether 20 familiar-novel labeled image pairings (shown in Table 3.1). The side at which familiar and novel pictures appeared was counterbalanced.

Four versions of the task were created, each item occurring once in each version with the mispronunciation types counterbalanced across the four versions; children never saw the same picture or heard the same label more than once. Each participant was randomly assigned to one of the versions. Participants were presented with 6 correctly pronounced familiar labels, 5 correctly pronounced novel labels, and 9 incorrectly pronounced familiar labels (followed by 10 correctly and 10 incorrectly pronounced

[1]Two real words produced by the manipulation (*Kuppe*, 'knoll', and *Wisch*, 'note') are unlikely to be known by the thirty-month-olds. Re-analyses with the exclusion of those two items did not change the overall results.

Table 3.1: Stimulus list, organized by familiar-novel word pairs and condition, noted with IPA (Word: labeled = the word that was labeled during trials, Corr. = correctly pronounced label, Δ1F = one-feature change, Δ2F, two-feature change, Δ3F = three-feature change introduced to the onset, Word: not labeled = the word that was not labeled during trials. Not labeled words are given only in English.)

Labeled	**Corr.**	**Δ1F**	**Δ2F**	**Δ3F**	**Not labeled**
Familiar					**Novel**
Bett (*bed*)	b	p	k	ʃ	tapir
Boot (*boat*)	b	d	z	ʃ	American pancake
Decke (*blanket*)	d	t	v	f	magenta
Dusche (*shower*)	d	t	p	f	microscope
Fahne (*flag*)	f	v	t	d	magnet
Fisch (*fish*)	f	p	z	g	ruler
Fuß(*foot*)	f	p	b	g	tarsier
Kaffee (*coffee*)	k	t	ʃ	v	coati
Pony (*pony*)	p	t	v	z	avocado
Schaf (*sheep*)	ʃ	t	d	g	static eliminator
Teddy (*ibid.*)	t	p	b	v	eyelash curler
Tisch (*table*)	t	d	b	v	sun dial
Sofa (*sofa*)	z	v	b	p	butter curler
Sonne (*sun*)	z	d	f	p	caviar
Suppe (*soup*)	z	d	t	k	weasel
Novel					**Familiar**
Dodo (*ibid.*)	d	–	–	–	cheese
oliv (*olive*)	o	–	–	–	scissors
Säge (*saw*)	z	–	–	–	comb
Sushi (*ibid.*)	z	–	–	–	baby
Yak (*ibid.*)	j	–	–	–	book

labels that belonged to the other study) in each version of the experiment. The proportion of correctly vs. incorrectly pronounced labels (55%) was similar to that of Experiment 1 in White and Morgan (2008) that employed the same conditions as the present study.

3.2.3 Procedure

Children were told that they were to watch a short movie, during which they should sit still and as a reward they could choose a booklet afterwards. After obtaining assent from the children and written informed consent from the caregiver, children were seated in their caregiver's lap and positioned such that their eyes were approximately 60 cm from the computer screen. Their pupil sizes were monitored by a Tobii 1750 corneal reflection eye-tracker (temporal resolution: 50 Hz, spatial accuracy: .5' to 1', recovery time after track loss: 100 ms.) All visual stimuli were shown centrally on a 17" (1280 x 1024) TFT screen with a size of 850 x 300 pixels (the two 300 x 300 pixel experimental pictures were separated by a 250 x 300 pixel gray strip) forming a horizontal viewing angle of 10.5° and a vertical viewing angle of 7.4°. The experiment started following the calibration period (five screen positions, ∼30 seconds).

The window of analysis consisted of the 3000 ms interval following the onset of the auditory stimulus (i.e., the naming phase). The trials were ordered such that those from the current study were presented in the first half and the those from the other study in the second half of the experiment. The experiment encompassed 8 blocks, each containing five trials (altogether 8 x 5 = 40 trials). The order of the experimental items was furthermore pseudo-randomized such that onsets were not repeated (e.g., *Bett* and *Boot*, did not follow each other), target onsets were not repeated (e.g., *Bett* and *Doot* did not follow each other as *Boot*, the correct form of *Doot*, shares an onset with *Bett*), correctness status was not repeated more than four times (e.g., *Bett*, *Decke*, *Pony*, and *Fisch* in a row was not a possible ordering). With the aim of keeping the children engaged and conveying a sense of progress throughout the experiment, a 'progression marker' was presented before each block and after the last one (altogether nine silent movie clips, featuring nine snails that initially line up on the left and one by one crawl to the right side of the screen). The clips were played in a loop until the experimenter pressed a key to start the next

block. On average, the experiment lasted 15 minutes.

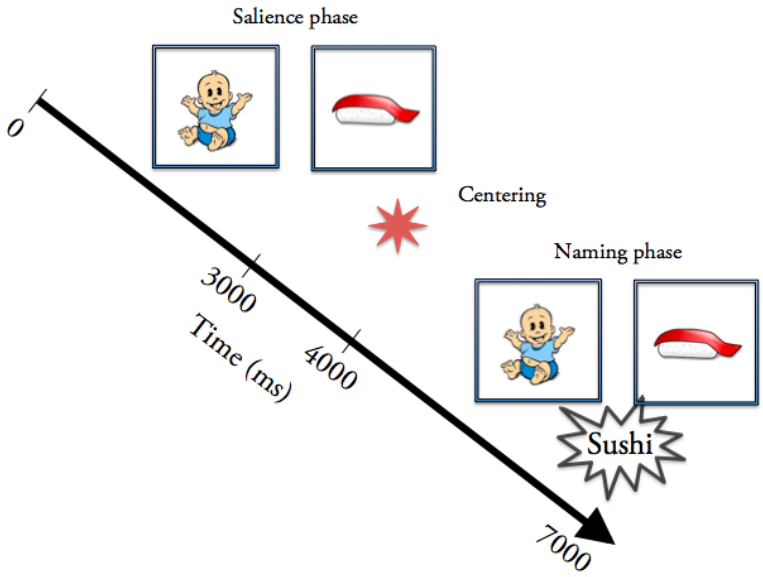

Figure 3.1: Trial structure (0–3000 ms = salience phase, 3000–4000 ms = centering, 4000–7000 ms = naming phase).

Each trial consisted of a salience phase, a centering, and a naming phase (illustrated in Figure 3.1). In the salience phase, a pair of target and distractor pictures were simultaneously presented on a gray background for 3000 ms, the target depicting a familiar referent and the distractor an unfamiliar referent. In order to reorient the children towards the center of the screen, a flashing red star was presented thereon for 1000 ms (centering). In the naming phase, the same pair of pictures as in the salience phase was accompanied by an auditory label for 3000 ms.

After the experiment, caregivers were asked to complete a questionnaire in order to estimate the child's vocabulary size and their familiarity with the experimental words. The questionnaire comprised the 600 FRAKIS items (Szagun et al., 2009), plus the labels of the purportedly unfamiliar objects. On average, the questionnaire took 20 minutes to complete.

3.3 Results

In order to ascertain that the experimental labels intended to be familiar were part of the participants' vocabulary, only words reported to be known in the parental questionnaire FRAKIS (Szagun et al., 2009) were included in the analyses ($M = 74\%$, $SD = 16.9$). Conversely, among the experimental labels that were intended to be unfamiliar (i.e., the distractor labels), only those reported as such were included (of the remaining trials: $M = 93.2\%$, $SD = 10.6$). Those participants who did not reach a threshold of 50% of successful trials (trials containing pupil measures from at least half the length of the trial, following Fritzsche & Höhle, 2015) were excluded from further analyses (eight participants). Two additional children were excluded due to providing large negative difference scores (proportion of target looks during naming phase - salience phase < -0.15) in the correct condition (following White & Morgan, 2008). On average, 88% of trials per participant were retained (35.14 / 40 trials).

The prediction that featural distance both negatively predicted target looking time and positively predicted pupil dilation was supported by observations (c.f., the bar plot summaries in Figures 3.2 and 3.3) as well as the respective analyses. In the looking time measure, linear mixed effects models were employed with random intercepts and slopes using the `lmer` function (estimates were chosen to optimize the log-likelihood criterion) in the `lme4 R` package (Bates, Maechler, Bolker, & Walker, 2014). `Degree of mispronunciation` (Correct / Δ1F / Δ2F / Δ3F / Novel), a within-subject factor was assigned a polynomial contrast[2] and was entered into the model as a fixed effect. Potentially confounding (sub-)lexical factors were included as control variables (word frequency, neighborhood density, and phonotactic probability, all calculated from the Clearpond database: Marian, Bartolotti, Chabal, & Shook, 2012), and children's vocabulary size that was estimated from the parental questionnaire. Participants ($N = 49$) and items ($N = 20$) were entered as random effects into the model. Overall proportion of looks towards the target in the naming phase – corrected for the proportion of looks in the salience phase – was used as the outcome

[2]Specifically, the first level of the contrast tested for a linear trend, the second for a quadratic trend, the third for a cubic trend, and the fourth for a quartic trend across the five conditions.

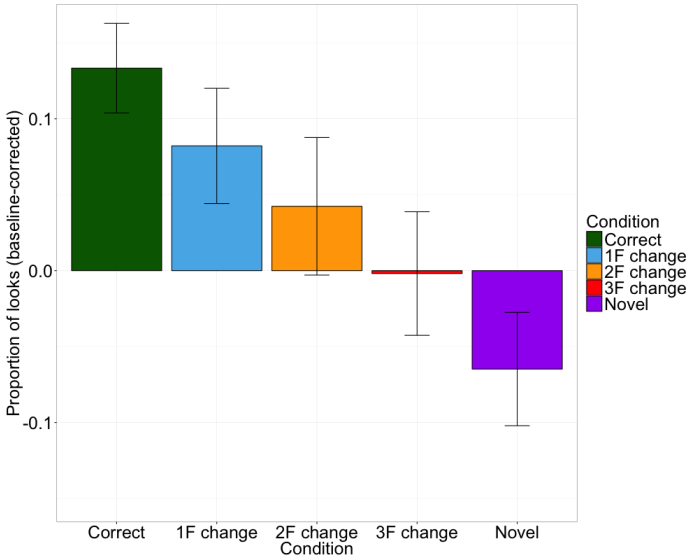

Figure 3.2: Mean proportion of looking time towards target in response to differing degrees of mispronunciation (error = SE).

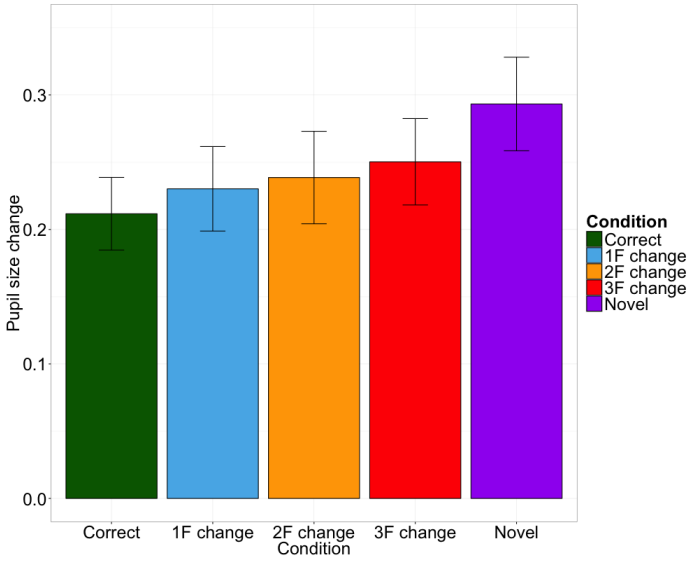

Figure 3.3: Mean pupil size change in response to differing degrees of mispronunciation (error = SE).

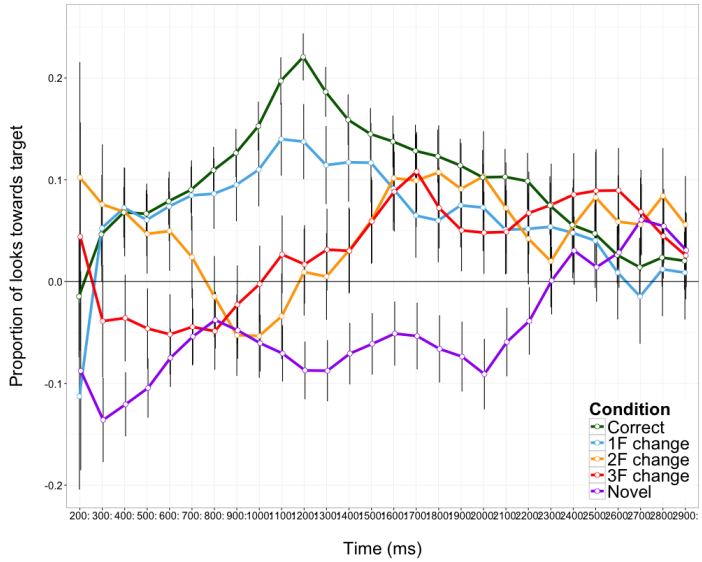

Figure 3.4: Proportion of looking time towards target over time in response to differing degrees of mispronunciation (error = SE).

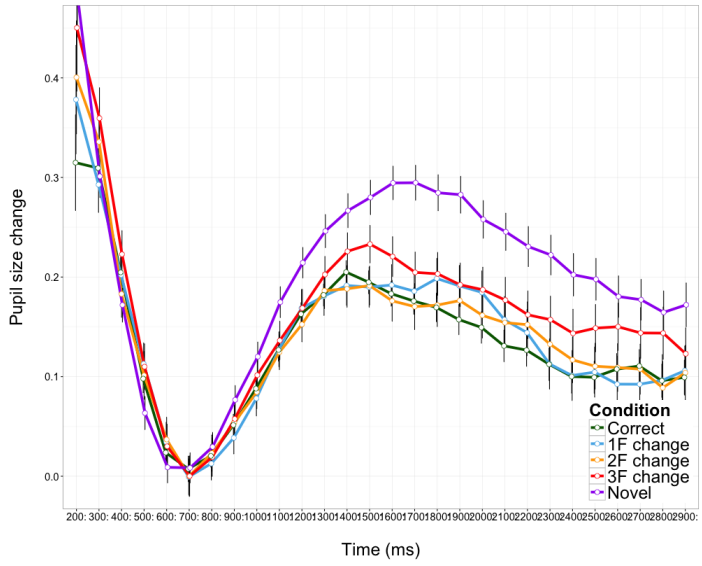

Figure 3.5: Pupil size change over time in response to differing degrees of mispronunciation (error = SE).

measure. The first 200 ms of the naming phase, i.e., the period immediately after the centering phase, did not yield sufficient data as the majority of the fixations took place centrally, a possible carry-over effect from the central fixation phase. The linear mixed effects models were built with maximally specified random structure as justified by the design (Barr et al., 2013; Jaeger et al., 2011). Each intercept and slope fitted by the model was adjusted by the effect of condition nested in participants. The most parsimonious model was chosen through comparisons using Likelihood Ratio Tests (Pinheiro, Bates, DebRoy, & Sarkar, 2007) via the `anova` function from the `stats` package (R Core Team, 2014) and contained `degree of mispronunciation` as fixed effect. In this model, a significant negative linear trend was obtained ($\beta = -0.15$, $SE = 0.04$, $t = -4.19$) in response to the degree of mispronunciation. All other trends (quadratic, cubic, quartic) were found non-significant. Phonotactic probability was found to be a marginally significant positive predictor of proportion of target looking time ($\beta = 0.02$ $SE = 0.01$, $t = 1.81$).

Post-hoc cluster-based permutation tests (Maris & Oostenveld, 2007) were employed to investigate the latency and the duration of contrasts. To test for target looking preference, time-course analyses were used to explore when looking preference significantly differed from zero in response to differing degrees of mispronunciation (c.f., Figure 3.4) using the `eyetrackingR` package (Dink & Ferguson, 2016). First, individual paired sample t-tests found the significant ($p < .05$) t-values across the whole time frame. Second, clusters (e.g., contiguous significant t-values) were identified, for which a cluster-level t-value was given as the sum of all single sample t-values within the cluster. Third, the significance of cluster-level t-values were assessed by generating Monte Carlo distributions ($N = 2000$) thereof and determining the probability of their occurrence given the distribution. Those clusters whose t statistic exceeded the threshold ($t = 2.8$, Bonferroni-corrected for multiple comparisons) were then tabulated for each contrast. The magnitude of contrasts in the identified clusters were then estimated by least square means (using the `lsmeans` function from the `lmerTest` package: Kuznetsova, Brockhoff, & Christensen, 2015). With this method, the following clusters were identified (using the `time_cluster_data` function in the `eyetrackingR` package: Dink & Ferguson, 2016): Steady target preference was observed

for the correct condition (300–2400 ms: $\beta = 0.13$, $SE = 0.03$, $t = 4.52$) and less robust but still significant target preference for the Δ1F condition (300–2200 ms: $\beta = 0.08$, $SE = 0.04$, $t = 2.17$), preferences flipped from target to distractor to target for the Δ2F condition (target preference – 300-600 ms: $\beta = 0.05$, $SE = 0.03$, $t = 1.81$, and 1500–2800 ms: $\beta = 0.06$, $SE = 0.04$, $t = 1.74$, distractor preference – 900–110 ms: $\beta = -0.05$, $SE = 0.03$, $t = -1.78$), preferences shifted from distractor to target for the Δ3F conditions (distractor preference – 500–800 ms: $\beta = -0.05$, $SE = 0.03$, $t = -1.76$, target preference: 1500–2300 ms: $\beta = 0.06$, $SE = 0.03$, $t = 2.04$), and distractor preference was observed for novel items (200–2300 ms: $\beta = -0.06$, $SE = 0.03$, $t = -1.82$) (positive t-values are linked to target preference and negative ones to distractor preference).

By looking at the time-course plot provided in Figure 3.4, differences were evident across all conditions. These observations were confirmed by time-course analyses that tested the significance of contrasts between each condition pair summarized in Table 3.2. Each pairwise comparison (i.e., comparisons between the correct and one-feature-change conditions, between the correct and two-feature-change conditions, etc.) was found to be significant at $p < .005$ level (Bonferroni-corrected for multiple comparisons). Some comparisons identified multiple significant time intervals (e.g., the comparison between the one-feature-change and the novel conditions). In two of those cases, some time intervals were marginally significant (i.e., the comparisons between the two-feature-change and novel conditions and that between the three-feature-change and novel conditions). Comparable results (i.e., significant contrasts across all condition pairs) were obtained when the function `time_cluster_data` was supplied with a formula containing a linear mixed effects model.

In the pupil dilation measure, a positive linear trend in pupil dilation in response to the degree of mispronunciation was obtained ($\beta = 0.05$, $SE = 0.02$, $t = 1.85$) in an analysis parallel to that of the linear mixed effects models of the looking time measure (c.f., Figure 3.3). The only modification to the model involved the outcome measure: overall mean pupil dilation (mm) in the naming phase, baseline-corrected by the minimum value at 700 ms. No other trends (quadratic, cubic, quartic) were found significant. In addition to `degree of mispronunciation`, `vocabulary`

Table 3.2: Significant contrasts across conditions in time-course analyses (Interval = time interval in the naming phase, $\sum t$ = cluster-level t value, p = p value associated with cluster-level t, Corr. = correctly pronounced familiar label, Δ1F = one-feature change, Δ2F, two-feature change, Δ3F = three-feature change introduced to the onset, Novel = novel label).

	Looking time			Pupil dilation		
Contrasts	Interval (ms)	$\sum t$	p	Interval (ms)	$\sum t$	p
Corr. vs. Δ1F	1200–1400	−3.31	*	–	–	–
Corr. vs. Δ2F	800–1500	−19.30	**	–	–	–
Corr. vs. Δ3F	400–1600	−28.69	***	1500–2900	2.33	†
Corr. vs. Novel	300–2300	-78.36	***	1300–2900	41.92	*
Δ1F vs. Δ2F	900–1200	−5.33	*	–	–	–
Δ1F vs. Δ3F	400–900	−11.83	**	–	–	–
Δ1F vs. Novel	300–1800 1900–2200	−43.77 −4.96	*** *	1300–2700	27.10	*
Δ2F vs. Δ3F	300–800	−10.10	*	–	–	–
Δ2F vs. Novel	300–800 1500–1900	−16.07 −6.90	** *	1300–2900	35.08	**
Δ3F vs. Novel	400–600 1100–1400 1600–1800	−3.18 −5.27 −4.48	† * †	1300–1500 1600–2400	3.31 16.97	† *

†: $p < .1$, *: $p < .05$, **: $p < .01$, ***: $p < 0.001$

`size` was also a significant positive predictor: $\beta = 0.06$, $SE = 0.02$, $t = 3.63$.

Judging by the respective time-course plot given in Figure 3.5, differences emerged between the correct and three-feature change conditions, and between the novel and all the other conditions. Across-condition time-course analyses (identical to the ones performed on the looking time data) supported these observations (the summary of which is provided in Table 3.2). The following contrasts reached significance: that between the novel and all the other conditions. The contrast between the correct and three-feature change conditions was found marginally significant. In line with the looking time models, results were again replicated by using a linear mixed effects model as argument in the function `time_cluster_data`.

3.4 Discussion

3.4.1 Looking behavior

The model output from the linear mixed effects model indicated that children's overall looking behavior was modulated by degree of mismatch such that the more featural overlap existed between the heard label and the correct target label, the more target looks were obtained. In cases when there was no overlap whatsoever with the correct label (i.e., the novel condition), the looking preference flipped to the distractor picture. Following previous work (e.g., Swingley & Aslin, 2000), we interpreted target (or distractor) looking preference as a sign for associating the heard label with the target (or distractor) picture; the earlier and the more prolonged the looking preference towards a picture in response to a given auditory label, the stronger the established association between the picture and the label. Therefore, this finding suggested gradient sensitivity to featural distance and, as such, the present study corroborated previous work conducted in intermodal preferential looking paradigms (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005). Sensitivity to the degree of mispronunciation (*contra* the findings of Swingley & Aslin, 2002 and Bailey & Plunkett, 2002) was possibly uncovered by using unfamiliar distractor pictures that can serve as plausible referents for the novel and mispronounced labels (Mani & Plunkett, 2011a; Ren & Morgan,

2011; White & Morgan, 2008; White et al., 2005).

Time-course analyses on looking preference further revealed target preference in response to the correct and Δ1F conditions and distractor preference in response to the novel condition, which is in line with earlier results that averaged over the naming phase (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005). Note that in the these conditions, looking preferences did not significantly shift, that is, children were more inclined to look at one picture over another for the duration of the naming phase.

Moreover, the results from time-course analyses extended on previous work in two respects. First, they enabled detecting significant looking preferences in intermediate conditions wherein children exhibited non-stable looking preferences, oscillating between distractor and target preference (the detection of which would not have been possible if looking time data would have been averaged across the entire time window). The dynamic shifts in their looking preference when presented with Δ2F and Δ3F labels suggests that children attempted to form links between those – largely mispronounced – labels and both pictures, yet stable link formation was disrupted by the ambiguity inherent in the mispronunciation manipulation of the current study: changed onset coupled with unchanged rhyme. In particular, Δ2F labels initially patterned with correct and Δ1F conditions by exhibiting target preference (as a possible sign of attempting to associate the label with the target image), then switched to distractor preference (a potential attempt to map the label with the distractor image), and finally shifted and remained with target preference. Since even in the featurally manipulated conditions, the fact that the rhyme of the word was always produced correctly may have facilitated the – albeit interrupted – retrieval of the correct word form and its mapping to the target picture. The Δ3F condition, on the other hand, seemed to follow a trajectory similar to the novel condition by exhibiting an initial distractor preference, which could signal an attempt to link the label with the distractor image. Eventually, however, looking preference shifted to the target image, which similarly to the shift in the Δ2F condition could have been caused by rhyme identity with the correct label. Thus, the distractor → target shift that occurred around 1000–1500 ms in response to the Δ2F and Δ3F conditions was probably due to delayed consolidation of

the largely mispronounced label with the correct lexical entry and in turn with the target picture.

The second set of novel findings involved time-course analyses of the looking time data that pitted conditions against each other and, as a result, identified significant differences *across all conditions* (c.f., Table 3.2). This is the first time that complete gradient sensitivity to the degree of mispronunciation has been observed as past research did not report differentiation between conditions containing large degrees of mispronunciation, Δ2F and Δ3F (Mani & Plunkett, 2011a; Ren & Morgan, 2011; Tamási et al., in press; White & Morgan, 2008; White et al., 2005). Time-course analyses in the present study found that differentiation between the looking patterns of correct pronunciation and small degree of mispronunciation (Δ1F) emerged relatively late at 1200 ms and only lasted for a short time (200 ms) in the naming phase. The function of featural distance between correct and mispronounced form positively predicted differentiation from the correct pronunciation. That is, the larger the featural distance, the earlier and longer the differentiation took place: Δ2F – at 800 ms for 700 ms; Δ3F – 400 ms for 1200 ms; novel label – at 300 ms for 3000 ms. In fact, this finding could be generalized to differences between any given condition pair: the less featural overlap across conditions, the longer differential response was obtained (one-step distance: 200–400 ms duration, two-step distance: 400–700 ms duration, three-step distance: 1200–1500 ms duration, four-step distance: 2000 ms duration. Latency did not follow such a strict ordering, but the trend was present nonetheless).

3.4.2 Pupillary response

Concerning the pupil dilation measure, the prediction that degree of mispronunciation affects magnitude of pupil dilation seemed to be borne out (in line with the findings of Tamási et al., in press) given the linear trend obtained by linear mixed effects models. Considering pupil dilation to be a direct measure of cognitive effort, finding pupillary differentiation across featural distances may be interpreted such that the more featural manipulations were introduced to the target label, the more cognitive resources were recruited to recover the appropriate lexical entry (Tamási et al., in press). In the novel condition – in which case a lexical entry cannot be

retrieved – children presumably establish a link between the novel label and the distractor picture (as suggested by their distractor preference in the looking time measure). Significant differences between the novel label condition and all other conditions suggest that pupil dilation reflects differences in the processing of somewhat familiar (correctly and incorrectly pronounced) labels and unrelated novel labels.

However, the positive trend found by linear mixed effects models in pupil dilation appeared to be driven by the significant differences between the correct and Δ3F conditions and those between the novel and all other conditions as no other contrasts were found significant in time-course analyses. Namely, no significant differences have been observed across correct and small-feature-change (Δ1F and Δ2F) conditions and across the mispronounced label conditions (Δ1F vs. Δ2F, Δ1F vs. Δ3F, Δ2F vs. Δ3F). To account for the relatively suppressed gradient response to degree of mispronunciation in contrast to previous findings (Tamási et al., in press), we review potential reasons that can stem from differences in design and suggest remedies thereof for future studies.

Unlike with the looking time measure, no direct comparison with past findings was possible as previous studies with children that used pupil dilation as outcome variable employed a relatively simpler, single-picture pupillometry paradigm (Fritzsche & Höhle, 2015; Tamási et al., in press, 2016b). In two of these studies, correct and mispronounced labels of the target referent were presented to the children along with the target picture, thus there was always a semantic link that could be established between the heard label and the target picture (Tamási et al., in press, 2016b). In these tasks, recall that sensitivity to the difference between correct and mispronounced labels was established – and Tamási et al. (in press) in particular found significant contrasts between pupillary reactions to small (Δ1F) and large degrees (Δ2F and Δ3F) of mispronunciation.

Some discrepancies between the previous single-picture pupillometry paradigms and the present intermodal preferential looking paradigm could have suppressed the emergence of differential response to featurally similar conditions. The first obvious divergence to point out is the inclusion of a distractor picture in the present study. It is important to reconsider the findings that have been obtained by single-picture pupillometry paradigms Kuipers and Thierry (2013) and Fritzsche and Höhle (2015). When they

presented monolingual children with labels that were semantically unrelated to the target picture (e.g., hearing *flower* while looking at a picture of a horse), no increase was found in the pupillary response relative to the response given to a semantically related label and referent (e.g., hearing *horse* while looking at a picture of a horse). However, in both of those studies, there was no alternative image to attach the label to as a single picture was shown in each trial. Therefore, monolingual children were unlikely to consider a novel label to belong to a referent whose name they were already familiar with due to the mutual exclusivity mechanism (Halberda, 2003). In the current study, contrastingly, a plausible referent was provided for those labels that did not quite match the target picture (i.e., a picture of an unfamiliar distractor item). For this reason, the same mechanism of mutual exclusivity can explain the results of the current study. In the presence of a distractor picture acting as plausible referent and upon presentation of a novel auditory label, larger pupillary response was exhibited, possibly indicating the cognitive cost of establishing the link between novel label and distractor picture.

A second, closely related divergence from the single-picture pupillometry paradigm presented in Chapter 4 is the addition of the novel label condition. In the present study, the novel condition resulted in a significantly larger degree of pupil dilation than all the other conditions. This suggested that resource consumption was highest when children were expected to attach a novel label to a novel distractor picture, thus establishing a link. It is possible that including such a condition that is semantically and phonologically unrelated to the target label may have interfered with the more subtle effect of small degrees of mispronunciation (i.e., fewer than two feature changes across conditions). In the single-picture pupillometry paradigm of Fritzsche and Höhle (2015), the authors presented half of the adult participants with correctly pronounced target labels, mispronounced target labels, and unrelated labels (e.g., *Tisch*, 'table'; *Kisch*, 'table' mispronounced; *Schaf*, 'sheep') along with a single target picture. For the other half of the participants, the first two conditions and the target pictures were the same, but the third condition was changed to unrelated non-words (e.g., *Tisch*, 'table'; *Kisch*, 'table' mispronounced; *Saaf*, 'sheep' mispronounced). Participants' pupillary responses in response to these conditions were recorded. In both groups, the mispronounced tar-

get label was associated with a larger pupil dilation than the correctly produced target label (i.e., mispronunciation effect). However, an interaction has been observed between the magnitude of the mispronunciation effect and word status of the unrelated condition such that when unrelated non-words were present, the mispronunciation effect grew weaker than when unrelated words were present. This finding exemplifies how an introduction of phonologically and semantically unrelated conditions and manipulations thereof may influence the magnitude of contrasts between the other, related conditions. Therefore, one way to assess whether more gradient pupillary response to degree of mispronunciation can be obtained in an intermodal preferential looking paradigm is to employ stimuli that are always semantically and phonologically related to the target picture (i.e., either the correct or mispronounced form of the target label), similarly to the stimuli used in single-picture pupillometry paradigms (Tamási et al., in press, 2016b).

The third divergence is the relative timing of visual and auditory stimuli. In single-picture pupillometry paradigms (Fritzsche & Höhle, 2015; Tamási et al., in press, 2016b), the onset of visual stimuli preceded that of the auditory stimuli by 1000 ms, while in the current study, the two stimuli were presented simultaneously in the naming phase. In the single-picture pupillometry studies, the silent presentation of visual stimuli for 1000 ms was introduced to provide the pupil with time to process seeing the picture as well as to adjust to the luminance of the picture. In the present study, processing of the visual and auditory stimuli in the naming phase was simultaneous (similarly to other intermodal preferential looking studies), making the disambiguation of whether visual or auditory stimuli the pupil dilation was in response difficult – though of course, first children are presented with the very same visual stimuli in the salience phase in order to minimize processing not related to the experimental (auditory) manipulation.

Another potential confound was that the pupil's adjustment to the particular luminance conditions commenced at the same time as the experimental manipulation. In fact, by looking at Figure 3.5, a contracting pupillary response can be observed for 700 ms prior to dilation, a pattern that has been noted in response to the visual stimuli in single-picture pupillometry paradigms (Tamási et al., in press, 2016b). The present analyses

included the same time window for looking time and pupil dilation analyses (200–3000 ms). When the time window was restricted to 700–3000 ms in the linear mixed effects models of pupil dilation, comparable results were obtained (linear trend: $\beta = 0.05$, $SE = 0.02$, $t = 2.02$). To minimize the effect of possible confounding factors and thus potentially obtain more significant pupillary contrasts in response to featurally similar conditions in future intermodal preferential looking studies, we suggest introducing in the naming phase a 1000 ms lag in the presentation of the auditory stimuli compared to that of the visual stimuli (the setup of the salience phase is proposed to remain unchanged). Asynchronous presentation – delaying the auditory label in relation to the picture – has been linked to lower processing load than synchronous presentation in other eye tracking studies (e.g., Althaus & Plunkett, 2015). Such a lag can be made more natural by embedding the auditory experimental item in a simple carrier phrase such as *Look! <item>* (e.g., Mani & Plunkett, 2011a). In addition to the afore-mentioned benefit – potentially obtaining more robust results in the pupil measure–, this modification would make baseline-correction more straightforward as well (by providing the option to choose a portion of the silent phase as baseline).

3.4.3 Comparing looking behavior and pupillary response

Analyzing looking behavior and pupillary response in one study provides a unique opportunity to compare the findings across the two measures. In contrasts with significant – or marginally significant – statistics (correct vs. Δ3F, correct vs. novel, Δ1F vs. novel, Δ2F vs. novel, Δ3F vs. novel), a systematic delay of around 1000 ms was recorded in the differentiations of pupillary response as opposed to those of the looking time measure. Such a delay would be expected given the findings of pupillometry literature on other cognitive tasks (reviewed in Beatty & Lucero-Wagoner, 2000). Furthermore, the duration of differentiation in the pupillary response seemed to be more uniform and thus less affected by featural overlap than that in the looking time measure. This observation is again consistent with previous research that characterize the pupillary response as stable (e.g., Klingner, 2010a).

3.5 Conclusions

The intermodal preferential looking paradigm is the most popular approach to assess early lexical knowledge. The current study, for the first time, analyzed pupil dilation data in conjunction with looking time data collected from a standard intermodal preferential looking paradigm. As a result, children's sensitivity to phonological overlap has been confirmed by the looking time measure and partially confirmed by the pupil dilation measure.

As for looking preference, each additional featural manipulation seemed to inhibit word recognition further. The recording of such nuanced information about children's looking behavior was made possible by employing time-course analyses. Changing one feature in the target label weakened target preference and thus the overall association between label and target picture, indicating disruption in retrieving the appropriate lexical entry. Changing two features introduced oscillation between target and distractor preference (target → distractor → target), suggesting interruption to establishing a link between the label and the target picture, and in turn interruption to recovering the correct lexical entry. Changing three features induced initial distractor preference that flipped to target preference, suggesting an attempt to link the label first with the distractor and then with the target picture and in turn, a delay in recovering the correct lexical representation. Finally, when no featural overlap between the target label and the heard label existed, the distractor was the preferred object to look at, indicating children's attempt at establishing a link between the two (c.f., mutual exclusivity). Taken together, these findings indicate complete gradient sensitivity to the degree of mismatch, confirming the notion that early lexical representations are fine-grained enough to encode sub-phonemic detail.

Furthermore, the pupil dilation measure has proven to be a valuable addition that can enrich intermodal preferential looking paradigms. The present findings were in line with past research that found phonological overlap to have influenced resource consumption as measured by pupillary response. More robust effects are expected to emerge in later studies following the implementation of a time lag between the visual and auditory stimuli in the naming phase. Even though the experiment employed a standard intermodal preferential looking paradigm and thus was not

specifically designed to detect significant effects in the pupil dilation measure, the fact that those were obtained regardless speaks to the resilience and robustness of the pupillary response – no matter the paradigm. We thus establish that pupillometry can be used in combination with intermodal preferential looking paradigms to provide an additional – dynamic and gradient – method to study children's cognitive processing, hence contributing to our understanding of lexical development.

Part II

Looking beyond the phoneme

Chapter 4

Consonant clusters in adults' and children's word recognition and production

Highlights:

- We studied consonant cluster processing with online procedures and production tasks.
- We assessed sensitivity to the contrast between correct vs. epenthesized clusters.
- Cluster type (homo-/heterorganicity) modulated the size of contrast in pupil dilation.
- In production, cluster type was consistent with type of phonological process observed.
- Results converge to show that cluster type affects adult and child word processing.

Abstract

We study consonant cluster processing to investigate whether homorganic and heterorganic clusters are structurally different. Arguably, homorganic consonants form a more unified cluster representation than heterorganic consonants by virtue of sharing their place of articulation. Such a distinction is suggested by a cross-linguistic asymmetry in the prevalence of epenthesis (rare in homorganic, common in heterorganic clusters) as well as articulatory and acoustic/perceptual reasons. Based on those considerations, the detection of epenthesis in homorganic clusters is more likely than in heterorganic clusters, which may differentially affect processing costs and hence the degree of pupil dilation. Results were consistent with this prediction. Complementing the perceptual results, proportions of phonological processes in production were consistent with structural difference. Comparable results were obtained with thirty-month-old children, indicating that such differences arise early. These findings converge to show that cluster type modulates word recognition and production from early in language development and into adulthood. (149 words)
Keywords: *Lexical development, lexical representations, consonant clusters, epenthesis, eye-tracking, pupillometry.*

4.1 The processing of consonant clusters in adults

4.1.1 Introduction

Consonant clusters – adjacent consonants within words (e.g., *s* and *t* in *stone*) – are cross-linguistically frequent and diverse structures. According to the World Atlas of Language Structures Online, approximately 87% of languages license (some type of) consonant clusters (Maddieson, 2013). This study addresses the question of how such prevalent structures are represented in the mental lexicon: Do consonants indeed form a single unit as the term *cluster* implies or is it a misnomer? We approach this question by evaluating speakers-listeners' perception and production patterns concerning different types of consonant clusters.

The principle that sequences of phonemes identical in one feature (say, place of articulation) *share* that feature permeates Autosegmental Phonology (Archangeli, 1984; Goldsmith, 1976; Hayes, 1986; E. Sagey, 1988; E. C. Sagey, 1986; Steriade, 1982) and various featural geometries (Clements, 1985; Halle & Vergnaud, 1980; McCarthy, 1986; Padgett, 1991; Selkirk, 1984). Applying this principle to consonant clusters, a basic distinction is whether the respective consonants have an identical place of articulation (and thus share their place of articulation) – homorganic clusters, e.g., *st*one – or not – heterorganic clusters, e.g., *gl*ass (Goldsmith, 1976; McCarthy, 1986; E. Sagey, 1988). Such a distinction can be represented as seen in Figure 4.1: homorganic consonant clusters sharing a place feature and being connected to the same place feature node (4.1A) and heterorganic clusters having separate place feature nodes (4.1B). Of interest here is whether such a distinction is encoded on a representational level, namely whether homorganic clusters form a unit based on their shared place features while heterorganic clusters remain separate units at this level of phonological representation.

Homorganic and heterorganic clusters exhibit an asymmetry that is widely documented across languages: schwa epenthesis[1] preferentially takes

[1] In the interest of consistency, the current study refers to the processes of schwa insertion for phonological, acoustic, or articulatory reasons – epenthesis, intrusion, and transitional vocoid insertion, respectively – as epenthesis throughout, while acknowledging the important distinctions between intrusive and epenthetic vowels on the acoustic phonetic / phonological level (Hall, 2006) and between transitional vocoids and

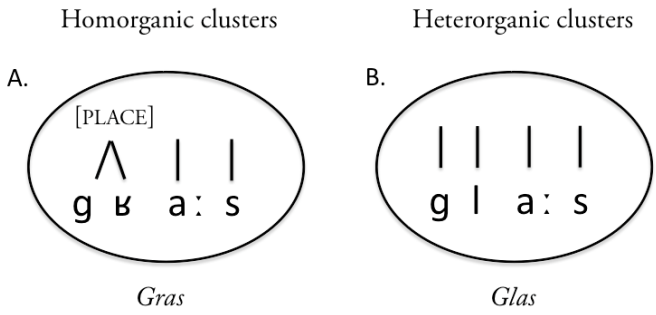

Figure 4.1: Hypothesis on the representational asymmetry between homorganic and heterorganic clusters.

place in heterorganic clusters as opposed to homorganic ones. Such preferential distribution of epenthesis in heterorganic clusters (versus homorganic clusters) can be observed in languages as diverse as Irish (Hickey, 1985), Dutch (Booij, 1999; C. C. Levelt et al., 2000), Scots Gaelic (Oftedal, 1956, cited by Hall, 2006), Ecuadorian Spanish (T. G. Bradley, 2006), Winnebago (Miner, 1989), as well as Trukese and Ponapean (de Lacy, 2002; Fischer, 1965).[2]

A representational asymmetry across the two cluster types (sketched in Figure 4.1) may explain why homorganic clusters are not as readily available to be broken up by epenthesis as heterorganic ones. If homorganic clusters form a unit, inserting a schwa in between the consonants may violate the structural integrity of the cluster (c.f., the ban on the crossing of association lines in autosegmental phonology and feature geometry, described by Goldsmith, 1976 and McCarthy, 1986). From this assumption it follows that homorganic clusters provide an incongruent setting for epenthesis. On the other hand, a schwa may be inserted freely in between heterorganic consonants as no violation of structural integrity would occur, the reason for which heterorganic clusters provide a congruent setting

epenthetic vowels on the articulatory phonetic / phonological level (Davidson & Stone, 2003).

[2] Some examples for epenthesis in heterorganic clusters include [bʁaten] → /beʁaten/ (German), [kalm] → /kal̲em/ (Dutch), [ʃalk] → /ʃal̲ek/ (Scots Gaelic), [g^{j}l^{j}aun] → /g^{j}^{i}l^{j}aun/ (Irish) (examples taken from Hall, 2006), [limra] → /lim̲ara/ (Ponapean) (de Lacy, 2002).

for epenthesis (see Figure 4.2A–B). Note that such representational asymmetry may stem from several sources, e.g., from differences in statistical distribution, perception, and production. The current study is not designed to adjudicate between those explanations, though it offers a review thereof in the General Discussion.

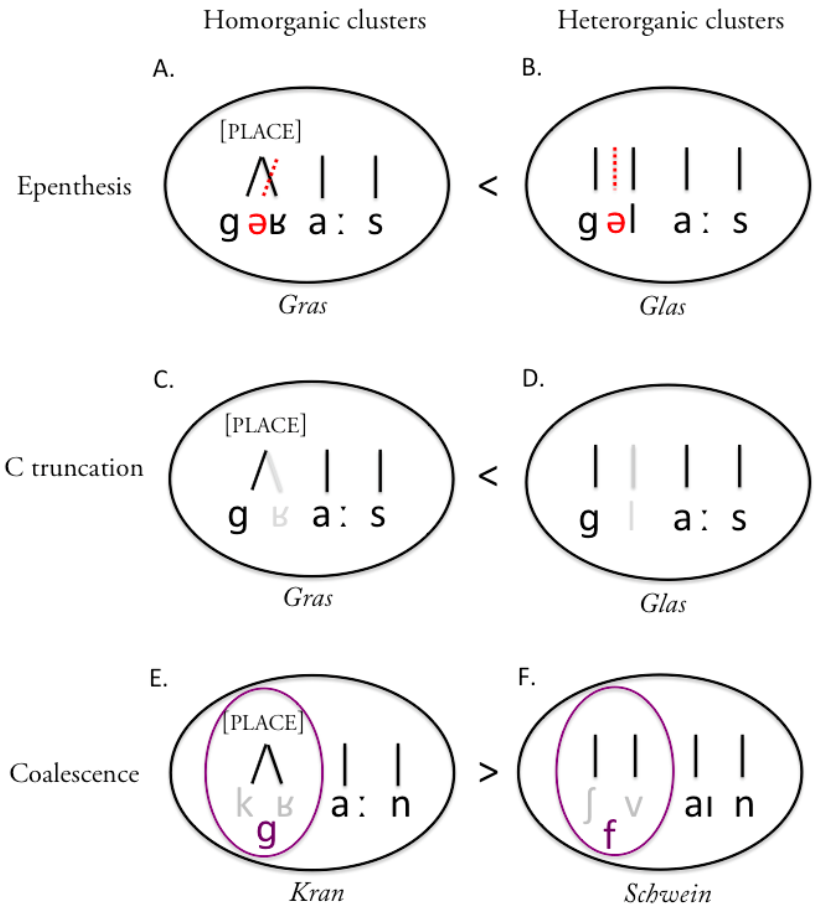

Figure 4.2: Preferential distribution of the three most common phonological processes relevant to cluster representation, presented on words containing homorganic and heterorganic clusters: epenthesis marked with red (A–B), single consonant (C) truncation marked with gray (C–D), and coalescence marked with purple (E–F).

In this study, we employ two experimental tasks to test the hypothesis that homorganic clusters form a unit in the mental lexicon, as opposed to heterorganic clusters, one tapping into perception, the other into production skills of participants. In the perceptual task, we test the hypothesis concerning the different representations of homorganic and heterorganic clusters by presenting listeners with correct productions of words with initial homorganic or heterorganic consonant clusters and with words in which a schwa had been inserted between the cluster consonants. While listening to these words and looking at the corresponding pictures, participants' pupillary response was recorded. Pupillometry is a minimally demanding online method that is based based on an involuntary reflex present already at birth (i.e., the pupillary response) and thus is suitable for all age groups (Karatekin, 2007). The degree of pupil dilation has been extensively used as an index of cognitive effort (reviewed in Beatty & Lucero-Wagoner, 2000).

We predict that when listeners are confronted with labels that contain epenthesized versions of clusters, it may require more cognitive effort to reconcile the epenthesis occurring in an incongruent context (i.e., in a homorganic cluster) with the correct representation compared to reconciling the correct label, than when the epenthesis occurs in a congruent context (i.e., in a heterorganic cluster). In other words, the difference between the cognitive effort required to process homorganic correct and epenthetic clusters are predicted to be larger than the difference between heterorganic correct and epenthetic clusters, which in turn yields a comparatively larger difference in the degree of pupil dilation (a model is sketched in Figure 4.3). Such a pattern ultimately corresponds to an interaction between condition (correctly produced and epenthesized items) and cluster type (homorganic and heterorganic clusters).

As for many other lexical and phonological processing tasks, it is necessary to address a number of potentially confounding variables. It may be expected that phonological, sub-lexical, and lexical characteristics affect cognitive load and one of its physiological correlates, pupil dilation. Apart from considering those factors as control variables, no specific predictions are to be put forth; they are included to confirm that the interaction between the critical variables (condition and cluster type) is significant over and above their contribution.

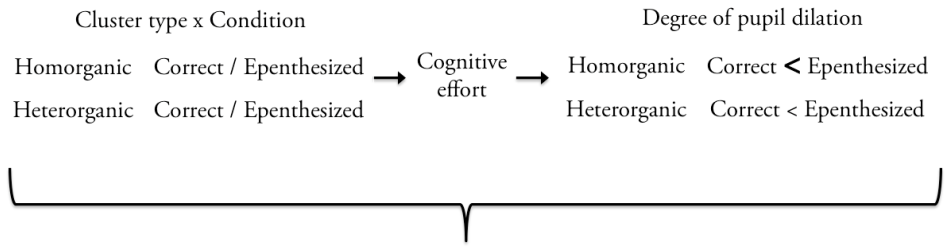

Figure 4.3: Predictions of the perceptual study.

Another way to assess the representation of clusters in the mental lexicon is via speech production data. Although admittedly more indirect due to potential limitations of the articulatory motor system (W. J. Levelt, 1992), results gained from production data can support or challenge those from perception data. Broadly, we investigate whether the distribution of phonological processes observed in production is consistent with the pattern observed in the perceptual data. The term *phonological process* refers to any deviation from the correct and canonical form in production. It most frequently occurs in children's production, but is evident in adult speech as well, for example in slip-of-the-tongue utterances. To this end, we also administer a separate production task with the same participants (and the same experimental items) to assess the proportion of phonological processes across homorganic and heterorganic clusters. We discuss the predictions on the likelihood of phonological processes across cluster types that have a relevant bearing on cluster representation: epenthesis, single consonant truncation, and coalescence (Barton et al., 1980; McLeod et al., 2001).

In general, we expect production to exhibit alterations and simplifications. Adult speakers are known to produce open transitions between consonants that result in transitional vocoids and even true schwa epenthesis given the right conditions (i.e., slow and/or emphatic speech) (Gafos, 2002; Hall, 2006; Steriade, 2009; Yun, 2014a). Specifically however, we also expect the proportion of phonological processes to be modulated by cluster structure. The following three predictions can be generated from

the hypothesis on the representational asymmetry between homorganic and heterorganic clusters.

First, if homorganic clusters organize into a more coherent complex unit than heterorganic clusters, epenthesis may violate the structural integrity of such a unit and therefore is predicted to occur less frequently than in heterorganic clusters that are proposed to be independently represented (see Figure 4.2A–B). Second, deleting one member of a purportedly higher-order unit of homorganic clusters may be unexpected: It is not clear why a truncation process would selectively target a part of and not the whole unit. There is no such prohibition against deleting a single segment in heterorganic clusters, and thus prevalence rates are predicted to reflect this (c.f., Figure 4.2C–D). Expecting such a pattern is typologically grounded, e.g., avoiding heterorganic clusters by deletion of one cluster consonant in Attic Greek (c.f., Section 7.5.1 in de Lacy, 2002). Third, coalescence among homorganic cluster consonants – i.e., feature sharing resulting in a single segment – is structurally more motivated and thus predicted to be more common than among heterorganic cluster consonants, assuming that homorganic clusters form a stronger unit than heterorganic ones (c.f., Figure 4.2E–F).

Given the hypothesis on representational asymmetry, we regard epenthesis and single consonant truncation as phonological processes that are consistent with a separate cluster representation (two independent segments) and coalescence as a phonological process consistent with a unified cluster representation (segments related in some fashion). Put differently, homorganic clusters are expected to undergo more phonological processes that are consistent with a unified cluster representation (assuming they form a unit) and heterorganic clusters to undergo phonological processes that are consistent with a separate cluster representation (assuming they are represented independently).

4.1.2 Method

4.1.2.1 Participants

Twenty-four adults (17 women), undergraduate students at the University of Potsdam who received course credit for their participation, were recruited. All participants reported normal or corrected-to-normal vision.

4.1.2.2 Stimuli

In anticipation of recruiting 30-month-old children along with adult participants, we adjusted the stimuli to be appropriate for participants in both age groups. For this reason, we employed words that are likely to be known by children. 17 easily depictable words with obstruent-initial onset clusters from FRAKIS (the German adaptation of the MacArthur-Bates Communicative Development Inventory, c.f., Szagun et al., 2009) were identified with a CCV, CCVC, or CCVCV syllable structure and word-initial stress. The words included nouns and adjectives. Due to our stringent phonological criteria, it was necessary to include 3 additional words reported to be known at 30 months of age (*Kran* 'crane', *Kröte* 'toad', and *grau* 'gray') were included from another source (Schröder, Gemballa, Ruppin, & Wartenburger, 2012).

Half of the words contained homorganic onset clusters, half heterorganic onset clusters. Critical German consonants, i.e., consonants that occur in the onset position of our stimuli, were classified by their place of articulation as follows: labial (i.e., /p/, /b/, /f/, /v/, and /m/), coronal (i.e., /n/, /t/, /d/, /l/, /s/, /z/, /ʃ/, and /ʒ/), and dorsal (i.e., /k/, /g/, /h/, and /ʁ/). Manner of the first consonant (stop / fricative) and the number of syllables (monosyllabic / disyllabic) were balanced across homorganic and heterorganic clusters. Each word was naturally produced with and without a schwa-epenthesis in an infant-directed manner by a female native speaker of German. The epenthesized versions of the labels resulted in forms that were no existing German words. The list of critical items is included in Table 4.1. The summary of various lexical and phonological characteristics from the Clearpond database (Marian et al., 2012) – word frequency (logged frequency per million), neighborhood density (mean frequency of phonological neighbors per million), and positional biphone probability (in the cluster, the probability that C1 is followed by C2) – are listed in Table 4.2. Apart from the critical items, 40 other easily depictable words from FRAKIS were included, 20 filler items that were always produced correctly, and 20 items with singleton onsets that deviated from the correct label by the exchange of one segment (both fillers and items with singleton onsets were reported in Chapter 2 in Tables 2.1 and 2.1, respectively).

Easily recognizable pictures that unambiguously matched the auditory

Table 4.1: Stimulus list, organized by cluster type, noted with IPA. Schwa epenthesis is shown in parentheses.

Word (*English*)	**IPA**
Homorganic clusters	
Gras (*grass*)	g(ə)ʁ
grau (*gray*)	g(ə)ʁ
grün (*green*)	g(ə)ʁ
Krabbe (*crab*)	k(ə)ʁ
Kran (*crane*)	k(ə)ʁ
Kröte (*toad*)	k(ə)ʁ
Schnecke (*snail*)	ʃ(ə)n
Stein (*stone*)	ʃ(ə)t
Stock (*stick*)	ʃ(ə)t
Stuhl (*chair*)	ʃ(ə)t
Heterorganic clusters	
blau (*blue*)	b(ə)l
Blume (*flower*)	b(ə)l
Brot (*bread*)	b(ə)ʁ
Clown (*ibd.*)	k(ə)l
Flasche (*bottle*)	f(ə)l
Fliege (*fly*)	f(ə)l
Frosch (*frog*)	f(ə)ʁ
Glas (*glass*)	g(ə)l
Knie (*knee*)	k(ə)n
Schwein (*pig*)	ʃ(ə)v

Table 4.2: Lexical and sub-lexical statistics of the stimuli, separated by cluster type: logged frequency (logFreq), neighborhood density (ND), and positional biphone probability (PBPP).

	logFreq	ND	PBPP
Homorganic	0.994 (*0.577*)	7.201 (*5.562*)	0.011 (*0.007*)
Heterorganic	1.429 (*0.223*)	6.528 (*4.241*)	0.006 (*0.002*)

label were selected and converted into a similar size (approx. 200 x 200 pixels displayed in a 300 x 300 pixel area). Four versions of the task were created, each picture occurring once in each version. Thus participants, having been randomly assigned to one version, never saw the same picture twice or heard the same label twice. In each version, participants were presented with 10 correct and 10 epenthesized items (5 homorganic and 5 heterorganic items in each condition). Altogether with fillers and unrelated items, each version contained 35 correctly and 25 incorrectly produced items.

4.1.2.3 Procedure

Collecting perceptual (pupillary response) data

Adults were told that their task was to watch a short movie, during which they should maintain their position. After providing informed consent, they were seated such that their eyes were approx. 60 cm from the computer screen. Changes in their pupil size were monitored by a Tobii 1750 corneal reflection eye-tracker in the ClearView software (Kruger, Schneider, & Westermann, 2006). All visual stimuli were paired up with their corresponding (correct or epenthesized) auditory labels. The picture was shown centrally on a 17" (1280 x 1024) TFT screen with a size of 300 x 300 pixels forming a horizontal and vertical viewing angle of 7.4°. The experiment started immediately following the calibration period (5 screen positions, approximately 30 seconds).

In each trial, a picture was presented and remained on screen for 4

seconds. One second after the picture appeared, the corresponding auditory label was played. The critical window of analysis was chosen to be the 3-second interval following the onset of the auditory stimulus. The experiment encompassed 12 blocks, each containing 5 trials (altogether 12 x 5 = 60 trials, of which 20 fillers and 20 unrelated). Before each block, an 'attention-getter' was presented (a short silent movie clip of animated cartoon characters and animals). The attention-getters were played in a loop until the experimenter pressed a key to start the next block. On average, the task lasted 15 minutes.

Collecting production data

After the perceptual task, adult participants were instructed to produce the labels of the experimental items (provided both in pictorial and orthographic format), and to do so twice: once in isolation and once with the definite article belonging to the label (e.g., *Krabbe* 'crab' and *die Krabbe* 'the crab'). The repetition was requested in order to maximize the chances of obtaining analyzable recordings of participants' utterances and doing so with variable phonological contexts.

4.1.3 Results

4.1.3.1 Pupillary response analysis

Successful trials were defined such that they contained pupil information from at least half the length of the trial. Based on this criterion, the proportion of successful trials was tabulated for each participant. Those participants who did not reach a threshold of 50% of successful trials (following Fritzsche & Höhle, 2015 and Tamási, McKean, Gafos, Fritzsche, & Höhle, in press) were excluded from further analyses. Based on this criterion, no adult participants were excluded. The mean number of successful trials was 19.5 out of 20 ($SD = 0.70$) in the experimental trials and 19.75 out of 20 ($SD = 0.85$) in the filler trials.

We employed linear mixed effects models with random intercepts and slopes using the `lmer` function (estimates were chosen to optimize the log-likelihood criterion) in the `lme4 R` package (Bates et al., 2014). Apart from handling unbalanced data sets and categorical variables better than

generalized linear models, linear mixed effects models allow the inclusion of continuous predictors and multiple random effects in the same model (Baayen et al., 2008; Bates, 2005).

The two critical within-subject factors, `Condition` and `Cluster type`, each with two levels (Correct / Epenthesized and Homorganic / Heterorganic, respectively) were entered as fixed effects into the model. Word frequency, neighborhood density, and positional biphone probability (estimated using the Clearpond database, c.f., Marian et al., 2012), summarized in Table 4.2 were included as control variables. In addition to those factors, manner of first consonant (stop / fricative) and number of syllables (monosyllabic / disyllabic) were entered into the model as control variables. All continuous variables were centralized and scaled for the analysis. Participants and items were entered as random effects into the model. Mean change in pupil diameter (i.e., the mean value extracted from the time window of 3000 ms after the auditory onset) was used as the outcome measure. It was calculated on a trial-wise basis and corrected for inter-subject and inter-trial variation by subtracting a baseline value (i.e., a mean value of a 100 ms interval before the onset of the auditory label). We repeated the calculations with two different (20 ms and 500 ms) baseline intervals and found that manipulating the baseline interval did not change the overall pattern. Each intercept and slope fitted by the model was adjusted by the effect of `Condition` and `Cluster type` nested in participants.[3] Since the levels of `Condition` and `Cluster type` were collinear, the correlation terms in the random effect structure was removed (Jaeger et al., 2011). The most parsimonious model was chosen by Likelihood Ratio Tests (Pinheiro et al., 2007) by using the `anova` function from the `stats` package (R Core Team, 2014).

Likelihood Ratio Tests determined that the most parsimonious model included `Neighborhood density` (β = -0.006, SE = 0.002, t = -2.27) as a negative predictor (denser neighborhood was associated with smaller degree of pupil dilation), `Positional biphone probability` (β = 0.004, SE = 0.001, t = 2.95) as a positive predictor (larger positional biphone probability was associated with larger degree of pupil dilation), `C1 manner`,

[3] Due to the possibility of overfitting and hence producing convergence errors, this model provided maximal specification warranted by the design described in Barr et al. (2013) and Jaeger et al. (2011).

i.e., the stop-fricative contrast (β = -0.006, SE = 0.002, t = -2.39) (stop-initial items are linked to lower degrees of pupil dilation than fricative-initial items), `Condition`, i.e., the correct-epenthesized contrast (β = 0.088, SE = 0.047, t = 1.84), as well as the interaction term `Condition` x `Cluster type` (β = -0.089, SE = 0.034, t = -2.62).[4] The interaction between `Condition` and `Cluster type` was driven by a difference such that a larger contrast was found between the homorganic correct and epenthesized clusters than between those of the heterorganic clusters (c.f., Figure 4.4). That is, the difference between homorganic correct and epenthesized clusters was significantly larger than the difference between heterorganic correct and epenthesized clusters.

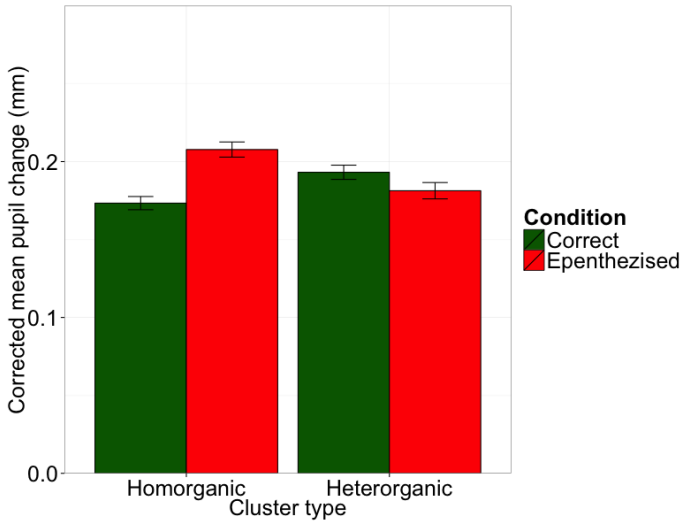

Figure 4.4: Adults' mean pupil size change in response to correct and epenthesized versions of homorganic and heterorganic clusters. Error bars represent 95% confidence intervals.

In what follows, we present a methodological approach complementary to pupillometry – item-wise analyses – in order to gain further insight in two respects: First, to gauge the effect of epenthesis on pupil size change within each individual item and thus assess item-specific contribution; and

[4] It was not possible to add more interaction terms to the model as they yielded rank deficiency.

second, to eliminate the possibility that the number of syllables and the manner of the first consonant confounded with the effect of cluster type (recall that it was not possible to include additional interaction terms in linear mixed effects models due to the ensuing rank deficiency). Item-wise analyses only contained `Condition` as predictor (akin to a series of *t*-tests, Bonferroni-correction was performed for multiple comparisons). Within each item, three possible patterns involving correctly produced and epenthesized items may emerge: pupil size associated with correctly produced items may be smaller than, equal to, or larger than that associated with epenthesized items. In line with the structural asymmetry hypothesis, more homorganic items were expected to exhibit the 'correct < epenthesized' pattern than heterorganic items, whereas fewer homorganic items were expected than heterorganic ones to be associated with pupil dilation patterns indistinguishable across the correct and epenthesized conditions. Pupil size in response to correctly produced items was not expected to be larger than epenthesized items in any of the cluster types.

The results of item-wise analyses confirmed those predictions (c.f., Table 4.3). The proportion of items that were associated with smaller degree of pupil dilation in the correct than in the epenthesized condition, i.e., exhibited the 'correct < epenthesized' pattern was 5 homorganic : 3 heterorganic items. Among those items associated with comparable pupil dilation across the correct and epenthesized conditions, i.e., exhibited the 'correct ≈ epenthesized' pattern, there were 2 homorganic : 6 heterorganic items (c.f., Table 4.3).

4.1.3.2 Production data analysis

Twenty of the 24 participating adults in the perception task provided production data. Most elicitations were repeated by the adults according to the instructions, albeit there was some noise-related data loss and, in one participant's case, failure to produce all experimental items twice. Taken together, 95% of the maximum 800 (20 adults x 20 labels x 2) elicitations were collected and analyzed. 38.15 out of 40 elicitations were produced per participant ($SD = 4.03$) and per item ($SD = 1.09$).

During the analysis, only the pronunciation of the cluster was considered, phonological processes in the vowel and/or in coda position were

Table 4.3: Summary of item-wise analyses on adults' perceptual data, organized by cluster type (left: homorganic; right: heterorganic) and degree of pupil dilation by condition (correct < epenthesized; correct ≈ epenthesized; correct > epenthesized). t thresholds were Bonferroni-corrected for multiple comparisons.

	Homorganic items	**Heterorganic items**
Correct < Epenthesized ($t > 2.58$)	grau Gras Krabbe Schnecke Stock	Flasche Fliege
Correct ≈ Epenthesized ($t \leq 2.58$ and $t \geq$ -2.58)	Kröte Stein Kran	Blau Blume Glas Knie Schwein Frosch
Correct > Epenthesized ($t <$ -2.58)	grün Stuhl	Brot Clown

not examined. Five percent of the clusters were produced with (varying lengths of) open transition in our sample, as determined by inspecting waveforms and spectrograms (using the Praat software, c.f., Boersma & Weenink, 2011).

Since in some German dialects, the fortis/lenis distinction tends to neutralize in the syllable onset position (Jessen & Ringen, 2002), neutralizations of this sort were not counted as phonological processes (16% of all elicitations). Other simplification processes such as spirantization (e.g., /kʁ/an → /kh/an) were identified in 10% of the elicitations and since they have no bearing on cluster structure, they were coded as 'other' phonological process in the analysis.

Table 4.4: χ^2 test statistics on adults' production data, including the omnibus 2 x 3 χ^2 test statistic, Fisher exact probability tests (2-tailed), *a priori* 2 x 2 contrasts ransacking and partitioning, categories contrasted, χ^2 test statistic, Fisher exact probability test (1-tailed), odds ratio (OR), and .95 confidence intervals.

Omnibus			χ^2	p			
			11.05	.004			
	Category 1	Category 2	χ^2	p	OR	.95 CI	
						Lower	Upper
Ransacking	*Correct*	*Epenthesis*	8.59	.003	.33	0.16	0.69
Partitioning	*Epenthesis*	*Epenthesis*	8	.002	0.34	0.16	0.71

The only type of phonological process found that was consistent with a separate cluster representation was epenthesis / open transition, i.e., no single consonant truncation was recorded. Moreover, no coalescence, a phonological process consistent with a unified cluster representation, was observed either. When homorganic and heterorganic clusters were considered separately (c.f., Figure 4.5), epenthesis was more prevalent among heterorganic clusters than homorganic ones (7.3% vs. 2.6%). Omnibus 2 x 3 and *a priori* contrast 2 x 2 χ^2 tests confirmed that `cluster type` significantly modulated the proportion of epenthesis (χ^2 statistics, Fisher exact probability tests, odds ratios, and .95 confidence intervals are given in the first section of Table 4.4). Following the suggestions laid out by Sharpe (2015), after obtaining a significant omnibus χ^2 test statistic, we

performed planned 2 x 2 χ^2 tests called ransacking (i.e., contrasting the correct and epenthesis categories) and partitioning (i.e., contrasting the collapsed correct + other and the epenthesis categories). For the calculations, we used the software available at Lowry (2004). The prevalence of epenthesis in heterorganic clusters as opposed to homorganic clusters persisted both in stop-initial and fricative-initial clusters (c.f., Figures 4.6 and 4.7).

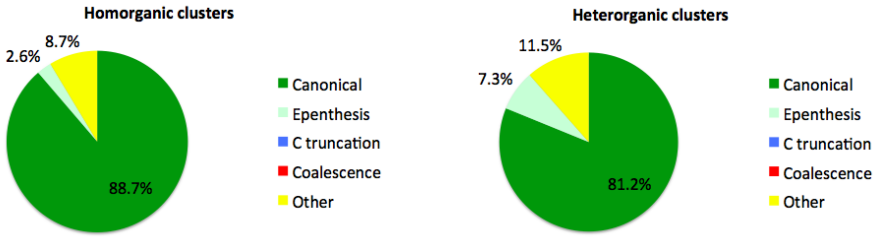

Figure 4.5: Adults' production of homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

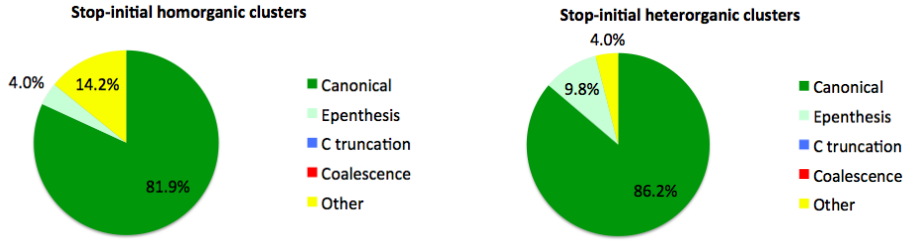

Figure 4.6: Adults' production of stop-initial homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

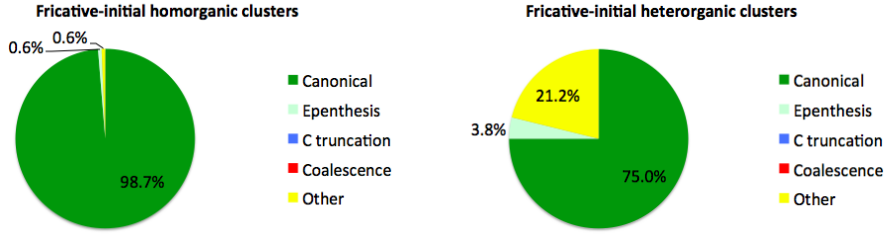

Figure 4.7: Adults' production of fricative-initial homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

4.1.4 Discussion

Our perception study tested whether listeners gave a differential response to epenthesis when occurring in an incongruent context, i.e., in homor-

ganic clusters versus in a congruent context, i.e., in heterorganic clusters. Specifically, we tested whether the processing of epenthesis in homorganic clusters require more cognitive effort than the processing of epenthesis in heterorganic clusters by measuring pupil dilation: a larger difference was expected to emerge in the former compared to the latter condition. Our prediction has been borne out; we detected an interaction between cluster type and condition that was significant over and beyond the contribution of controlling factors.

The production task provided results consistent with the hypothesis on structural differences between the cluster types. Significantly more epentheses were recorded during the production of heterorganic clusters than that of homorganic clusters, which may be taken as support of the notion of relatively more independent representation of heterorganic clusters. As no single consonant truncation or coalescence were recorded with adults, no other phonological processes that affected cluster representation have been observed. Even though the asymmetry observed in the prevalence of epenthesis in production is consistent with our predictions, the source of such an asymmetry may be external, such as motor constraints (i.e., in heterorganic clusters, moving from one articulator to another is associated with a greater likelihood for intervening material to appear). Potential sources of representational asymmetry will be reviewed in the General Discussion.

The findings presented converge to support the claim that homorganic clusters are more tightly connected in the mental lexicon, and behave as if they formed a unit, unlike heterorganic clusters. Such findings indicate sensitivity to the structural difference between homorganic and heterorganic clusters.

4.2 The processing of consonant clusters in children

4.2.1 Introduction

In this section, we ask how and when such sensitivity to the structural difference by homorganicity comes about. Is it already present in young children or does it arise later on due to lexical development and/or motor practice? We regard this exploratory study as the first step in investigat-

ing cluster acquisition from the perspective of homorganicity, integrating online perception data via pupillometry and speech production data.

Several generalizations have been made with regards to cluster acquisition. There is evidence that cluster acquisition is guided by structural constraints such as those of prosody, cluster and word position (Goad, Rose, Kager, Pater, & Zonneveld, 2004; C. R. Marshall & van der Lely, 2009; Treiman, 1991; Treiman & Cassar, 1996) and sonority (Ohala, 1999), as well as non-structural constraints, i.e., phonetics (Demuth & McCullough, 2009) and lexical characteristics such as frequency (C. C. Levelt et al., 2000) and phonotactic probability (Mattys et al., 1999; Saffran & Thiessen, 2003).

First, word-medial and word-final clusters are usually produced earlier and more accurately than onset clusters, excepting s + C clusters (Chambless, 2004; Kirk, 2008; Kirk & Demuth, 2005; C. R. Marshall & van der Lely, 2009). In some cases, morphological, frequency or phonetic effects may reverse this order, e.g., in French (Demuth & McCullough, 2009).

Second, consonant clusters are preserved better in stressed positions than unstressed ones (C. R. Marshall & van der Lely, 2009). Another related effect is position within a cluster. The leftmost consonant has better chances of survival when the cluster is truncated (Goad et al., 2004; Treiman, 1991; Treiman & Cassar, 1996).

Third, cluster acquisition is governed by sonority restrictions. As shown by the linguistic output of consonant truncations, it is generally the less sonorous cluster consonant that survives (*try* → *ty*) in order to keep the sonority distance between onset and rhyme maximal (Barlow, 2003, 2005; Fikkert, 1994; Gierut, 1999; Ohala, 1999; Yavas & Gogate, 1999). To assess the scope of sonority restrictions in language development, children have been tested on their skills with various cluster types including those that are unattested in their language. Children were found to perform better with sonority-abiding clusters (e.g., stop-liquid clusters such as *tr* and *bl*) than with those that violate sonority restrictions (e.g., liquid-stop clusters such as *bd* and *lb*) (Ohala, 1999; Pertz & Bever, 1975) and moreover, the sonority profile of the cluster predicts performance (i.e., identity judgment) therewith (Berent et al., 2011).

Fourth, phonetic reduction of non-prominent phonetic material in adult speech may be one of the reasons that word-initial clusters appear sooner

in production than word-medial and final ones. For example, the reduction of codas and unstressed syllables in French can result in phonetic reduction (Demuth & McCullough, 2009).

Fifth, the frequency of certain cluster types as well as cluster tokens in child-directed speech predicts the order of acquisition of clusters. The trend is that the more exposure a cluster type or token receives in the speech input, the earlier and more reliably it is acquired (Demuth & McCullough, 2009; C. C. Levelt et al., 2000).

Sixth, phonotactic probability seems to play a role in how clusters are extracted from fluent speech, as shown by word segmentation tasks using the head-turn preference paradigm (Mattys et al., 1999; Saffran & Thiessen, 2003). Younger than one-year-old infants seem to differentiate by the permissibility and the frequency of consonant clusters in habituation tasks (the design of which is discussed in Section 1). For instance, 9-month-old infants are able to quickly recognize voicing differences in word-medial cluster consonants, i.e., whether the first consonant is voiceless and the second is voiced (e.g., the speech stream contained items such as *dakdot* and *gopguk*), or vice versa (e.g., *todkad*, *kogpub*) (Saffran & Thiessen, 2003). Using a similar habituation task, 9-months-olds are furthermore shown to react differentially to clusters that co-occur more frequently within words than across word boundaries (e.g., 'ft', 'vn') and to clusters with the reverse pattern of probability (e.g., 'fh', 'mk') (Mattys et al., 1999). Such phonotactic knowledge presumably facilitates word segmentation from a continuous speech stream, preparing the infants to build their lexicon (Werker & Gervain, 2013).

Although these studies provide valuable information about the course of cluster development, our understanding thereof is yet to be expanded in several respects. The present study proposes to investigate how cluster acquisition is affected by homorganicity, a question that has not been addressed before. As much as it was possible being restricted to words familiar to the children, we attempted to control for the factors discussed above that may influence cluster processing (lexical position, lexical stress, manner of the first consonant). Since adults were presented with the same stimuli as the children, these factors were described in detail in Section 4.1.2. Lexical position was kept constant by using only word-initial onset clusters. Likewise, stress always fell on the first syllable. Given the

limitation of choosing words familiar to the children, we could not entirely control for the effect of sonority.[5] Instead, we balanced the manner of the first consonant (by including 6 stop- or 4 fricative-initial clusters in both heterorganic and homorganic groups). To further address the effect sonority may play in the perception task, item-wise analyses are conducted. We minimized acoustic and phonetic variation in our stimuli by recording a single speaker using child-directed speech and carefully selecting our stimulus set (making sure that no phonetic information is truncated and that intonation is kept even). In order to account for the effect of (sub-)lexical factors, word frequency, neighborhood density, positional biphone probability, manner of first consonant, and number of syllables were included as control variables in the analysis (c.f., Section 4.1.3.1). In addition to the variables employed in the adult model, vocabulary size was included as it may affect the detailedness of lexical representations in children (Munson, Edwards, & Beckman, 2005; Munson et al., 2011).

Apart from the novel perspective of homorganicity, we propose to assess the state of children's lexical representations with a more comprehensive approach. The majority of experimental studies on cluster acquisition have analyzed production data, collected either naturalistically or experimentally (with notable exceptions of the judgment task employed by Pertz & Bever, 1975: 'which of these two clusters are easier, more likely to occur or more usual in the world's languages?' and the offline perception task by Berent, Harder, & Lennertz, 2011: 'do these two words sound the same?'), as a means of assessing children's proficiency with clusters. Since as of yet, no online experiment has tested children's processing of clusters in a perceptual task, it is an open empirical question whether cluster representations are already mature and adult-like at a younger age than suggested by their production output. More specifically to our study, we ask how early structural differences between homorganic and heterorganic cluster representations emerge in language development – Do young children who are in the early stages of the protracted process of cluster acquisition show adult-like sensitivity to the structural difference between

[5] Three of the homorganic clusters are /ʃ/ + stop and 1 of the heterorganic clusters is /ʃ/ + fricative, clusters that are regarded as sonority non-compliant (as the sonority value of the first cluster consonant is not strictly lower than the following consonant), while the remainder of the clusters (7 homorganic, 9 heterorganic) is sonority-compliant, c.f., Ohala (1999).

homorganic and heterorganic clusters?

Cluster production is reported to emerge in 2-year-olds (Lleó & Prinz, 1996). Despite such an early start, the attempt to produce consonant clusters remains rare in comparison to that of singletons (Stoel-Gammon, 1987), which may explain the protracted mastery of some clusters until well into school-age (Ingram, Pittarn, & Newman, 1985; Smit, 1993). At 30 months of age, German children can be expected to be familiar with consonant clusters, as evidenced by their reliable production of at least some clusters (Fox & Dodd, 1999). However, in accordance with past studies on cluster production (Dyson & Paden, 1983; Fox & Dodd, 1999; Lleó & Prinz, 1996; McLeod et al., 2001; Smit, 1993; Watson & Scukanec, 1997), 30-month-olds are expected to use many phonological processes including epenthesis, single consonant truncation, and coalescence, enabling us to test the hypothesis whether the respective proportions of phonological processes are modulated by cluster type. While coalescence and single consonant truncation is typically only present in the earliest stages of cluster acquisition, epenthesis is pervasive throughout the whole developmental trajectory (Barton et al., 1980; Dyson & Paden, 1983; McLeod et al., 2001) and as we have also seen in our adult production data, it may even surface later in adult speech (c.f., Yun, 2014a).

Our dependent measure for the perceptual task, pupillometry, is especially well suited to test young children. Being based on the pupillary reflex that is present from birth, pupillometry is able to detect infants' sensitivity to incongruence in the auditory domain (Hochmann & Papeo, 2014) as well as young children's sensitivity to (differing degrees of) mispronunciation (Fritzsche & Höhle, 2015; Tamási et al., in press).

4.2.2 Method

4.2.2.1 Participants

Forty-eight 30-month-old monolingual German children were recruited (*M = 30, SD = 0.56*) (26 girls) from the BabyLAB Participant Pool at the University of Potsdam. Caregivers reported no developmental and sensory disabilities. We assessed children's familiarity with the experimental items by using the parental report FRAKIS (Szagun et al., 2009). According to the report, children were fairly familiar 79.9% (*SD = 16.9*) with the

correct experimental labels. The children's reported vocabulary size (M = 410; SD = 112) was within the FRAKIS normed range for 30 month-old German-speaking children (M = 439, Szagun et al. 2009). Due to providing insufficient data, 5 children were excluded from the analyses (see Results).

4.2.2.2 Stimuli

The stimuli were identical to what were administered to the adults.

4.2.2.3 Procedure

Collecting perceptual (pupillary response) data

Children were told that they were to watch a short movie, during which they should sit still and as a reward, they could choose a booklet afterwards. After obtaining assent from the children and informed consent from the caregiver, children were seated in their caregiver's lap and positioned such that their eyes were approximately 60 cm from the computer screen. The remainder of the procedure was identical of the adult participants.

Upon completion of the task, caregivers were asked to complete a questionnaire in order to estimate the child's vocabulary size and their familiarity with the experimental words. The questionnaire comprised the FRAKIS word list (Szagun et al., 2009), 3 additional critical items plus 9 filler items (altogether 612 words). The questionnaire took approximately 20 minutes to complete.

Collecting production data

The elicitation task that immediately followed the perceptual task was presented in a game form. Children were shown each picture that they were presented with in the perception task and were invited to produce their label twice. Since initial results showed that children's speech tended to be too low-amplitude for useful analysis, they were then additionally encouraged to produce the experimental items in a clear and loud fashion (to a stuffed animal that was purportedly hard of hearing).

4.2.3 Results

4.2.3.1 Pupillary response analysis

In order to ensure that the children knew the words used in the experiment, only those trials that included words (and their mispronunciations) reported to be known in the parental questionnaire were considered in the analysis of the data. Apart from this modification, an analysis identical to the adults' was performed on the children's pupillary responses. Based on the criteria detailed in the adult results section (p. 8.), 5 children were excluded from further analyses. The mean number of successful trials was 16.81 out of 20 ($SD = 1.63$) in the experimental trials and 17.38 out of 20 ($SD = 1.94$) in the filler trials.

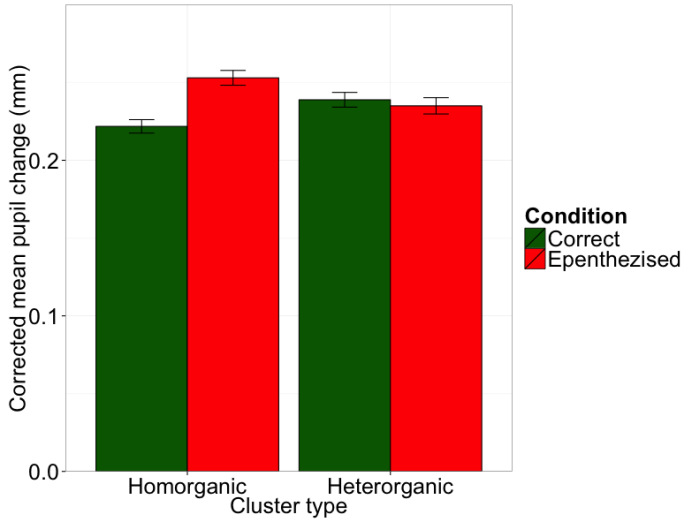

Figure 4.8: Children's mean pupil size change in response to correct and epenthesized versions of homorganic and heterorganic clusters (only words reported to be familiar included). Error bars represent 95% confidence intervals.

Vocabulary size was estimated from the parental questionnaire and was entered as additional control variable into the model. Likelihood Ratio Tests determined that the most parsimonious model included the negative predictor `Neighborhood density` (denser neighborhood was as-

sociated with smaller degree of pupil dilation: β = -0.028, SE = 0.003, t = -7.67), the positive predictor `Positional biphone probability` (larger positional biphone probability was associated with larger degree of pupil dilation: β = 0.048, SE = 0.019, t = 2.48), and `Condition`, i.e., the correct-epenthesized contrast (β = 0.368, SE = 0.047, t = -7.81) as main effects, and the interaction term `Condition` x `Cluster type` (β = -0.214, SE = 0.097, t = -2.17). As expected, the `Condition` x `Cluster type` interaction was such that the contrast between correct and epenthetic items was larger among homorganic clusters than among heterorganic clusters (c.f., Figure 4.8).

Table 4.5: Summary of item-wise analyses on children's perceptual data, organized by cluster type (left: homorganic; right: heterorganic), and degree of pupil dilation by condition (correct < epenthesized; correct ≈ epenthesized; correct > epenthesized). t thresholds were Bonferroni-corrected for multiple comparisons.

	Homorganic items	**Heterorganic items**
Correct < Epenthesized ($t > 2.58$)	grün	blau
	Krabbe	Glas
	Kran	Frosch
	Kröte	
	Stein	
	Stock	
Correct ≈ Epenthesized ($t \leq 2.58$ and $t \geq$ -2.58)	Gras	Blume
	Schnecke	Clown
	Stuhl	Knie
		Flasche
		Fliege
		Schwein
Correct > Epenthesized ($t <$ -2.58)	grau	Brot

Item-wise analyses parallel to the ones on the adult data were performed with children (c.f., Table 4.5). Six homorganic and 3 heterorganic

items were associated with lower degrees of pupil dilation in the correct than in the epenthesized condition. On the other hand, 3 homorganic and 6 heterorganic items were linked to comparable degrees of pupil dilation across the correct and epenthesized conditions (c.f., Table 4.5).

4.2.3.2 Production data analysis

Of the 43 participants whose perceptual data were retained, 4 children's production data could not be included in the production data analysis. 2 children failed to complete the production task due to extreme shyness, 1 was too soft-spoken for the recorder to capture, and 1 had a cold which rendered his speech denasalized. Even though the participants were instructed to produce each item twice, few items were actually repeated by children (39 items – 5.8% – were realized twice without change and 12 items – 1.8% – with different phonological processes. Those items that were repeated as per the instructions, were counted as half in the analyses.) Altogether, 627 target labels were elicited (678 with repetitions), which is 87% of the possible 780 (39 children x 20 items). On average, 16.07 elicitations (*SD = 3.54*) were recorded per children and 31.35 elicitations (*SD = 2.72*) per item. In parallel to our approach in the adults' production data analysis, voicing neutralizations were regarded as part of normal idiolectal variation and thus were coded as 'canonical' (5% of all elicitations). Other simplifications and neutralizations (phonological processes that have no bearing on cluster structure) were categorized as 'other' (22% of all elicitations).

The overall proportion of the types of phonological processes fits well with reports for this age group in the literature for German and English (Fox & Dodd, 1999; Lleó & Prinz, 1996; McLeod et al., 2001; Smit, 1993; Watson & Scukanec, 1997). The breakdown of those proportions by cluster type is shown in Figure 4.9.

Prior to analyzing children's production data, we did not anticipate that coalescence could not surface in fricative-initial homorganic clusters (c.f., Figure 4.11). The only possible simplifying process targeting those clusters is consonant truncation. For example, clusters like /ʃt/ can only be simplified as [ʃ] or [t] as there is no intermediate segment in between fricative-initial homorganic cluster consonants. For this reason, we only assessed the effect of cluster type on phonological processes in stop-initial

clusters.

Table 4.6: χ^2 test statistics on children's production data, including the omnibus 2 x 5 χ^2 test statistic, Fisher exact probability tests (2-tailed), *a priori* 2 x 2 contrasts ransacking and partitioning, categories contrasted, χ^2 test statistic, Fisher exact probability test (1-tailed), odds ratio (OR), and .95 confidence intervals.

Omnibus			χ^2	p			
			13.71	.008			
	Category 1	Category 2	χ^2	p	OR	.95 CI	
						Lower	Upper
Ransacking	*Correct*	*Epenthesis*	2.75	.093	0.28	0.06	1.38
	Correct	*C truncation*	2.25	.067	0.66	0.40	1.09
	Correct	*Coalescence*	2.34	.063	1.78	0.91	3.47
Partitioning	$\overline{\textit{Epenthesis}}$	*Epenthesis*	2.90	.085	0.27	0.06	1.34
	$\overline{\textit{C truncation}}$	*C truncation*	5.94	.007	0.56	0.36	0.88
	$\overline{\textit{Coalescence}}$	*Coalescence*	4.29	.019	1.99	1.08	3.68

After restricting the scope of investigation to stop-initial clusters (c.f., Figure 4.10), three observations can be made. First, heterorganic clusters were a more likely target for epenthesis (3.8% vs. 1.1%) and single consonant truncation (37.4% vs. 24.9%) than homorganic clusters. Second, homorganic clusters tended to undergo more coalescence (17.8% vs. 9.8%) than heterorganic clusters. These observations were statistically confirmed by omnibus 2 x 5 and *a priori* contrast 2 x 2 χ^2 tests (χ^2 statistics, Fisher exact probability tests, odds ratios, and .95 confidence intervals are tabulated in Table 4.6). In parallel to the adult production data analysis, after obtaining a significant omnibus χ^2 test statistic, planned 2 x 2 χ^2 tests were conducted, contrasting all three critical phonological processes – epenthesis, single consonant truncation and coalescence – with either the correct category (ransacking) or with all the other categories collapsed (partitioning), c.f., Sharpe (2015).

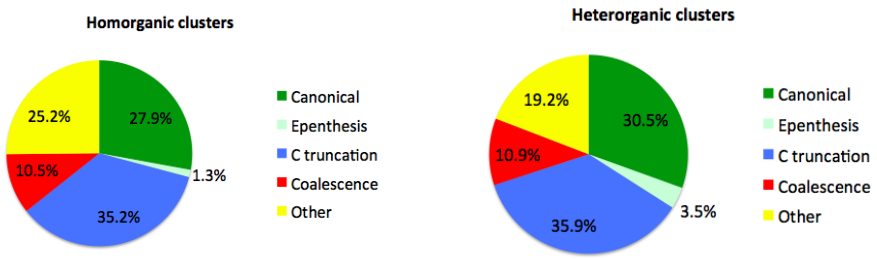

Figure 4.9: Children's production of homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

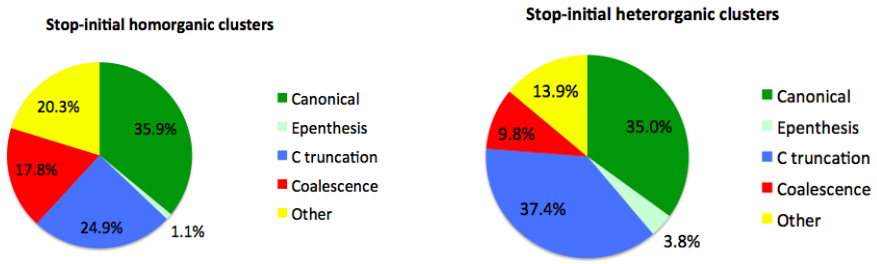

Figure 4.10: Children's production of stop-initial homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

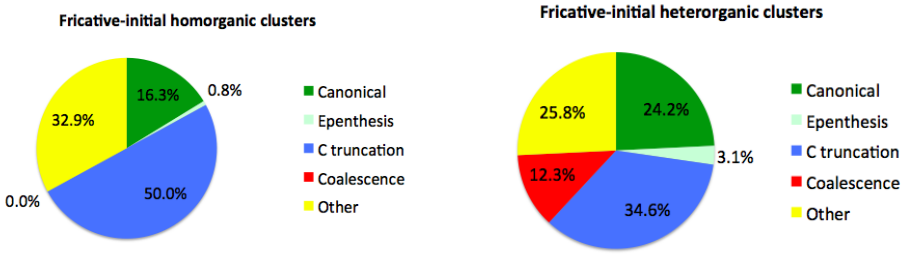

Figure 4.11: Children's production of fricative-initial homorganic (left) and heterorganic clusters (right), categorized by type of phonological process.

4.2.4 Discussion

Our study found larger differences in pupil size between correct and epenthesized homorganic clusters than between correct and epenthesized heterorganic clusters when controlling for a number of potentially confounding sub-lexical and lexical factors. It thus provides evidence that pupillometry is able to capture that children's pupillary response is affected by the cluster type. These results are in line with the hypothesis that homorganic and heterorganic clusters are encoded differently in the mental lexicon.

The production data is also suggestive of a structural difference across homorganic and heterorganic clusters as the break-down of phonological processes associated with homorganic and heterorganic clusters are markedly different. However, it is important to note again that production data, due to its very nature cannot be a definitive indicator of how lexical representations are structured. Nevertheless, we find that in children's production, heterorganic clusters are significantly more prone to epenthesis and single consonant truncation, phonological processes that are consistent with a separate cluster representation, while homorganic clusters are more likely to undergo coalescence, a phonological process consistent with a more unified cluster representation.

4.3 General discussion

In this paper, we proposed a possible mechanism behind the cross-linguistic asymmetry between homorganic and heterorganic clusters, namely that the difference is rooted in the nature of representations. The results of the above presented studies suggest that speaker-listeners possess intricate knowledge about the representation of consonant clusters.

Regarding our hypothesis on structural asymmetry, results obtained across the two age groups were convergent. In the perceptual task, the pupil size difference between correct and epenthesized homorganic clusters was larger than between correct and epenthesized heterorganic clusters, resulting in a `Condition` x `Cluster type` interaction for both adults and children. Moreover, the production data of the two age groups were indicative of a difference between homorganic and heterorganic clusters. In the production task, we studied the question of homorganic/heterorganic contrast by tallying the prevalence of phonological processes that are relevant

for the structure of cluster representations: epenthesis, single consonant truncation, and coalescence (Barton et al., 1980). Findings from this task were concordant with those of the perceptual task: Heterorganic clusters were more likely to undergo phonological processes consistent with a separate cluster representation, a pattern that was observed in both the adult participants' (epenthesis) and the child participants' production (epenthesis plus single consonant truncation). In turn, homorganic clusters were more likely to be produced with coalescence, a phonological process consistent with a unified cluster representation, as suggested by children's production data. Such findings are consistent with our hypothesis that homorganic clusters form a unit whose interruption via epenthesis is more noticeable than inserting a schwa in presumably separately represented heterorganic clusters.

The effect of homorganicity was obtained while attempting to control for potentially confounding factors (identified by previous studies on cluster processing: position, prosody, sonority, phonetics and other (sub-)lexical factors) as much as possible. Concerning the item-wise analyses conducted on the perceptual data of adults and children, it is worth noting that no confound specific to the phonetic and phonological characteristics of the items was identified. First, syllable number and manner of the first consonant cut across the outcome categories in both age groups (i.e., there was no one-to-one correspondence between a characteristic such as monosyllabicity and an outcome category such as 'correct < epenthesized'). Second, multiple instantiations of the same cluster (e.g., /ʃt/ in *Stein*, *Stock*, and *Stuhl*) did not occur together in the same outcome category. Third, sonority non-compliant clusters (/ʃt/ and /ʃv/) occur in all outcome categories, suggesting that they do not form a coherent group. Fourth, the overlap between adult and child item-wise summaries was not substantial: only 6 items (*Blume*, *Brot*, *Krabbe*, *Knie*, *Stock*, and *Schwein*) ended up in the same outcome category across the two age groups, a set whose members shared no common characteristics. Thus, the results of the item-wise analyses support the notion that the contrast between homorganic and heterorganic clusters cannot entirely be captured by unrelated phonetic and phonological properties already described in the literature, but is at least in part due to abstract structural differences, i.e., homorganicity.

Recall that in the children's analysis, only those words reported to be

familiar were entered into the model. To further strengthen our conclusions, we employed another exclusion method, enabled by the elicitation task using the same experimental stimuli. When only those words that were produced in any manner by the children in the elicitation task were introduced to the analysis, a similar `Condition` x `Cluster type` interaction was obtained (β = -0.636, SE = 0.32, t = -2.01), which we regard as additional confirmation of children's sensitivity to the structural difference. To our knowledge, no study as of yet has used this method of assessing children's familiarity with experimental items, made possible by the collection of perceptual and production information about the same word from each participant.

Our findings do not address the source of the representational asymmetry between homorganic and heterorganic clusters. According to previous research on clusters, it is likely to have both an articulatory and an acoustic/perceptual basis. First, moving from one articulator to another (as in the production of heterorganic clusters) increases the chance of open transition, possibly coupled with an insertion of a transitional vocoid (Browman & Goldstein, 1990, 1992; Davidson & Stone, 2003; Gafos, 2002; Gafos, Hoole, Roon, & Zeroual, 2010). Second, heterorganicity in a cluster may be a cause for acoustic discontinuity (either due to a presence of an audible release or an intensity rise), hence more epenthesis/insertion is expected in heterorganic clusters from an acoustic phonetic point of view as well (Steriade, 2009; Wilson & Davidson, 2015; Yun, 2012, 2014b). For these reasons, heterorganic clusters may provide a congruent context for epenthesis, while homorganic clusters may be, by and large, incongruent with epenthesis. These possible explanations for the emergence of the homorganic vs. heterorganic contrast suggest themselves to further investigation: Do articulatory and acoustic phonetic properties (e.g., burst intensity and length in stop-initial clusters, length of epenthetic vowel) influence the percept of epenthesis? Furthermore, speaker-listeners may encounter more epenthesized heterorganic clusters (e.g., [səw]ine) in the ambient language than they do epenthesized homorganic clusters (e.g., [sət]one), making the epenthesized version a phonological variant / part of the exemplar of heterorganic clusters, but not of homorganic clusters (Bybee, 2003; Pierrehumbert, 2002). To our knowledge, no studies have investigated the question of frequency thus far, so this too may prove to

be a promising avenue of research.

In conclusion, our study is the first to investigate consonant cluster processing and representation with an online perception task, pupillometry, along with a production task. We sought to determine whether homorganicity influenced adult and child processing of consonant clusters above and beyond other phonetic and phonological factors. In accordance with our predictions, we found that adults' and children's performance amounted to support the claim – originally inspired by typological data – that homorganic clusters were tied closer together in lexical representations than heterorganic ones, from the early stages of cluster acquisition and into adulthood.

Chapter 5

Conclusions and further research questions

This chapter contains a discussion of the studies on the detailedness of lexical representations that were carried out using a single-picture pupillometry paradigm (Chapters 2 and 4), a preferential looking paradigm coupled with pupillometry (Chapter 3), and speech production analysis (Chapter 4). In conducting this research, we aimed to test whether infants were sensitive to sub-phonemic detail and differences in cluster type. We furthermore intended to establish pupillometry as a viable method using two different paradigms to assess word processing and thus broadening the methodological spectrum available to infant language researchers. In the following sections, the major findings and their implications will be revisited and directions for further research will be considered. The appendix evaluates the sensitivity and robustness of the outcome measures of pupil dilation and looking time that were considered for analysis.

5.1 Major conclusions

5.1.1 Sub-phonemic detail encoded in early words

In the introductory chapter, we reviewed the available evidence for sensitivity to the degree of mispronunciation in adult and preschooler populations. Such sensitivity has been interpreted as an indication for the presence of sub-phonemic detail in lexical representations. Then we considered whether gradient sensitivity can be demonstrated for infants given minimally demanding conditions. We discussed studies that used online methodologies to study the granularity of infant lexical representations.

Seemingly contradictory results were obtained, some research finding no evidence for infant gradient sensitivity (Bailey & Plunkett, 2002; Swingley & Aslin, 2002), while more current research did (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005). We argued that methodological advancements that took place in the subsequent studies (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White & Morgan, 2008; White et al., 2005) were responsible for detecting gradient sensitivity. Fuller control over phonetic and phonological variables in the stimuli, the introduction of the fixation point, stricter exclusion criteria, and unfamiliar distractor images were all instrumental in enabling infant gradient sensitivity to surface (c.f., Table 1.1 in the introductory chapter).

Table 5.1: Summary table of studies presented in Chapters 2 and 3.[1]

Study	**Tamási et al. (in press)**	**Tamási et al. (2016a)**
Paradigm	SPPP	IPLP
Age (months)	30	30
Manipulation	Corr / 1F / 2F / 3F	Corr / 1F / 2F / 3F / Nov
Position in word	onset	onset
Stimulus set	featurally balanced	featurally balanced
Stimulus creation	PoA / MoA / V type & direction counterb.	PoA / MoA / V type & direction counterb.
D familiarity	N/A	no
Preset ISI	yes	yes
Fixation point	before blocks & naming	before blocks & naming
Trial structure	4 s, naming at 1 s	7 s, naming at 4 s
Exclusion crit.	familiar words, >50% info in trials, >50% trials	familiar words, naming score, 2 images fixated, >50% info in trials, >50% trials
Analyses	time-course: PD	time-course: PTL & PD
Results	Corr | 1F | 2F, 3F	Corr | 1F | 2F | 3F | Nov (PTL) Corr, 1F, 2F | 3F | Nov (PD)
Interpretation	gradient sensitivity	complete gradient sensitivity (PTL) partial sensitivity (PD)

[1]Abbreviations: IPLP = intermodal preferential looking paradigm, Corr = correct word form, 1-3F = one-three feature change, Nov = novel word form, PoA = place of articulation, MoA = manner of articulation, V = voicing, counterb. = counterbalanced, *D* = distractor, N/A = not applicable, ISI = inter-stimulus interval, PTL = proportion of target looks, *T* = target, SPPP = single-picture pupillometry paradigm, PD = pupil dilation, time-course = cluster-based permutation tests.

Our studies implemented and extended the improvements introduced by White and Morgan (2008). Following the format of the summary table discussed in the introductory chapter, the research presented in Chapters 2 and 3 is summarized in Table 5.1. Sensitivity to the degree of mispronunciation has been demonstrated with a minimally demanding paradigm, single-picture pupillometry (Tamási et al., in press, discussed in Chapter 2). As a direct consequence of adopting this paradigm and using one picture that is always semantically related to the heard input, concerns about visual and lexical competition originating from the distractor picture and/or the novel label became moot, further minimizing memory and attention requirements.

To enhance the generalizability of the finding on gradient sensitivity, we developed a featurally more balanced stimulus set by using vocabulary items familiar to slightly older, 30-month-old children. With this stimulus set, it became possible to counterbalance both feature type and direction of feature change. In addition, cluster-based permutation tests (i.e., time-course analyses) allowed for pinpointing the emergence of effects over the course of the trial.

Recall that none of the previous mispronunciation detection studies have found evidence for complete gradient sensitivity (only a trend in the expected direction was detected by (White & Morgan, 2008) and Tamási et al., in press). Tamási et al. (2016a) used a standard intermodal preferential looking paradigm with the slight modification of the manipulation and stimulus set developed by Tamási et al. (in press) (i.e., the addition of the novel condition, similarly to the design of White & Morgan, 2008). Due to the simultaneous analysis of looking time and pupil dilation information, more stringent exclusion criteria were adopted relative to studies that only used one type of outcome measure. With these modifications and more fine-grained (time-course) analyses, it became possible to detect complete gradient sensitivity in the looking time measure. Time-course analyses were furthermore informative about children's millisecond-by-millisecond preference changes, which enabled capturing the oscillating behavior between target and distractor preference in response to large (two- and three-feature) degrees of mispronunciations. Averaging those preferences over a time window may have concealed the contrast between the two conditions in previous studies (Mani & Plunkett, 2011a; Ren & Morgan, 2011; White

& Morgan, 2008).

Overall, sensitivity to the degree of mismatch has been demonstrated by our studies in Chapter 2 and 3. More specifically, sensitivity to small (one-feature change) and large degrees of mispronunciation (two- and three-feature change) has been obtained by the study presented in Chapter 2 and complete gradient sensitivity (i.e., significant contrasts between one-, two-, and three-feature change) by the study in Chapter 3. We interpreted sensitivity to degree of mismatch to indicate that cognitive effort necessary for the integration with the appropriate lexical representation is affected by the experimental manipulation. As such, we take these findings as support for the notion that infant lexical representations encode fine-grained sub-phonemic detail.

5.1.2 Cluster type encoded in early and mature words

The third study probed the extent of detailedness in early lexical representations containing consonant clusters. Particularly, we addressed the question whether clusters were represented holistically or in a more specific fashion. In doing so, we tested two types thereof, homorganic and heterorganic clusters.

Previous findings from production studies on cluster acquisition showed that clusters are avoided or simplified in the early stages of language development (Barton et al., 1980; Dyson & Paden, 1983; Fox & Dodd, 1999; Lleó & Prinz, 1996; McLeod et al., 2001; Smit, 1993; Stemberger & Treiman, 1986; Watson & Scukanec, 1997). Other offline studies showed that children are unable to consciously access and manipulate clusters and indeed anything below the syllabic level (Gathercole et al., 1991; Treiman, 1983, 1991; Treiman & Cassar, 1996). These findings can be construed as support for the notion that clusters are represented holistically.

However, as pointed out in the introductory chapter, offline studies may underestimate infants' and children's lexical knowledge by assessing children's capabilities on the basis of their consciously generated production output. Early specificity accounts would advocate that detail is present in lexical representations at the initial stages of language acquisition, though it may not be accessible in tasks involving production and metalinguistic knowledge. For this reason, we complemented the previously employed production-based methodologies with single-picture pupil-

lometry, a perception-based minimally demanding method.

Chapter 4 examined adults' and 30-month-old children's sensitivity to the contrast between correct and epenthesized forms of homorganic and heterorganic cluster types. The amount of cognitive resources spent to establish a link between the – correctly or incorrectly pronounced – word form and the lexical representation was measured by the participants' pupillary response. Thus we were able to assess how epenthesis might affect the recognition of words containing clusters.

A differential response was observed in both age groups between correct and epenthesized items, i.e., mispronounced items were associated with a larger degree of pupil dilation than correctly produced items. This result fits with the concept detailed in Section 1.1.2 that the processing of mispronounced items – i.e., matching the input with the appropriate lexical representation – requires more cognitive effort than that of correctly pronounced items. In this case, the ability to distinguish between correct and incorrect representations was evidenced by differential pupillary response (similarly to Fritzsche & Höhle, 2015 and Tamási et al., in press). Critically for this study, we furthermore observed in both age groups a difference in the magnitude of pupil dilation change between correct and epenthesized forms such that the contrast was larger within homorganic clusters (i.e., consonants produced with the same organ) than within heterorganic ones (i.e., consonants produced with a different organ).

Apart from the single-picture pupillometry task, participants were invited to take part in a production task as well. In elicitation speech, heterorganic clusters were more likely to undergo phonological processes than homorganic clusters that are consistent with a representation of a cluster as a sequence of two independent units (evident in both adults' and children's production). In turn, homorganic clusters were more likely to be produced with phonological processes consistent with a more coherent cluster representation (evident in children's production of stop-initial clusters).

Results from the perceptual and production tasks across the two age groups converge to show that homorganic clusters are represented in a more coherent fashion than heterorganic clusters from the earliest stages of cluster acquisition and into adulthood. This suggests that cluster type – homorganicity – affects how clusters are represented not just for mature

language users, but also for infants. Thus, *contra* the holistic and supporting the early specificity account, we are able to conclude that information about cluster type is encoded in lexical representations early on. Based on these findings, it would be premature to answer the question whether clusters are represented as one or two units in lexical representations. At the moment, we only have relational, but not absolute information – homorganic clusters behaving more as units than heterorganic ones.

5.1.3 Summary

In Section 1.1.2, we reviewed how using online methodologies can uncover infant sensitivity to mispronunciations by obtaining a differential response to correctly and incorrectly pronounced words. This suggests that infants, similarly to adults, process correct word forms more efficiently than mispronounced word forms, which is consistent with the encoding of phonemic detail in their lexical representation. Subsequent sections set the ground for the general question whether infants can detect granularity in mispronunciations. That is, whether infants differentially react to relatively small (i.e., one-feature change, epenthesis in heterorganic clusters) vs. to relatively large degrees of mispronunciation (i.e., two- and three-feature change, epenthesis in homorganic clusters).

Throughout the dissertation, we have seen that infant sensitivity is a dynamic phenomenon that may be detected given careful methodological decisions on multiple dimensions including the choice of paradigm (the less cognitively demanding, the greater the chance of success), experimental manipulation, stimuli selection, procedure, exclusion criteria, and type of analysis. The methodological choices made either implicitly or explicitly have a crucial bearing on the experimental outcome sometimes in unforeseen ways (e.g., using an unknown distractor: introducing a plausible referent for mispronunciations). Therefore, the ability to detect children's sensitivity to more nuanced detail – sub-phonemic information and cluster type – is contingent on methodological choices.

Our results, obtained by the online methodologies pupillometry and eye tracking, are consistent with the gradient sensitivity of infants. Overall, we recorded a weaker pupillary response to small compared to large degrees of mispronunciation (Chapters 2, 3, and 4) as well as weaker target looking preference to small than to large degrees of mispronunciation

(Chapter 3). This result can be interpreted such that infants are able to recover the appropriate lexical entry with less cognitive effort when presented with small vs. large degrees of mispronunciation, showing greater flexibility and tolerance towards those forms that better resemble the correct pronunciation.

We argued in Chapter 2 and 3 against the possibility that the results can be accounted for by general surprise. First, it is not clear how surprise would be be able to predict the degree of change in pupil dilation (as opposed to a form-related explanation, i.e., degree of featural change introduced to the onset). Second, incongruous labels themselves do not automatically invoke increased degree of pupillary response. When presented with a label that is semantically unrelated to the pictured referent, infants do not exhibit enhanced pupil dilation as it might be expected if pupil dilation indexed surprise in the task (Fritzsche & Höhle, 2015; Kuipers & Thierry, 2013). Instead, their lack of increased pupillary response can be interpreted such that monolingual infants do not attempt to integrate the heard input with the pictured referent (Fritzsche & Höhle, 2015; Kuipers & Thierry, 2013).

Therefore, the most plausible explanation for infants' performance in our tasks if early words contain information on the sub-phonemic level as well as information beyond the phonemic level. As such, the findings on infant gradient sensitivity are not compatible with the holistic, but instead with the early sensitivity hypothesis. Language acquisition models are thus needed to be updated to reflect and incorporate the experimental findings suggesting that 30-month-old infants are able to encode sub-phonemic detail (Chapters 2 and 3, corroborating previous research) and cluster type in an adult-like fashion (Chapter 4).

5.1.4 Methodological contributions of the dissertation

In the past thirty years or so, 'devious and clever' ways have been developed to study infant language processing (intermodal preferential looking: Golinkoff et al., 1987, head turn preference: Jusczyk & Aslin, 1995, switch: Stager & Werker, 1997, implicit naming: Mani & Plunkett, 2010a). Using the single-picture pupillometry paradigm, gradient sensitivity in infants, and to cluster type in both adults and infants have been detected (Chapter 2 and 4, respectively), building on previous research demonstrating sensitivity to mispronunciation (Fritzsche & Höhle, 2015). Based on these findings, pupillometry in a single-picture pupillometry paradigm has proven to be a sensitive measure and thus provides a viable alternative to other paradigms used in child language research.

Chapter 3 explored whether pupil dilation can be used in the intermodal preferential looking paradigm in conjunction with traditional outcome measures involving looking time. Given that partial sensitivity to sub-phonemic detail was detected in the pupil dilation measure (c.f., Chapter 3), pupillometry has the potential to enrich the intermodal preferential looking paradigm. Although, it seems that more work is needed to increase its efficiency, which probably can be achieved via the introduction of stimulus onset asynchrony between the visual and auditory stimuli in the naming phase. In the future, it is advisable to leave sufficient time for the pupil to adjust to the visual stimuli (700-1000 ms) prior to the presentation of the auditory labels that is the experimental manipulation. This way, one can expect to obtain more robust results in the pupil dilation measure using the intermodal preferential looking paradigm as well.

Time-course analyses allowed the looking time measure in the intermodal preferential looking paradigm to detect fine-grained, complete gradient sensitivity to featural overlap (Chapter 3). Such a result might have been achieved in previous studies (Mani & Plunkett, 2011a; White & Morgan, 2008), had participants' dynamic preference changes been analyzed.

Finally, the fact that differential response to cluster type has been detected in both perceptual and production data (Chapter 4) strengthens the claim that cluster type plays a role in how clusters are structured in lexical representations. Collecting perceptual and production data from the same participants using the same experimental items has great potential in understanding the perception-production link.

5.2 Directions for further research

Naturally, several issues were left unaddressed by the dissertation. As stated in the introductory section and elsewhere, it is yet to be clarified what the precise nature of sub-phonemic detail and representational asymmetry between the different cluster types are. It is unclear whether infant lexical representations encode abstract phonological or acoustic information (or perhaps both). When it comes to manipulating vowels, acoustic distance seems to be a better predictor of infants' performance than featural distance (Mani & Plunkett, 2011a). Quantifying acoustic distance between consonants, however, it not as straightforward as between vowels. A possible way may be to use confusion matrices in order to determine how likely each consonant is confused with one another based on non-linguistic, i.e., offline forced-choice perceptual tasks (Allen, 2005; Christiansen & Greenberg, 2008; Miller & Nicely, 1955; Phatak & Allen, 2007; Phatak, Lovitt, & Allen, 2008). To our knowledge, there is no confusion matrix made for German as of yet. However, once available, it may be possible to estimate to what extent acoustic and/or phonological distance can independently account for infant gradient sensitivity.

On a related note, further analyses are needed to explore the effect of individual type of feature change (place of articulation, manner of articulation, voicing) and the direction of feature change (e.g., from labial to coronal place of articulation or vice versa) on the pupillary response. Tamási et al. (in press) presented in Chapter 2 and Tamási et al. (2016a) presented in Chapter 3 balanced those factors across the stimulus set and were not designed to gauge their effects. However, there is some indication from adult and infant language research that certain feature changes and/or direction thereof may play a prominent role in language processing (e.g., Fikkert, 2010; Fort, Martin, & Peperkamp, 2014; Kirov & Wilson, 2012; Lahiri & Reetz, 2002). Further, little is known about potential synergies and interactions between features (Christiansen & Greenberg, 2008).

Our findings presented in Chapter 4 do not address the source of the representational asymmetry between homorganic and heterorganic clusters. Consonants that share their place of articulation may be represented differently from consonants that do not for the following – potentially overlapping – reasons: (1) *Frequency – asymmetry in realization.* Probably

related to the typological asymmetry, many languages exhibit a preferential distribution of epenthesis in heterorganic (vs. homorganic) clusters. The production output in a given language may contain more heterorganic epenthesized tokens than homorganic ones (though not empirically studied yet, it is consistent with predictions posited by examplarist and frequentist models, e.g., Bybee, 2003; Pierrehumbert, 2002). (2) *Articulation – asymmetry in gestural timing.* The act of engaging two articulators in succession, as in the production of heterorganic clusters, may introduce open transitions and thus transitional vocoids, as opposed to engaging only one articulator, i.e., when producing homorganic clusters (Browman & Goldstein, 1990, 1992; Davidson & Stone, 2003; Gafos, 2002; Gafos et al., 2010). (3) *Perception – asymmetry in acoustic discontinuity.* Heterorganic clusters are more likely to contain acoustic discontinuity, which may be reinterpreted as epenthesis by the listener (Steriade, 2009; Wilson & Davidson, 2015; Yun, 2012, 2014b).

As reviewed in the introductory chapter, lexical restructuring accounts posit that specificity in early lexical representations emerge dynamically, mainly driven by two interdependent processes: the steady expansion of vocabulary and the gradual increase of speech processing efficiency (Garlock et al., 2001; McKean, Letts, & Howard, 2013; Pierrehumbert, 2002; Werker & Curtin, 2005). Converging evidence from these studies suggests that, as vocabulary grows, neighborhood density, phonotactic probability and word frequency may become factors that determine the specificity of lexical representations. The acquisition of vocabulary may necessitate more efficient ways to differentiate words from each other: In particular, dense lexical neighborhood and high phonotactic probability have been proposed as facilitators for the emergence of sub-lexical, i.e., syllabic and, subsequently, phonemic structure (De Cara & Goswami, 2003; Garlock et al., 2001; Hollich et al., 2002; Hoover et al., 2012; Stokes, 2010, 2013; Storkel, 2002, 2009) along with word familiarity, age of acquisition and frequency (Barton, 1976; Barton et al., 1980; Goodman et al., 2008).

Table 5.2: Summary table of lexical and sub-lexical factors in studies included in the dissertation.[1]

Study	Tamási et al. (in press)	Tamási et al. (2016a)		Tamási et al. (2016b)	
Chapter	Chapter 2	Chapter 3		Chapter 4	
DV	PD	PTL	PD	PD	PD
Age	infant	infant	infant	infant	adult
Vocab	(+)	n.s.	+	n.s.	N/A
LogFreq	n.s.	n.s.	n.s.	n.s.	n.s.
ND	−	n.s.	n.s.	−	−
PP	n.s.	(+)	n.s.	+	+

Table 5.2 provides a summary of how lexical factors contributed to model fit in each of the studies included in the dissertation. Since only models on perception studies from Chapter 4 included lexical and sublexical factors, models on production studies are not listed in the table. Even though the present experiments were not designed to evaluate those factors by employing them as control variables only, some considerations are offered below. The studies presented in Chapter 4 provided no evidence for a qualitative change over development in how words are represented as the same lexical effects were found in children and adults. In other online studies in which age has been included as a factor, mixed results were obtained, some finding correlation between age and performance (Mani & Plunkett, 2007; Swingley & Aslin, 2000) while others not (Bailey & Plunkett, 2002; Mani & Plunkett, 2010b; Werker et al., 2002). It is possible that critical changes in word recognition take place at earlier stages of lexical development (c.f., Werker et al., 2002; Zesiger et al., 2011).

In most of the statistical models, vocabulary size has not been found to be a significant predictor of 30-months-olds infants' performance (only in the study presented in Chapter 3, vocabulary size positively predicted

[1]Abbreviations: DV = dependent/outcome variable, PD = pupil dilation, PTL = proportion of target looking, Vocab = vocabulary size, ND = neighborhood density, PP = phonotactic probability (positional biphone probability in Chapter 4), LogFreq = logged frequency, − = negative predictor, + = positive predictor, (+) = marginally positive predictor, n.s. = not significant, N/A = not applicable.

pupil dilation – though not looking preference and in the study presented in Chapter 2 it was a marginally significant positive predictor). Finding no effect or only a weak effect is consistent with Werker et al. (2002) that reported vocabulary size to be correlated with the head turn preference of younger (14-month-old), but not older (20-month-old) infants. Most online studies, however, reported no correlation between looking preference and (receptive or productive) vocabulary size (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley, 2005; Swingley & Aslin, 2000; Zesiger et al., 2011).

Likewise, (logged) word frequency did not appear to contribute to the model fit of our studies, neither for 30-month-old infants, nor for adults. This finding is compatible with Bailey and Plunkett (2002) that reported no frequency and age of acquisition effect in the performance of 18- and 24-month-olds.

Neighborhood density negatively predicted pupil dilation in three out of four models, i.e., words in denser neighborhoods were associated with smaller pupil size in both age groups, hinting at more efficient processing (the fifth looking time model in Chapter 3 was not significant). This result may be accommodated in the lexical restructuring literature as better performance is expected from preschooler-aged children with words that have been pressured to be represented in finer detail, that is, words from dense neighborhoods (c.f., Garlock et al., 2001; Storkel, 2009; Walley et al., 2003, but see Vitevitch & Luce, 1999 and Chen & Mirman, 2012 for contradictory findings with adults). Previous online studies, however, found no statistical relationship between performance and neighborhood density (Bailey & Plunkett, 2002; Swingley & Aslin, 2002).

On the other hand, phonotactic probability tended to positively predict pupil dilation, i.e., words with frequently occurring clusters were associated with larger pupil size, suggesting increasing cognitive demand. It may be expected that neighborhood density and phonotactic probability exert opposite effects on preschool-aged (Storkel, 2009) and adult spoken word recognition (Luce & Large, 2001; Vitevitch & Luce, 1999). Although the direction of the effects cannot be easily accounted for, especially that of positional biphone probability. Moreover, no developmental change regarding lexical factors from infants to adults was observed (i.e., the same effects of neighborhood density and phonotactic probability were obtained

regardless of age).

Overall, divergent outcomes have been observed regarding the effects of most lexical factors, the reason of which remains largely unclear. Future work needs to evaluate the possibility that some lexical effects might be transient and only apply to the first stage of word learning process, before the vocabulary spurt (up until 18 months of age) (Werker et al., 2002; Zesiger et al., 2011). Naturally, our results cannot speak to the earliest stages of language acquisition. Therefore, future studies should tease apart the individual and joint effects of age and lexical factors on infant and adult word processing, including possible interactions for instance between neighborhood density and phonotactic probability (c.f., Luce & Large, 2001). A closer look at those factors using online methodologies may shed more light on how detail in lexical representations emerge.

Appendix

This section offers a review of how sensitive different outcome measures were to our experimental manipulation in the task using the intermodal preferential looking paradigm presented in Chapter 3. Apart from assessing variables in looking preference (proportion of target looking time, latency of first look towards the target/distractor, longest look towards the target/distractor), measures derived from pupil dilation (automatically collected during eye-tracking sessions): mean and peak dilation, peak velocity of the pupil, and associated latency values were considered. Critical time windows, baseline-correction, smoothing, and exclusion criteria; practices that have critical bearing on the outcome, are discussed.

Proportion of target looking time is the most widely reported measure in tasks employing the intermodal preferential looking paradigm (Golinkoff et al., 2013). It is calculated by dividing the total looks towards the target picture by the total looks towards target and distractor in the naming phase. The salience phase can be used to gauge the baseline preferences of children looking at the familiar vs. the novel image without any labels. Luche et al. (2015) makes the case that proportion of looking time should be corrected for those baseline preferences by subtracting the proportion of target looks in the salience phase from the proportion of target looks in the naming phase. For this reason, the experiment presented in Chapter 3 reports proportion of target looking time baseline-corrected for proportion of target looking time in the salience phase (also following White & Morgan, 2008). This measure is predicted to be proportional to the strength of the association between the heard input and the picture, mediated by the activation of the respective lexical representation (Golinkoff et al., 2013). For example, the more looks toward the picture of a baby are registered in

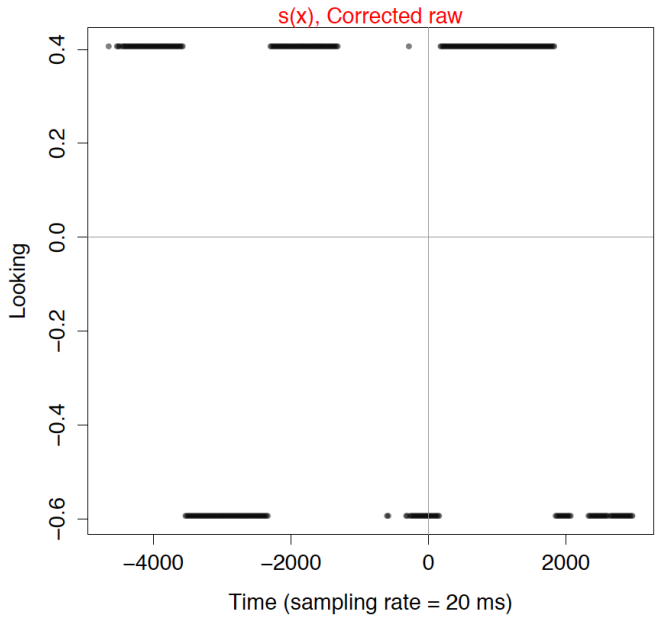

Figure 5.1: Corrected looking preferences over time in a randomly selected correct trial. Time (x-axis) is zeroed on the auditory label onset, looking preferences (y-axis) are baseline-corrected for the proportion of target looks in the salience phase. Positive values indicate looking towards the target, negative ones looking towards the distractor.

response to the heard input *vaby*, the stronger link between the heard input and target picture can be assumed, caused by activation of the lexical representation <baby>. A randomly selected correct trial whereby corrected looking preferences are plotted over time is shown in Figure 5.1. In this figure, target looking times are shown in the upper portion, distractor looking times in the lower portion. The values are baseline-corrected by subtracting the proportion of target looking time.

Latency of first look towards the target was used in the original intermodal preferential looking paradigm (Golinkoff et al., 1987) as an outcome variable and it is also widely reported. The prediction is that the faster children can reliably orient towards the target picture, the faster and stronger the lexical access takes place (Golinkoff et al., 2013). Since

our task reported in Chapter 3 included a novel label that could not be readily matched with the target picture, latency of first look towards the distractor was also included as an outcome measure in our investigation. If infants were to exhibit distractor preference while presented with a novel or mispronounced auditory label, that would indicate that infants have fast-mapped the input to be the label of the distractor picture (via the mutual exclusivity mechanism discussed in the introductory chapter).

Longest look towards the target is the longest uninterrupted stretch of time of target looks in response to the auditory label. Similarly to proportion of target looks and latency of first look towards target, it has been hypothesized that longest look can be used as a proxy of the strength of the match between heard input and target image (Luche et al., 2015). Again in line with the latency measures, a longest look towards the distractor is discussed as a measure.

Linear contrasts have been included in linear mixed effects models, the setup of which has been detailed in Chapter 3. All other contrasts were evaluated with post-hoc cluster-based permutation tests with Bonferroni-corrections for multiple comparisons, also discussed in Chapter 3. These models were run with different outcome measures, five related to looking time and five related to pupil dilation. The significance table of looking time-related outcome measures is included in Table 5.3.

Overall, proportion of target looking time proved to be by far the most sensitive measure as all of the examined contrasts were significant. The other four outcome measures altogether yielded five marginally significant contrasts (longest look toward the target: 1, longest look toward the distractor: 2, latency to first look toward target: 1, latency of first look toward distractor: 1).

Regarding pupillary response, the average pupil dilation is by far the most reported measure, either raw or uncorrected (Beatty & Lucero-Wagoner, 2000). In line with the suggestion to baseline-correct the proportion of target looking time, baseline-correction is suggested for pupil dilation as well especially when luminance changes are to be controlled for. In our experiments presented in Chapters 2, 3, and 4, we employed trial-wise baseline correction for this reason. In each trial, the values from the time interval immediately preceding the naming phase were collected and were subtracted from the raw values of pupil size. As varying the time

Table 5.3: Significance table of contrasts relevant to the study reported in Chapter 3. The looking time outcome measures are proportion of target looking time (row 1), longest look toward target (row 2), longest look toward distractor (row 3), latency to first look towards target (row 4), and latency to first look towards distractor (row 5). Abbreviations: Lin = linear contrast, C = correct condition, Δ1-3 = one-to-three-feature change, N = novel condition, C/N = contrast between the correct and novel conditions.

	Lin	C/N	C/Δ3	Δ1/N	C/Δ2	Δ1/Δ3	Δ2/N	Δ3/N
Prop. of T look	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆
Longest T look		†						
Longest D look		†		†				
Latency to $T1$			†					
Latency to $D1$						†		

†: $p < .1$, *: $p < .05$

interval (20 ms, 100 ms, 500 ms pre-onset) produced comparable results, we chose the medium interval 100 ms to be the baseline. In order to better understand the outcome measures in pupil dilation, the same trial whose looking preferences were shown above in Figure 5.1 is shown, this time which the pupillary response over time in Figure 5.2.

The first plot in Figure 5.2 shows the change of corrected raw pupil size over time. Time on the x-axis is zeroed on the auditory label onset in all plots. Raw pupil values on the y-axis are baseline-corrected by substracting the average of values in the 100 ms pre-onset time interval. Note the sparse series approximately 200 nm below the dense one. Upon checking its prevalence, we find it is systematically present in all of our recordings regardless of age and experimental manipulation. It seems to be synchronous in the two eyes. For these reasons, it is assumed by an measurement artefact introduced by the eye tracker. Due to its uniformity and predictability, we excluded the sparse series from analysis by filtering data points that are 0.05 mm smaller than the preceding value in the series. The filtered values of pupil size change over time are included in the second plot of Figure 5.2.

In order to calculate the peak value of pupil dilation, smoothing was employed. The smoothing function `smooth.spline` from the basic `R` package (R Core Team, 2014) was specified with the default features: smoothing parameter – [-1.5, 1.5], absolute precision – 0.0004, relative precision – .08, maximal number of iterations – 500, smoothing method = *gam*. After inspecting the evolution of the pupillary curve in individual trials (c.f., Plot 3 of Figure 5.2), we operationalized finding the peak dilation to be used in the analysis as follows. The first local maximum that exceeded 80% of the absolute maximum dilation was chosen as the peak dilation point. If no dilation point met this criterion among the first three local maxima, the largest one of the three maxima was chosen. The peak dilation point is shown in blue in Plot 3 of Figure 5.2. The y coordinate of the peak dilation point corresponds to peak dilation in mm, the x coordinate to the latency of peak dilation in ms.

Velocity measures are typically used in articulatory phonology, for instance to characterize tongue movements (Guenther, 1995). In the case of pupillary response, it may be hypothesized that the velocity with which the pupil reacts to a stimuli has a relationship with cognitive effort. Ve-

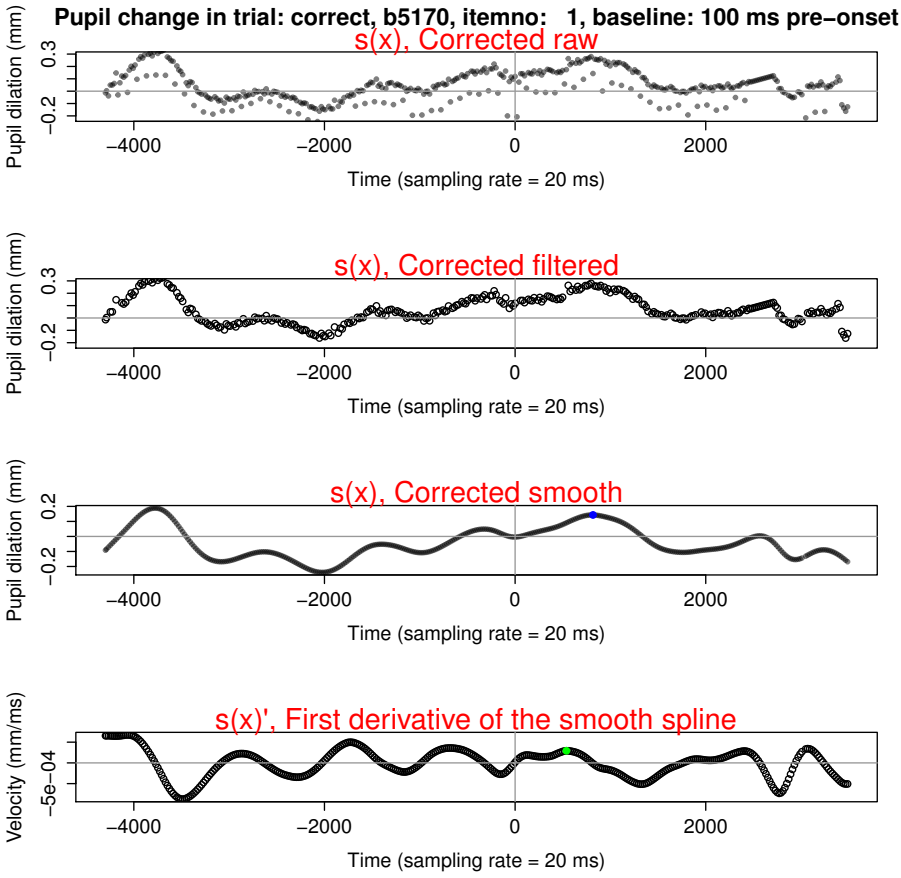

Figure 5.2: Corrected pupil size change over time in a randomly selected correct trial. Time (x-axis) is zeroed on the auditory label onset in all plots. Plot 1: Raw pupil values (y-axis) are zeroed on the baseline-correction of pupil dilation values in the time interval 100 ms pre-onset. Plot 2: Filtered pupil values, baseline-correction same as in Plot 1. Plot 3: Smoothed pupil values, baseline-correction same as in Plot 1. Blue dot shows peak pupil dilation. Plot 4: First derivative of the smoothed pupil values, baseline-correction same as in Plot 1. Green dot shows peak velocity of the pupil.

Table 5.4: Significance table of contrasts relevant to the study reported in Chapter 3. The pupil dilation outcome measures are mean pupil dilation (row 1), peak pupil dilation (row 2), latency to peak pupil dilation (row 3), peak velocity of pupil (row 4) and latency to peak velocity of the pupil (row 5). Abbreviations: Lin = linear contrast, C = correct condition, Δ1-3 = one-to-three-feature change, N = novel condition, C/N = contrast between the correct and novel conditions.

	Lin	C/N	C/Δ3	Δ1/N	C/Δ2	Δ1/Δ3	Δ2/N	Δ3/N
Mean dilation	⋆	⋆	†	⋆			⋆	
Peak dilation	†	⋆						
Latency to peak								
Peak velocity								
Latency to peak v.								

†: $p < .1$, *: $p < .05$

locity was calculated by taking the first derivative of the smooth spline generated from the pupil dilation values using the `predict` function in `R` (R Core Team, 2014). Plot 4 Figure 5.2 shows the velocity change over time in response to the experimental manipulation. The green dot in Plot 4 indexes peak velocity point of the pupil, the y coordinate of which corresponding to the peak velocity in mm/ms and the x coordinate the latency to peak velocity in ms.

The significance table of pupil-dilation related outcome measures is summarized in Table 5.4. Similarly to looking time measures, models of the most widely reported outcome yielded the most significant contrasts. Models with mean pupil dilation contained four significant contrasts, the linear contrast and the ones between the correct and novel, one-feature change and novel, and two-feature change and novel conditions and one marginal contrast, the one between the correct and three-feature change conditions. One marginal and one significant contrast was observed with the peak pupil dilation measure. There were no significant contrasts found in the two peak velocity measures, peak velocity of the pupil and latency

to peak velocity.

Calculations of these alternative measures proved to be less sensitive than measures of central tendency. Among the considered looking time-related measures, proportion of target looking time yielded the largest number of significant contrasts, and thus was found to be the most sensitive. Second, among the pupil dilation-related measures, mean pupil dilation was deemed the most sensitive. Both of these measures are popular ways of characterizing the outcome in preferential looking paradigms and in pupillometry studies, respectively. It is possible that for the other measures under consideration more filtering is warranted to be able to detect the effect of the experimental manipulation.

References

Allen, J. B. (2005). Consonant recognition and the articulation index. *The Journal of the Acoustical Society of America*, *117*(4), 2212–2223. doi: 10.1121/1.1856231

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. doi: 10.1037/e536982012-548

Althaus, N., & Plunkett, K. (2015). Timing matters: The impact of label synchrony on infant categorisation. *Cognition*, *139*, 1–9. doi: 10.1016/j.cognition.2015.02.004

Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science*, *16*(6), 980–990. doi: 10.1111/desc.12071

Archangeli, D. B. (1984). *Underspecification in Yawelmani phonology and morphology* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*(1), 63–80. doi: 10.1016/j.jecp.2009.10.002

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Baayen, R. H. (2008). *Analyzing linguistic data.* Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi: 10.1016/j.jml.2007.12.005

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282. doi: 10.1016/s0885-2014(02)00116-8

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173. doi: 10.1017/s0305000904006567

Barlow, J. A. (2003). Asymmetries in the acquisition of consonant clusters in Spanish. *The Canadian Journal of Linguistics*, *48*(2), 179–210. doi: 10.1353/cjl.2004.0024

Barlow, J. A. (2005). Sonority effects in the production of consonant clusters by Spanish-speaking children. In *Selected proceedings from the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 1–14).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Barton, D. (1976). Phonemic discrimination and the knowledge of words in children under three years. *Papers and Reports on Child Language Development (Linguistics, Standford University)*, *11*, 61–68.

Barton, D., Miller, R., & Macken, M. A. (1980). Do children treat clusters as one unit or two? *Papers and Reports on Child Language Development*, *18*, 105–137.

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27–30.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lme4` (R-package-version 1.1-6)

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., Vol. 2, pp. 142–162). Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511546396

Berent, I., Harder, K., & Lennertz, T. (2011). Phonological universals in early childhood: Evidence from sonority restrictions. *Language Acquisition*, *18*(4), 281–293. doi: 10.1080/10489223.2011.580676

Blount, B. G., & MacKay, D. G. (1991). The organization of perception and action: A theory for language and other cognitive skills. *Language*, *67*(1), 187. doi: 10.2307/415576

Boersma, P., & Weenink, D. (2011). *Praat: doing phonetics by computer (Version 5.2. 23)[Computer program]. Retrieved May 4, 2011.*

Booij, G. (1999). *The phonology of Dutch* (Vol. 5). Oxford, UK: Oxford University Press. doi: 10.2307/416139

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. doi: 10.1111/j.1469-8986.2008.00654.x

Bradley, T. G. (2006). Spanish complex onsets and the phonetics-phonology interface. In F. Martınez-Gil & S. Colina (Eds.), *Optimality-Theoretic Studies in Spanish Phonology* (pp. 15–38). John Benjamins, Amsterdam. doi: 10.1075/la.99.02bra

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 341–376. doi: 10.1017/cbo9780511627736.019

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155–180. doi: 10.1159/000261913

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511612886

Chambless, D. (2004). Asymmetries in initial and medial cluster acquisition. In *Proceedings of the 28th Annual Boston University Conference on Language Development (BUCLD), ed. by Alejna Brugos, Linnea Micciula, and Christine Smith* (pp. 86–97).

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205–215. doi: 10.1017/s0305000900013180

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*, 727–735. doi: 10.1017/s0305000900010023

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological review*, *119*(2), 417. doi: 10.1037/a0030049

Chomsky, N. (1964). *Current issues in linguistic theory. The structure of language, ed. by J. A. Fodor and J. J. Katz, 50-118.* Englewood Cliffs, NJ: Prentice-Hall. doi: 10.1075/cilt

Christiansen, T. U., & Greenberg, S. (2008). Cross-spectral synergy and consonant identification. *The Journal of the Acoustical Society of America*, *123*(5), 3850. doi: 10.1121/1.2935680

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*(01), 225–252. doi: 10.1017/s0952675700000440

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44–56.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*(4), 463–480. doi: 10.1006/jmla.1997.2535

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3-5-year-olds. *Developmental Science*, *15*(5), 697–713. doi: 10.1111/j.1467-7687.2012.01173.x

Croft, W., & Vihman, M. (2003). Radical templatic phonology and phonological development. In C. Goddard & P. Lee (Eds.), *Handbook of cognitive linguistics.* Oxford, UK: Oxford University Press. doi: 10.1515/9783110292022

Curtin, S., & Archer, S. L. (2015). Speech perception. In E. L. Bavin & L. R. Naigles (Eds.), *The Cambridge Handbook of Child Language* (pp. 137–158). Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9781316095829.007

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–731. doi: 10.1111/j.1467-7687.2009.00814.x

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. In *Handbook of psycholinguistics* (pp. 249–283). Elsevier BV. doi: 10.1016/b978-012369374-7/50009-2

Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in

consonant cluster production: an ultrasound study. In *Proceedings of the West Coast Conference on Formal Linguistics* (Vol. 22, pp. 165–178).

de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, *16*(1), 1–17. doi: 10.1017/s0305000900013404

De Cara, B., & Goswami, U. (2003). Phonological neighbourhood density: Effects in a rhyme awareness task in five-year-old children. *Journal of Child Language*, *30*(03), 695–710. doi: 10.1017/s0305000903005725

de Haan, M. (2007). *Infant EEG and event-related potentials*. Informa UK Limited. doi: 10.4324/9780203759660

de Lacy, P. V. (2002). *The formal expression of markedness* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Demuth, K., & McCullough, E. (2009). The longitudinal development of clusters in French. *Journal of Child Language*, *36*(02), 425–448. doi: 10.1017/s0305000908008994

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*(41), 16027–16031. doi: 10.1073/pnas.0705270104

Dink, J., & Ferguson, B. (2016). eyetrackingR [Computer software manual]. Retrieved from `http://www.eyetracking-R.com` (R package version 0.1.6)

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, *42*(02), 447–465. doi: 10.1017/s0305000914000063

Dyson, A. T., & Paden, E. P. (1983). Some phonological acquisition strategies used by two-year-olds. *Communication Disorders Quarterly*, *7*(1), 6–18. doi: 10.1177/152574018300700102

Eilers, R. E., & Oller, D. K. (1976). The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, *3*(03), 319–329. doi: 10.1017/s0305000900007212

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306. doi: 10.1126/science.171.3968.303

Ellis, N. C. (2008). Implicit and explicit knowledge about language. In *Encyclopedia of language and education* (pp. 1878–1890). Springer. doi: 10.1007/978-0-387-30424-3_143

Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language*, *90*(1-3), 378–392. Retrieved from `http://dx.doi.org/10.1016/s0093-934x(03)00449-8` doi: 10.1016/s0093-934x(03)00449-8

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*(5), 1376–1383. doi: 10.1111/j.1467-8624.2010.01479.x

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*(2-3), 245–264. doi: 10.1177/00238309030460020901

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. *Invariance and Variability in Speech Processes*, *1*, 36–51.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439. doi: 10.2307/412864

Fernald, A., McRoberts, G. W., & Swingley, D. (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn & B. Höhle (Eds.), *Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition. Volume 1* (pp. 97–123). John Benjamins Publishing Company. doi: 10.1075/lald.23.08fer

Fikkert, P. (1994). *On the acquisition of prosodic structure* (Unpublished doctoral dissertation). Radboud University Nijmegen.

Fikkert, P. (1995). Models of acquisition: How to acquire stress. In *Proceedings of North Eastern Linguistic Society 25, GLSA, University of Massachusetts.*

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory phonology* (Vol. 10, pp. 227–260). Berlin, Germany: de Gruyter. doi: 10.1515/9783110224917.3.227

Fischer, J. L. (1965). The stylistic significance of consonantal sandhi in

Trukese and Ponapean. *American Anthropologist*, *67*(6), 1495–1502. doi: 10.1525/aa.1965.67.6.02a00090

Forster, K. (1989). Basic issues in lexical processing. In *Lexical representation and process.* (pp. 15–38). Cambridge, MA: MIT Press.

Fort, M., Martin, A., & Peperkamp, S. (2014). Consonants are more important than vowels in the Bouba-kiki effect. *Language and Speech*, *58*(2), 247–266. doi: 10.1177/0023830914534951

Fox, A. V., & Dodd, B. J. (1999). Der Erwerb des phonologischen Systems in der deutschen Sprache. *Sprache-Stimme-Gehör*, *23*(4), 183. doi: 10.1055/s-00000082

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*(1-2), 1–20. doi: 10.1016/0010-0277(87)90002-3

Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): Data from American English and Mexican Spanish infants. *Journal of Child Language*, *35*(01), 77–98. doi: 10.1017/s0305000907008264

Fritzsche, T., & Höhle, B. (2015). Phonological and lexical mismatch detection in 30-month-olds and adults measured by pupillometry. *Proceedings of ICPhS XVIII*.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269–337. doi: 10.1023/a:1014942312445

Gafos, A. I., Hoole, P., Roon, K., & Zeroual, C. (2010). Variation in overlap and phonological grammar in Moroccan Arabic clusters. *Laboratory Phonology X, Mouton de Gruyter, Berlin/New York*, 657–698. doi: 10.1515/9783110224917.5.657

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468–492. doi: 10.1006/jmla.2000.2784

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. doi: 10.1016/s0010-0285(02)00003-8

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's rep-

etition of nonwords. *Applied Psycholinguistics*, *12*(03), 349. doi: 10.1017/s0142716400009267

Gerken, L., Murphy, W. D., & Aslin, R. N. (1995). Three- and four-year-olds' perceptual confusions for spoken words. *Perception & Psychophysics*, *57*(4), 475–486. doi: 10.3758/bf03213073

Gierut, J. A. (1999). Syllable onsets, clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research*, *42*(3), 708–726. doi: 10.1044/jslhr.4203.708

Goad, H., Rose, Y., Kager, R., Pater, J., & Zonneveld, W. (2004). Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. *Constraints in phonological acquisition*, 109–157. doi: 10.1017/cbo9780511486418.005

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. doi: 10.1037/0033-295x.105.2.251

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518. doi: 10.1016/0749-596x(89)90009-0

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1211–1238. doi: 10.1037/0278-7393.18.6.1211

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic analysis*, *2*, 23–68.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. doi: 10.1017/s030500090001271x

Golinkoff, R., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: what have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339. doi: 10.1177/1745691613484936

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count?

parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(03). doi: 10.1017/s0305000907008641

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283. doi: 10.3758/bf03204386

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594–621. doi: 10.1037/0033-295x.102.3.594

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34. doi: 10.1016/s0010-0277(02)00186-5

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, *23*(03), 387–429. doi: 10.1017/s0952675706000996

Halle, M., & Vergnaud, J.-R. (1980). Three dimensional phonology. *Journal of Linguistic Research*, *1*(1), 83–105.

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, *19*(4), 463–481. doi: 10.1016/s0163-6383(96)90007-7

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, *14*(4), 439–456. doi: 10.1080/15250000902996532

Hayes, B. (1986). Assimilation as spreading in Toba Batak. *Linguistic Inquiry*, *17*(3), 467–499.

Hebden, J. C. (2003). Advances in optical imaging of the newborn infant brain. *Psychophysiology*, *40*(4), 501–510. doi: 10.1111/1469-8986.00052

Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology*, *115*(1), 16–29. doi: 10.1016/j.jecp.2012.12.013

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development*, *17*(3), 359–377. doi: 10.1080/15248372.2015.1135801

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350. doi: 10.1126/science.132.3423.349

Hewlett, N. (1990). Processes of development and production. In *Devel-*

opmental speech disorders (pp. 15–38). Edinburgh, UK: Churchill Livingstone. doi: 10.1017/s003329170003720x

Hickey, R. (1985). The interrelationship of epenthesis and syncope: evidence from Dutch and Irish. *Lingua*, *65*(3), 229–249. doi: 10.1016/s0024-3841(85)90199-8

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning. In *Becoming a word learner: A debate on lexical acquisition* (pp. 136–164). Oxford University Press (OUP). doi: 10.1093/acprof:oso/9780195130324.003.006

Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science*, *25*(11), 2038–2046. doi: 10.1177/0956797614547918

Höhle, B., van de Vijver, R., & Weissenborn, J. (2006). Word processing at 19 months and its relation to language performance at 30 months: A retrospective analysis of data from German-learning children. *International Journal of Speech-Language Pathology*, *8*(4), 356–363. doi: 10.1080/14417040600970614

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2002). Lexical neighborhood effects in 17-month-old word learning. In *Proceedings of the 26th annual Boston University Conference on Language Development* (Vol. 1, pp. 314–23).

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye-tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press. doi: 10.1080/17470218.2015.1098709

Hoover, J. R., Storkel, H. L., & Rice, M. L. (2012). The interface between neighborhood density and optional infinitives: Normal development and Specific Language Impairment. *Journal of Child Language*, *39*(4), 835. doi: 10.1017/s0305000911000365

Houston, D. M., Stewart, J., Moberly, A., Hollich, G., & Miyamoto, R. T. (2012). Word learning in deaf children with cochlear implants: Effects of early auditory experience. *Developmental Science*, *15*(3), 448–461. doi: 10.1111/j.1467-7687.2012.01140.x

Ingram, J., Pittarn, J., & Newman, D. (1985). Developmental and sociolinguistic variation in the speech of Brisbane schoolchildren. *Australian Journal of Linguistics*, *5*(2), 233–246. doi: 10.1080/

07268608508599346

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. *Developmental Science*, *12*(4), 670–679. doi: 10.1111/j.1467-7687.2008.00805.x

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*(2), 281–320. doi: 10.1515/lity.2011.021

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, *19*(02), 189–218. doi: 10.1017/s0952675702004311

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. *Phonological Development: Models, Research, Implications*, 17–64.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21(1–2)*, 3–28.

Jusczyk, P. W. (1997). The beginnings of word recognition in infancy. *The Journal of the Acoustical Society of America*, *102*(5), 3189. doi: 10.1121/1.420872

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. doi: 10.1006/cogp.1995.1010

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, *21*(1), 50–54. doi: 10.3758/bf03199467

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 56–88. doi: 10.1080/14639220210159735

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*. doi: 10.1126/science.154.3756.1583

Kahneman, D., Tursky, B., Shapiro, D., & Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology*, *79*(1, Pt.1), 164–167. doi: 10.1037/h0026952

Karatekin, C. (2004). Development of attentional allocation in the dual

task paradigm. *International Journal of Psychophysiology*, *52*(1), 7–21. doi: 10.1016/j.ijpsycho.2003.12.002

Karatekin, C. (2007). Eye-tracking studies of normative and atypical development. *Developmental Review*, *27*(3), 283–348. doi: 10.1016/j.dr.2007.06.006

Kiparsky, P., & Menn, L. (1977). On the acquisition of phonology. *Language learning and thought*, *47778*.

Kirk, C. (2008). Substitution errors in the production of word-initial and word-final consonant clusters. *Journal of Speech, Language, and Hearing Research*, *51*(1), 35–48. doi: 10.1044/1092-4388(2008/003)

Kirk, C., & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language*, *32*(04), 709–734. doi: 10.1017/s0305000905007130

Kirov, C., & Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 587–592).

Klatte, M., Lachmann, T., & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise Health*, *12*(49), 270. doi: 10.4103/1463-1741.70506

Klingner, J. (2010a). Fixation-aligned pupillary response averaging. In *Proceedings of the 2010 Symposium on Eye-Tracking Reseach & Applications* (pp. 275–282). doi: 10.1145/1743666.1743732

Klingner, J. (2010b). *Measuring cognitive load during visual tasks by combining pupillometry and eye-tracking* (Unpublished doctoral dissertation). Stanford University.

Kruger, J., Schneider, J., & Westermann, R. (2006). ClearView: An Interactive Context Preserving Hotspot Visualization Technique. *Visualization and Computer Graphics*, *12*(5), 941–948. doi: 10.1109/tvcg.2006.124

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*. doi: 10.3109/02699209308985546

Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience*, *5*. doi: 10.3389/fnhum.2011.00061

Kuipers, J.-R., & Thierry, G. (2013). ERP-pupil size correlations reveal how bilingualism enhances cognitive flexibility. *Cortex*, *49*(10), 2853–2860. doi: 10.1016/j.cortex.2013.01.012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest' [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lmerTest` (R package version 2.0-29)

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. doi: 10.1177/1745691611427305

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory phonology*, *7*, 637–675. doi: 10.1515/9783110197105.637

Levelt, C. C., Schiller, N. O., & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition*, *8*(3), 237–264. doi: 10.1207/s15327817la0803_2

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1–22. doi: 10.1016/0010-0277(92)90038-j

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. doi: 10.1037/h0020279

Lleó, C., & Prinz, M. (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment. *Journal of Child Language*, *23*(01), 31–56. doi: 10.1017/s0305000900010084

Loewenfeld, I. E. (1993). *The pupil: Anatomy, physiology, and clinical applications* (Vol. 2). Ames, Detroit: Wayne State University Press. doi: 10.1136/bjo.85.1.121e

Lowry, R. (2004). *VassarStats: Website for statistical computation.* Retrieved from vassarstats.net. Vassar College.

Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon: Research on Speech Perception* (Unpublished doctoral dissertation). Indiana University.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5-6), 565–581. doi: 10.1080/01690960143000137

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*(1), 1–36. doi:

10.1097/00003446-199802000-00001

Luche, C. D., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the intermodal preferential looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*, *40*, 151–172. doi: 10.1016/j.infbeh.2015.05.005

Maddieson, I. (2013). *Syllable structure.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/chapter/12`

Mani, N., Coleman, J., & Plunkett, K. (2008). Phonological specificity of vowel contrasts at 18-months. *Language and Speech*, *51*(1-2), 3–21. doi: 10.1177/00238309080510010201

Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: an event-related potential study. *Developmental Science*, *15*(1), 2–11. doi: 10.1111/j.1467-7687.2011.01092.x

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. doi: 10.1016/j.jml.2007.03.005

Mani, N., & Plunkett, K. (2010a). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*(7), 908–913. doi: 10.1177/0956797610373371

Mani, N., & Plunkett, K. (2010b). Twelve-month-olds know their 'cups' from their 'keps' and 'tups'. *Infancy*, *15*(5), 445–470. doi: 10.1111/j.1532-7078.2009.00027.x

Mani, N., & Plunkett, K. (2011a). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, *38*(3), 606–627. doi: 10.1017/s0305000910000243

Mani, N., & Plunkett, K. (2011b). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206. doi: 10.1016/j.cognition.2011.06.013

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloSOne*, *7*(8), e43230. doi: 10.1037/e505772014-047

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1),

177–190. doi: 10.1016/j.jneumeth.2007.03.024

Marshall, C. R., & van der Lely, H. K. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia. *Language*, *85*(1), 39–57. doi: 10.1353/lan.0.0081

Marshall, S. (2002). The index of cognitive activity: measuring cognitive workload. In *Proceedings of the IEEE 7th conference on human factors and power plants.* Institute of Electrical & Electronics Engineers (IEEE). doi: 10.1109/hfpp.2002.1042860

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. doi: 10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1376–1392. doi: 10.1037/0096-1523.22.6.1376

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. doi: 10.1016/0010-0285(78)90018-x

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494. doi: 10.1006/cogp.1999.0721

Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from trace simulations. *Journal of Memory and Language*, *71*(1), 89–123. doi: 10.1016/j.jml.2013.09.009

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 207–263. doi: 10.1086/ci.1986.12.issue-4

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi: 10.1016/0010-0285(86)90015-0

McKean, C., Letts, C., & Howard, D. (2013). Functional reorganization in the developing lexicon: separable and changing influences of lexical and phonological variables on children's fast-mapping. *Journal of Child Language*, *40*(02), 307–335. doi: 10.1017/s0305000911000444

McLeod, S., Doorn, J. v., & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech-Language Pathol-*

ogy, *10*(2), 99. doi: 10.1044/1058-0360(2001/011)

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42. doi: 10.1016/s0010-0277(02)00157-9

Mehler, J., Dupoux, E., & Segui, J. (1990). *Constraining models of lexical access: The onset of word recognition.* Cambridge, MA, USA: The MIT Press.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. *Language Production*, *2*, 3–50. doi: 10.1017/cbo9780511980503.009

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & Cognition*, *25*(1), 47–56. doi: 10.3758/bf03197284

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3. doi: 10.1037/0022-0663.91.1.3

Metsala, J. L., Stavrinos, D., & Walley, A. C. (2009). Children's spoken word recognition and contributions to phonological awareness and nonword repetition: A 1-year follow-up. *Applied Psycholinguistics*, *30*(01), 101–121. doi: 10.1017/s014271640809005x

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. *Word Recognition in Beginning Literacy*, 89–120.

Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, *26*(4), 305–308. doi: 10.3758/bf03337665

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352. doi: 10.1121/1.1907526

Miner, K. L. (1989). Winnebago accent: The rest of the data. *Anthropological Linguistics*, 148–172. doi: 10.1353/anl.2010.0007

Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 496. doi: 10.1037/a0020989

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. doi: 10.1037/h0027366

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language & Hearing Research*, *48*(1).

Munson, B., Edwards, J., & Beckman, M. E. (2011). Phonological representations in language acquisition: Climbing the ladder of abstraction. *Handbook of Laboratory Phonology*, 288–309. doi: 10.1093/oxfordhb/9780199575039.013.0012

Naigles, L. R., & Tovar, A. T. (2011). Portable Intermodal Preferential Looking (IPL): investigating language comprehension in typically developing toddlers and young children with autism. *Journal of Visualized Experiments*, *70*, e4331–e4331. doi: 10.3791/4331

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*(4), 522–537.

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*(2), 271–279. doi: 10.1016/j.cogdev.2006.10.007

Nieuwenhuis, S., Geus, E. J. D., & Aston-Jones, G. (2011). The anatomical and functional relationship between the p3 and autonomic components of the orienting response. *Psychophysiology*, *48*(2), 162–175. doi: 10.1111/j.1469-8986.2010.01057.x

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, *87*(6), 2705. doi: 10.1121/1.399061

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, *32*(1), 120. doi: 10.1044/jshr.3201.120

Oakes, L. M. (2011). Advances in eye tracking in infancy research. *Infancy*, *17*(1), 1–8. doi: 10.1111/j.1532-7078.2011.00101.x

Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, *32*(6), 397–422. doi: 10.1016/s0021-9924(99)00018-0

Padgett, J. E. (1991). *Stricture in feature geometry* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, *80*(3), 384–402. doi: 10.1353/lan.2004.0141

Pertz, D. L., & Bever, T. G. (1975). Sensitivity to phonological universals in children and adolescents. *Language*, 149–162. doi: 10.2307/413156

Peterson, B. S., & Ment, L. R. (2001). The necessity and difficulty of conducting magnetic resonance imaging studies on infant brain development. *PEDIATRICS*, *107*(3), 593–594. doi: 10.1542/peds.107.3.593

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *The Journal of the Acoustical Society of America*, *121*(4), 2312–2326. doi: 10.1121/1.2642397

Phatak, S. A., Lovitt, A., & Allen, J. B. (2008). Consonant confusions in white noise. *The Journal of the Acoustical Society of America*, *124*(2), 1220–1233. doi: 10.1121/1.2913251

Pierrehumbert, J. (2002). Probabilistic phonology: Discrimination & robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228).

Pinheiro, J. C., Bates, D., DebRoy, S., & Sarkar, D. (2007). Linear and nonlinear mixed effects models. *R-Package-version*, *3*, 57.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. Springer New York. doi: 10.1007/978-1-4419-0318-1

Pulvermuller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. doi: 10.1038/nrn2811

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers.

Cognitive Psychology, *59*(1), 96–121. doi: 10.1016/j.cogpsych.2009.02.002

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *The Quarterly Journal of Experimental Psychology*, *63*(4), 772–783. doi: 10.1080/17470210903104412

Ren, J., & Morgan, J. L. (2011). Sub-segmental details in early lexical representation of consonants. In *Proceedings of the 17th International Congress of Phonetic Sciences.*

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*(3), 484–494. doi: 10.1037/0012-1649.39.3.484

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. *Handbook of child psychology*. doi: 10.1002/9780470147658.chpsy0202

Sagey, E. (1988). On the Ill-Formedness of Crossing Association Lines. *Linguistic Inquiry*, 109–118.

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology* (Unpublished doctoral dissertation). Massachussetts Institute of Technology, Cambridge.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. doi: 10.1016/s0010-0277(03)00139-2

Schröder, A., Gemballa, T., Ruppin, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods*, *44*(2), 380–394. doi: 10.3758/s13428-011-0164-y

Selkirk, E. O. (1984). On the major class features and syllable theory. In *Language sound structure* (pp. 107–136). Cambridge, MA, USA: The MIT Press.

Sharpe, D. (2015). Your Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research & Evaluation*, *20*(8), 2. doi: 10.4135/9781412950596.n433

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. doi: 10.1002/wcs.1323

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech and Hearing Research*, *36*(5), 931. doi: 10.1044/jshr.3605.931

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382. doi: 10.1038/41102

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, *28*(2), 164–188. doi: 10.1016/0749-596x(89)90042-9

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language*, *25*(2), 163–180. doi: 10.1016/0749-596x(86)90027-6

Steriade, D. (1982). *Greek prosodies and the nature of syllabification* (Unpublished doctoral dissertation). Universite Laval.

Steriade, D. (2009). The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. In P. Kiparsky, K. Hanson, & S. Inkelas (Eds.), *The nature of the word: studies in honor of Paul Kiparsky* (pp. 178–202). Cambridge, MA, USA: The MIT Press. doi: 10.7551/mitpress/9780262083799.001.0001

Stoel-Gammon, C. (1987). Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools*, *18*(4), 323–329. doi: 10.1044/0161-1461.1804.323

Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research*, *53*(3), 670–683. doi: 10.1044/1092-4388(2009/08-0254)

Stokes, S. F. (2013). The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language*, 1–24. doi: 10.1017/s030500091300010x

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, *29*(2), 251–274. doi: 10.1017/s0305000902005032

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, *36*(291-321). doi: 10.1017/s030500090800891x

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. *Invariance and Variability in Speech Processes*, 58–84.

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, *66*(1), 18–41. doi: 10.1006/jecp.1997.2375

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*, *46*(2-3), 265–294. doi: 10.1177/00238309030460021001

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*(5), 432–443. doi: 10.1111/j.1467-7687.2005.00432.x

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*(2), 252–269. Retrieved from `http://dx.doi.org/10.1016/j.jml.2008.11.003` doi: 10.1016/j.jml.2008.11.003

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology*, *52*(7), 1011–1023. doi: 10.1037/dev0000114

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. doi: 10.1016/s0010-0277(00)00081-0

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. doi: 10.1111/1467-9280.00485

Szagun, G., Schramm, S. A., & Stumper, B. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt: Pearson Assessment. doi: 10.3726/978-3-653-03521-6/20

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (in press). Pupillometry registers toddler's sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology*. doi: 10.1016/j.jecp.2016.07.014

Tamási, K., McKean, C., Gafos, A., & Höhle, B. (2016a). Children's sensitivity to degrees of mispronunciation: Enriching the preferential looking paradigm with pupillometry. *Manuscript in preparation*.

Tamási, K., McKean, C., Gafos, A., & Höhle, B. (2016b). Consonant

clusters in adults' and children's word recognition and production. *Manuscript in preparation*.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. Retrieved from `http://dx.doi.org/10.1126/science.7777863` doi: 10.1126/science.7777863

Thelen, E. (1996). *A dynamic systems approach to the development of cognition and action*. Cambrige, MA, USA: MIT press. doi: 10.1162/jocn.1995.7.4.512

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, *15*(1-3), 49–74. doi: 10.1016/0010-0277(83)90033-1

Treiman, R. (1991). Children's spelling errors on syllable-initial consonant clusters. *Journal of Educational Psychology*, *83*(3), 346. doi: 10.1037//0022-0663.83.3.346

Treiman, R., & Baron, J. (1983). Phonemic-analysis training helps children benefit from spelling-sound rules. *Memory & Cognition*, *11*(4), 382–389. doi: 10.3758/bf03202453

Treiman, R., & Breaux, A. M. (1982). Common phoneme and overall similarity relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research*, *11*(6), 569–598. doi: 10.1007/bf01067613

Treiman, R., & Cassar, M. (1996). Effects of morphology on children's spelling of final consonant clusters. *Journal of Experimental Child Psychology*, *63*(1), 141–170. doi: 10.1006/jecp.1996.0045

Vihman, M. (2010). Phonological templates in early words. *Laboratory Phonology 10*, *4*(4), 261. doi: 10.1515/9783110224917.3.261

Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics*, *45*(4), 683–725. doi: 10.1515/ling.2007.021

Vihman, M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, *50*(3), 336–353. doi: 10.1016/j.jml.2003.11.004

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of*

Memory and Language, *40*(3), 374–408. doi: 10.1121/1.423698

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development*, *3*(2), 137–165. doi: 10.1016/0885-2014(88)90016-0

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*(3), 286–350. doi: 10.1006/drev.1993.1015

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, *16*(1-2), 5–20. doi: 10.1023/a:1021789804977

Walley, A. C., Smith, L. B., & Jusczyk, P. W. (1986). The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Memory & Cognition*, *14*(3), 220–229. doi: 10.3758/bf03197696

Waterson, N. (1971). Some views on speech perception. *Journal of the International Phonetic Association*, *1*(02), 81–96. doi: 10.1017/s0025100300000293

Watson, M. M., & Scukanec, G. P. (1997). Phonological changes in the speech of two year olds: A longitudinal investigation. *Infant-Toddler Intervention: The Transdisciplinary Journal*, *7*(1), 67–77. doi: 10.3109/17549507.2012.663936

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. doi: 10.1207/s15473341lld0102_4

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. doi: 10.1207/s15327078in0301_1

Werker, J. F., & Gervain, J. (2013). *Speech perception in infancy* (P. D. Zelazo, Ed.). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199958450.013.0031

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672. doi: 10.1037/0012-1649.24.5.672

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception:

Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. doi: 10.1016/s0163-6383(02)00093-0

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*(4), 221–226. doi: 10.1177/0963721412449459

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. doi: 10.1016/j.jml.2008.03.001

White, K. S., Morgan, J. L., & Wier, L. (2005). When is a 'dar' a 'car'? Effects of mispronunciation and referential context on sound-meaning mappings. In *Proceedings of the 29th Annual Boston University Conference on Language Development* (pp. 651–662).

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4), 362–378. doi: 10.1016/j.jml.2013.01.003

Wilson, C., & Davidson, L. (2015). Acoustic characteristics of open transition in nonnative consonant cluster production. In *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5).

Yavas, M. S., & Gogate, L. J. (1999). Phoneme awareness in children: A function of sonority. *Journal of Psycholinguistic Research*, *28*(3), 245–260. doi: 10.1007/10936.1573-6555

Yeung, H. H., Chen, L. M., & Werker, J. F. (2013). Referential labeling can facilitate phonetic learning in infancy. *Cognitive Development*, *85*(3), 1036–1049. doi: 10.1111/cdev.12185

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–418. doi: 10.1111/j.1467-7687.2008.00789.x

Yun, S. (2012). Perceptual similarity and epenthesis positioning in loan adaptation. In *Proceedings of the Chicago Linguistic Society* (Vol. 48).

Yun, S. (2014a). English -uh-insertion and consonant cluster splittability. In *Proceedings of the Annual Meetings of Phonology* (Vol. 48).

Yun, S. (2014b). The role of acoustic cues in nonnative cluster repairs. In *Proceedings of the 31st West Coast Conference on Formal Linguistics* (Vol. 31).

Zesiger, P., Lozeron, E. D., Lévy, A., & Frauenfelder, U. H. (2011). Phonological specificity in 12- and 17-month-old French-speaking infants. *Infancy*, *17*(6), 591–609. doi: 10.1111/j.1532-7078.2011.00111.x

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3. doi: 10.1037/0033-2909.131.1.3