Pupillometry registers toddlers' sensitivity to degrees of mispronunciation.

Tamasi, Katalin; McKean, C; Gafos, A; Fritzsche, Tom; Höhle, Barbara

Published in:
Journal of Experimental Child Psychology

DOI:
10.1016/j.jecp.2016.07.014

Journal of Experimental Child Psychology 153 (2017) 140–148

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

Pupillometry registers toddlers' sensitivity to degrees of mispronunciation

Katalin Tamási [a,*], Cristina McKean [b], Adamantios Gafos [a], Tom Fritzsche [c], Barbara Höhle [a]

[a] International Doctorate for Experimental Approaches to Language and the Brain (IDEALAB), University of Potsdam (DE), Newcastle University (GB), University of Groningen (NL), University of Trento (IT), and Macquarie University (Sydney, AU)
[b] School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne NE1, UK
[c] Department of Linguistics, University of Potsdam, 14476 Potsdam, Germany

ARTICLE INFO

Article history:
Available online 29 September 2016

Keywords:
Phonological development
Featural distance
Lexical representations
Mispronunciation detection
Pupillometry
Eyetracking

ABSTRACT

This study introduces a method ideally 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* ∼ *daby* ∼ *faby* ∼ *shaby*), we tested 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-segmental detail and by the degree of mismatch. Thus, we

* Corresponding author.
E-mail address: tamasi@uni-potsdam.de (K. Tamási).

http://dx.doi.org/10.1016/j.jecp.2016.07.014

establish that pupillometry provides a viable alternative to paradigms that require overt behavioral response in increasing our understanding of the development of lexical representations.

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, Fennell, Swingley, & Werker, 2009). However, the precise degree of this specificity in early words requires further investigation. Whereas some studies have found evidence for children's ability to detect differing degrees of mismatch (Mani, Mills, & Plunkett, 2012; Ren & Morgan, 2011; White & Morgan, 2008), others have not (Bailey & Plunkett, 2002; Swingley & Aslin, 2002). This study sought to determine children's sensitivity to degree of mispronunciation with a tool that 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—during 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, Siqueland, Jusczyk, & Vigorito, 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 because children might 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- to 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), and height and backness in vowels (e.g., *bed–bid, brush–brash*) (Mani et al., 2012). Moreover, children as young as 14 months 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 possible only if early words contain sub-segmental detail.

So far, only a handful of studies, all using the preferential looking paradigm, have considered the question of whether children younger than 3 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, 2011; Ren & Morgan, 2011; White & Morgan, 2008), whereas 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. Their 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 toward the picture of "keys" was greater than that toward 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 toward the target, but to a lesser extent than in the correct condition. When two-feature changes were introduced ("deys"), infants exhibited a nonsignificant target preference, and when three-feature changes were introduced ("zeys"), they exhibited a nonsignificant distractor preference (the two- and 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, children were able to do so when the onset differed by one feature, but less successfully than with the correct label (as evidenced by the drop in the proportion of looking times toward the target). Furthermore, 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. In addition, a similar graded sensitivity in looking time has been observed when manipulating vocalic rather than consonantal featural distance (Mani & Plunkett, 2011). In that study, 24-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 a target preference, whereas larger two- and three-feature deviations yielded a weak distractor preference.

We highlight here an important shared methodological characteristic of the studies described above that 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 affected not only 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, 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 psychosensory reflex, that is, pupil dilation (Loewenfeld, 1993). In pupillometry, instead of recording the pattern of gaze fixations, the eyetracking equipment is used to measure the 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, pupillometry 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 that may provide a more appropriate way in which 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 distractor image) (Klingner, 2010). Third, pupillometry is inexpensive and easy to learn. Although electrophysiological and brain imaging techniques generally avoid the shortcomings that looking time paradigms introduce, such techniques require specialized equipment and expertise. The eyetracking 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 mispronounced forms.

Furthermore, in contrast to previous studies, our study employed a featurally balanced consonantal set, also cross-balanced for feature type and change. Words with a diverse set of initial consonants were selected and then systematically manipulated not just by the number but also by the type and direction of feature changes (see Table 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; cf. Vihman & Croft, 2007).

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

Effect of mispronunciation: 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.

Method

Participants

A total of 48 30-month-old monolingual German children (26 girls) were recruited (M_{age} = 30 months, SD = 0.56) from the BabyLAB Participant Pool at the University of Potsdam. Caregivers reported no developmental or sensory disabilities. We assessed 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 Communicative Development Inventories [CDI]; cf. Szagun, Schramm, & Stumper, 2009). Participants were reported to be familiar with the majority of (correct) experimental items (M = 79.9%, SD = 16.9). 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).

Stimuli

To identify words likely to be known by toddlers, 20 easily depictable words with CVC (consonant–vowel–consonant) 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', [ʃa:f], 'sheep'), 20 items with one feature change (e.g., [ta:f], manner of articulation change), 20 items with two feature changes (e.g., [da:f], manner of articulation and voicing change), and 20 items with three feature changes (e.g., [ga:f], manner of articulation, voicing, and place of articulation change). Mispronunciations resulted in nonwords for children.[1] Type (i.e., voice, manner, place) and direction of feature change were counterbalanced. From each word, three mispronunciations were created by changing one, two, or three features. These phonologically related items (e.g., [ʃa:f], "sheep", [ta:f], [da:f], [ga:f]) formed an item family. An additional 40 easily depictable words from FRAKIS were included: 20 filler items that were always produced correctly (listed in online supplementary material)

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

Table 1
Stimulus list, organized by condition (Correct = correctly pronounced onset, Δ1F = one-feature change, Δ2F = two-feature change, Δ3F = three-feature change), noted with IPA (International Phonetic Alphabet, Laver, 1994).

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

and 20 items related to another study (20 items with onset clusters—also listed in supplementary material—to be reported in Tamási, McKean, Gafos, & Höhle, 2016). 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 1. Easily recognizable color drawings depicting a referent of the original word were converted to a similar size (~200 × 200 pixels displayed in a 300 × 300 pixel area). For more information on luminance, refer to the supplementary material. Four versions of the task were created, with each item family occurring once in each version and with the four conditions counterbalanced across the four versions; children never saw the same picture or heard the same label more than once.

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 afterward. After obtaining assent from the children and written informed consent from the caregivers, children were seated in the 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 eyetracker. All visual stimuli were shown centrally on a 17-inch (1280 × 1024) TFT screen with a size of 300 × 300 pixels forming a horizontal and vertical viewing angle of 7.4°. The experiment started following the calibration period (five screen positions, ~ 30 s).

In each trial, a picture was presented and remained on the screen for the duration of the trial (4 s). The picture appeared, and 1 s later the corresponding (correctly or incorrectly produced) auditory label was played. The critical window of analysis was the 3-s interval following the onset of the auditory stimulus. The experiment encompassed 12 blocks, with each block containing 5 trials (altogether 60 trials, including 20 fillers and 20 unrelated). Before each block, an "attention-getter" was presented (i.e., 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 min.

After the experiment, caregivers were asked to complete a questionnaire in order to estimate children's vocabulary size and children's 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 min to complete.

Results

To ensure that the words used in the experiment were part of children'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 ($n = 5$ participants). (Additional information on the pre-processing steps is given in the supplementary material.) The mean number of successful trials was 17.19 of 20 ($SD = 1.89$) in the experimental trials and 17.38 of 20 ($SD = 1.94$) in the filler trials. The mean number of successful trials per experimental condition was 4.30 of 5 ($SD = 0.10$).

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, Maechler, Bolker, & Walker, 2014). Featural distance, a within-participant 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 changes. 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 3-s window of analysis) was used as the outcome measure. The most parsimonious model contained featural distance, vocabulary size, and neighborhood density as fixed effects (for model description, see supplementary material).

Mean pupil dilation in each condition is presented in Fig. 1 (left: bar plot; right: 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

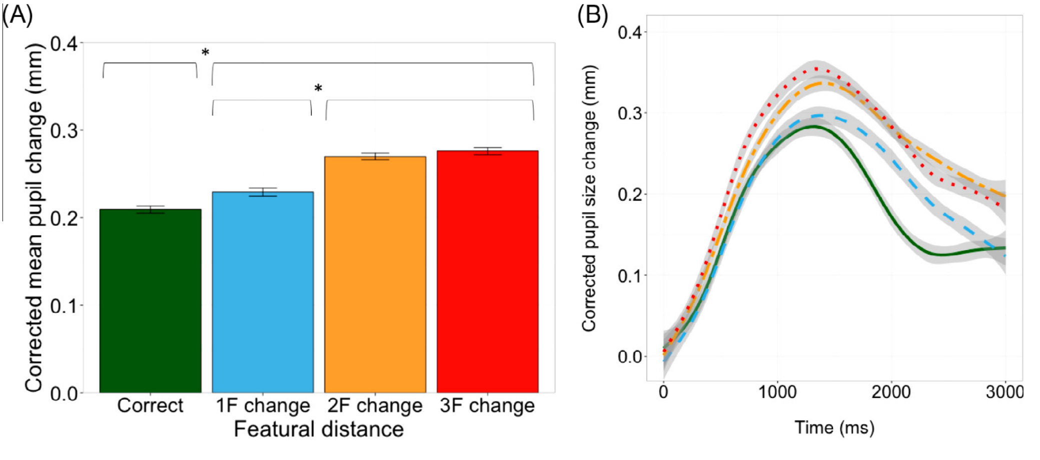

Fig. 1. (A) Mean pupil size change in response to differing degrees of mispronunciation (only familiar words included). Significant contrasts ($t > 1.96$) between the correct and mispronounced items and between the one-feature (1F) and two- and three-feature (2F and 3F, respectively) changes are marked with asterisks. Error bars represent the standard errors built around the mean. (B) Mean pupil size change over time in response to differing degrees of mispronunciation (correct: solid green; 1F change: dashed blue; 2F change: dot-dash orange; 3F change: dotted red). A 95% confidence interval was built around the fitted values, shown with gray shading.

observations. In the model, featural distance significantly contributed to model fit, $\chi^2(3) = 10.19$, $p < .017$. We subsequently found the correct versus mispronounced contrast and the one-feature change versus 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 ($\beta_3 = 0.005$, $SE = 0.023$, $t = 0.21$). Further information (including time-course analyses) is provided in the supplementary material.

Discussion

Our findings indicate that pupillometry is a viable method in lexical representation research because 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 with 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 shared between the correct and mispronounced forms, 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 that was not expected as a target label. However, we would argue that a form-related explanation is more plausible; that is, 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.

However, the results do not support complete graded sensitivity given that 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, 2011; 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 might not be linearly additive (i.e., the difference between the two- and three-feature changes might 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 forms 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 incorrect forms correlates with the degree of featural distance. It is possible that (as argued by Mani & Plunkett, 2011, 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, 2011).

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, 2011; Ren & Morgan, 2011; White & Morgan, 2008), suggesting that early lexical representations encode sub-segmental 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. Therefore, it 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.

Acknowledgments

This research was supported by Grant EMJD 520101-1-2011-1-DE-ERA to B. Höhle and K. Tamási and Grant ERC AdG 249440 to A. Gafos. The authors thank the participants and their families; the BabyLAB personnel (Katja Schneller, Carolin Jäkel, Steffi Meister, Vanessa Löffler, Elisabeth Metz, and Lennard Gottmann) for participant recruitment and running; participants of MFM22, EWCN33, and the IDEALAB Winter School 2014; and members of the Phonetics Lab (University of Potsdam), the Language Acquisition Colloquium (University of Potsdam), the Phonology and Phonetics Group (Newcastle University), and the Child Language Seminar (Newcastle University) for discussions and feedback.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2016.07.014.

References

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265–1282.
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.
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.
Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.
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: De Gruyter.
Fritzsche, T., & Höhle, B. (2015). Phonological and lexical mismatch detection in 30-month-olds and adults measured by pupillometry. In Proceedings of the 18th International Congress of Phonetic Sciences (Paper No. 0339), Glasgow, UK.
Hochmann, J.-R., & Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science, 25*, 2038–2046.
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*, 50–54.
Karatekin, C. (2007). Eye-tracking studies of normative and atypical development. *Developmental Review, 27*, 283–348.
Klingner, J. M. (2010). Measuring cognitive load during visual tasks by combining pupillometry and eye-tracking. Unpublished doctoral dissertation, Stanford University.
Kuipers, J. R., & Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5*. http://dx.doi.org/10.3389/fnhum.2011.00061.
Laver, J. (1994). The phonetic alphabet of the International Phonetic Association. *Principles of Phonetics* (pp. 593–595). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/cbo9781139166621.023.
Loewenfeld, I. E. (1993). *The pupil: Anatomy, physiology, and clinical applications* (Vol. 2) Ames: Iowa State University Press.
Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: An event-related potential study. *Developmental Science, 15*, 2–11.
Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language, 38*, 606–627.
Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Crosslinguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One, 7*(8), e43230.

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 (pp. 1586–1589), Hong Kong, China.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484.
Szagun, G., Schramm, S. A., & Stumper, B. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt, Germany: Pearson Assessment.
Tamási, K., McKean, C., Gafos, A., & Höhle, B. (2016). Consonant clusters in adults' and children's word recognition and production. Manuscript in preparation.
Vihman, M., & Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology. *Linguistics, 45*, 683–725.
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114–132.
Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*, 412–418.