Measuring sensitivity to phonological detail in monolingual and bilingual infants using pupillometry

Katalin Tamási[a], Thilanga D. Wewalaarachchi[b], Barbara Höhle[a] and Leher Singh[b]

[a]International Doctorate for Experimental Approaches to Language and Brain
[b]National University of Singapore, Department of Psychology
{tamasi,hoehle}@uni-potsdam.de, thilanga.d.w@gmail.com, leher.singh.nus@gmail.com

Abstract

This study shows for the first time that mispronunciation detection in 24-month-old mono- and bilingual children can be assessed using pupil data from a preferential looking study. Mispronounced words (specifically, consonant, vowel and tone changes) resulted in larger degrees of pupil dilation for bilingual children than correctly produced words, whereas monolingual children's pupillary responses provided no evidence of sensitivity to mispronunciations. Between-group comparisons revealed that pupil dilation of bilingual children in response to correct labels was lower than monolingual children. Overall performance reversed with consonant-changed labels, but was comparable with vowel- and tone-changed labels. Taking degree of pupil dilation as a proxy of cognitive effort, we argue that in comparison to monolinguals, bilingual children seem to require fewer resources for processing correct labels, while if anything, more resources for processing mispronounced labels in order to activate the corresponding lexical entry. This finding – converging on past research showing enhanced bilingual sensitivity to detect mismatch and mispronunciations [1] – further supports the notion that certain aspects of bilingual early words are represented in greater detail than those of monolingual words.

Index Terms: bilingualism; mispronunciation detection; lexical development; eye tracking; pupillometry

1. Introduction

One of the first steps in language acquisition is to identify the building blocks of words in the target 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 word learning and word recognition. The ability to detect a small yet contrastive change is crucial to building up a rich lexicon that contains minimally different words (e.g., *tap* and *cap*). Previous research tested children's lexical knowledge by presenting them with correct and incorrect forms of words (e.g., *ball* and *dall*), showing that between 11 and 24 months, children learn to differentiate between the two (preference for correct labels in head-turn preference paradigms: [2]; noticing phonemic change in habituation paradigms: [3], longer target looks with correct label in preferential looking paradigms: [4]; difference in event-related brain potential signature: [5]). These findings suggest that early lexical representations are specific, i.e., containing sufficient information to allow the learner to match with the input.

The majority of mispronunciation detection studies to date have relied on looking behavior. Eye tracking studies yield a valuable body of data, only a fraction of which are routinely considered. This study extends the methodology conventionally used in eye tracking studies by complementing the preferential looking paradigm with a measure automatically collected using such a paradigm: pupil dilation. Pupil dilation in children has been linked to cognitive effort, surprise, and novelty [6], making it an appealing tool for infant research. Recently, pupillometry was demonstrated to be sensitive to acoustic (dis-)similarity: [7], semantic violation: [8], and mispronunciations: [9, 10, 11].

Using single-picture pupillometry paradigms – presenting a single visual stimulus per trial –, 30-month-old children have been shown to give a differential pupillary response to correctly pronounced labels and their mispronunciations: the general finding being that mispronounced labels were associated with larger degrees of pupil dilation than correct labels [9, 10]. 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 reconstruct the correct form and map onto the corresponding lexical representation) than doing so with the correct label. Such finding is consistent with earlier studies that demonstrated the specificity of lexical representations [2, 3, 4, 5].

Most recently, 30 month-old children's pupil dilation data collected from a preferential looking paradigm have been found to exhibit a linear trend in response to the degree of mispronunciation [11]. The methodological import of the study is showing that pupillary response data can act as an additional measure to looking behavior in preferential looking paradigms as even in the presence of the distractor picture, the degree of pupil dilation remains associated with the cognitive effort of reconstructing the correct form of the target label.

Notwithstanding that growing up in a multilingual household is the global norm, differences between mono- and multilingual language development are not well understood: Only a handful of studies to date have considered multilingual (in particular, bilingual) lexical knowledge. Whether early bilingual lexical representations are more or less detailed / stable than monolingual ones given the inherent complexity of the multilingual language environment is an open empirical question. One can approach this question by determining how the mispronunciation detection skills of mono- and bilingual children differ, if at all. If bilingual children were more sensitive to phonological contrasts than monolinguals, it would enable building finer-grained lexical representations (and/or have better access thereto). In contrast, attenuated bilingual ability in mispronunciation detection could indicate weaker, less stable lexical representations, possibly due to relatively less experience with and exposure to each language in comparison to monolinguals.

Using the switch task, studies have found that both bilingual and monolingual 17- to 20-month olds dishabituated after the presentation of the phonetically different stimulus provided the speakers' language background matched theirs (i.e., bilingual infants learned more successfully from bilingual speakers and monolinguals from monolingual speakers) [12, 13, 14]. These findings show that both mono- and bilinguals are similarly attuned to phonetic detail and use such information to guide their learning of minimally different words. On the flip side, children from either a mono- or bilingual background may be disadvantaged when tested on non-matching stimuli, suggesting that bi- and monolingual infants are similarly constrained by the set of phonetic characteristics present in their environment [12, 13, 14].

In support of bilingual advantage, Mandarin-English bilinguals were found to be able to detect tone contrasts (e.g., a contrast between Tones 1 vs. 4) in newly learned words earlier than Mandarin monolinguals (12-13 vs. 17-18 months) [1]. Even with a non-tone language background, tonal sensitivity can be gained by bilinguals earlier than by monolinguals (11-12 vs. 17-18 months) [15]. To account for the bilingual success in these tasks, the authors cited cognitive adaptations that may accompany phonological conflict introduced by exposure to a tone- and non-tone language. This invites the possibility that, instead of being a source of confusion, under certain conditions the multilingual environment may in fact boost the identification of contrastive properties and thus the formation of word-object associations.

Similarly to [11], this study analyzed pupil dilation captured in a preferential looking paradigm (the eye tracking data is reported in [16]). In accord with the findings of past studies using pupillometry, 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 [9, 10, 11]. Critically for this study, given the existing literature on mono- vs. bilinguals' different abilities, the mispronunciation effect was predicted to be modulated by language background.

2. Method

Sixty two-year-old infants (1;11;3 – 2;2;1, M = 2;0;6, 27 boys) participated, 13 of which did not provide data analyzable by the eye tracker due to track loss, calibration error, and noncompliance. Infants were categorized as either Mandarin-English bilinguals (Mandarin exposure: 50-75%) or as Mandarin monolinguals (Mandarin exposure: $>$90%). Participants who did not fit those categories (N = 6) were excluded. Finally, those participants who did not reach a threshold of 50% of successful trials (those trials that contain pupil information from at least half the length of the trial) were excluded from further analyses, leaving 14 bilingual and 18 monolingual participants. On average, 81% of trials per participant were retained.

A total of 36 different target words (and 2 correctly produced practice words) were recorded by a Mandarin native speaker. All labels were monosyllabic nouns preceded by the exclamation 'Look!' in Mandarin Chinese. Auditory stimuli were normalized for amplitude and presented at 70 dB using Praat (version 5.3.63, [16]). Eighteen of these target words were correctly produced, 6 contained a vowel, 6 a consonant change (made to the word onset) and 6 a tone change. Each phonemic change constituted a single-feature deviation from the correct form. Out of the four lexical tones used by Mandarin Chinese (high level [Tone 1],rising [Tone 2], dipping [Tone 3], and falling [Tone 4]), three of which were used in this study. Tone mispronunciation included change between Tones 1 and 2, 2 and 4, and 1 and 4 (with the direction of change counterbalanced). Easily recognizable color pictures depicting a referent of the original word were converted such that all pictures were of a similar size (approximately 200 x 200 pixels displayed in a 320 x 320 pixel area). The areas of interest included the 450 x 450 pixel area around each picture.

In each trial, a pair of target and distractor images were simultaneously presented on each side of the screen on a white background for 2500 ms, the target depicting a familiar item and the distractor an unfamiliar item. After the center fixation phase, the same pair of pictures was shown for 2500 ms, accompanied by a Mandarin auditory label for the target image that was either correctly pronounced or mispronounced (consonant / tone / vowel change). 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. Prior to the experiment, 2 practice items were presented in order to familiarize the child with the task. The practice items were not included in the analysis. Altogether, participants were presented with 20 correctly and 18 incorrectly pronounced items in each version of the experiment.

The study was conducted on a Tobii 60 XL eye tracker (version 3.2.1) coupled with a 24" LCD monitor used for the presentation of stimuli and the recording of infant eye gaze and an experimenter computer (Dell Optiplex 755). Children were seated 70–80 cm away from the screen. The auditory stimuli were presented via the in-built speakers on both sides of the LCD monitor.

3. Results

The prediction that language background modulated the mispronunciation effect was supported by the analyses. Linear mixed effects models were employed with random intercepts and slopes (estimates were chosen to optimize the log-likelihood criterion) [18]. `Mispronunciation type` (correct / consonant change / tone change / vowel change) and `language background` (bilingual / monolingual) were introduced as fixed effects, subjects and items as random effects, and corrected pupil size change (the mean of the 100 ms interval prior to the auditory label onset subtracted from the raw pupil value) was used as the outcome measure. `Mispronunciation type` was treatment-coded such that the correct condition was compared to all other conditions. Each intercept and slope fitted by the model was adjusted by the effect of `mispronunciation type` nested in participants [19]. The correlation term for `mispronunciation type` in the random effect structure was removed [20]. The most parsimonious model was chosen through comparisons using Likelihood Ratio Tests [21] using the `anova` function from the `stats` package [22]. Thus the interaction model was chosen over the main effects model (χ^2(3) = 8.54, $p < .036$). This model contained a significant `mispronunciation type` x `language background` interaction, driven by the correct vs. consonant and correct vs. vowel change contrasts in bilinguals (β_1 = 0.02, SE = 0.01, t = 1.98, β_2 = 0.03, SE = 0.01, t = 2.21, respectively, the correct vs. tone change contrast was not significant: t $<$ 1.23) and no significant contrasts in monolinguals (all ts $<$ 1.10).

Time-course analyses (post-hoc cluster-based permutation tests [23, 24]) were used to explore when significant differences emerged across the factor levels (4 levels of `mispronunciation type` and 2 levels of `language background`). 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

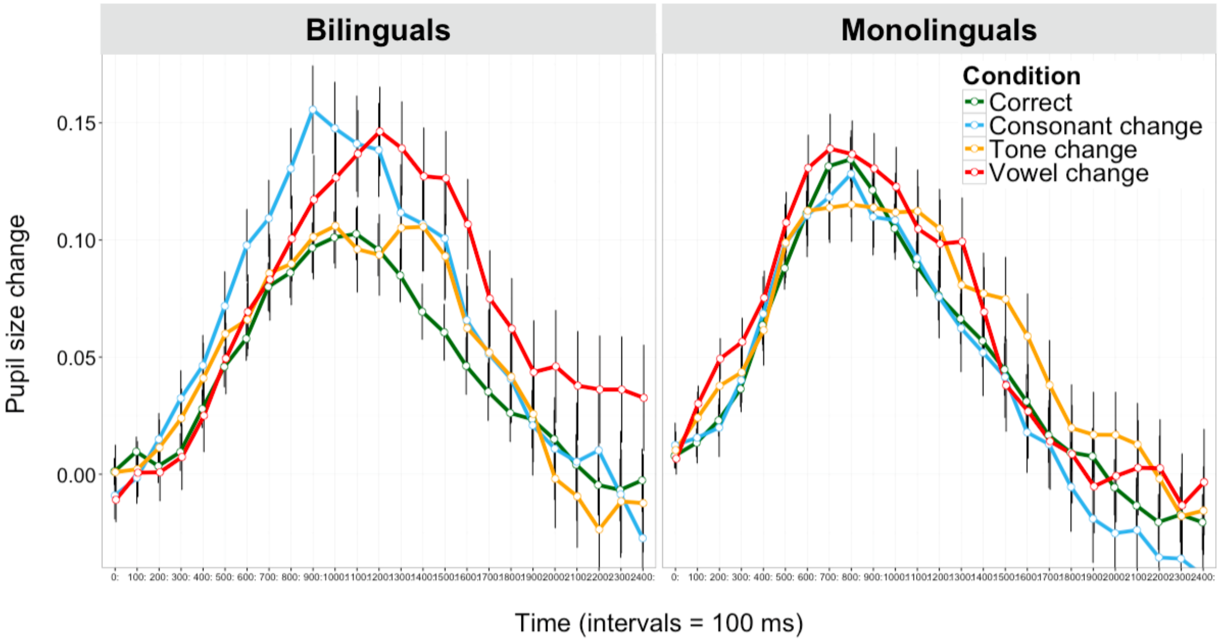

Figure 1: *Corrected pupil size change by mispronunciation type and language group (auditory label onset = 0 ms, error = 95%CI).*

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. With this method, significant clusters in the bilingual group were identified. Positive cluster-level t-values signal that all mispronounced conditions were associated with larger pupil dilation than the correct condition among bilinguals (lines 1–3 in Figure 1). No significant clusters in the monolingual group were found (c.f., lines 4–6 in Figure 1). Furthermore, comparisons across the bi- and monolinguals yielded significant clusters in all conditions (c.f., lines 7–9 in Figure 1). Positive cluster-level t-values indicate that monolinguals' pupils dilated significantly more than bilinguals', while negative values signal significant differences in the opposite direction. Thus in the consonant change condition, bilinguals exhibited a higher degree of pupil dilation than monolinguals and in the vowel and tone change conditions, both directions can be observed.

4. Discussion

This study supports previous research [1, 15] that has found a bilingual advantage in word recognition: The significance of the correct vs. mispronounced contrast in the bilingual and lack thereof in the monolingual group indicate that bilinguals exhibited more sensitivity to segmental change than their monolingual peers, which corroborates previous work showing bilingual children to be more sensitive to contrasts than monolinguals [1, 15]. Time-course post-hoc analyses confirmed those findings from linear mixed effects models and additionally found the correct vs. tone contrast to be significant in the bilingual group.

Time-course analyses revealed that hearing the correct form induced greater degrees of pupil dilation for monolingual vs. bilingual speakers. Contrastively, hearing consonant-change mispronunciation produced the opposite pattern and when processing the other mispronounced labels, the overall pupil dilation provided by the two language groups were comparable.

Table 1: *Summary table of time-course analyses (Interval = time interval in the naming phase, $\sum t$ = cluster-level t, p = p-value associated with cluster-level t, Corr. = correct label, ΔC = consonant change, ΔV = vowel change, ΔT = tone change).*

Contrasts	Interval	$\sum t$	p
Bilinguals			
Corr. vs. ΔC	800–1500	16.60	*
Corr. vs. ΔV	1100–1600	38.12	**
Corr. vs. ΔT	1400–1500	3.31	*
Monolinguals			
Corr. vs. ΔC	–	–	n.s.
Corr. vs. ΔV	–	–	n.s.
Corr. vs. ΔT	–	–	n.s.
Bi- vs. Monolinguals			
Corr.	200–900	14.83	*
ΔC	900–1800	−18.91	**
ΔV	100–900	18.55	**
ΔV	1200–1900	−11.86	*
ΔT	500-700	3.41	*
ΔT	1400-1500	−3.23	†

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

Considering pupil dilation to be a direct measure of cognitive effort [9, 10, 11], those findings can be interpreted such that bilinguals use fewer cognitive resources to activate the respective lexical representation than monolinguals. In a similar vein, bilinguals require more resources than monolinguals to activate the lexical entry when the label is only a partial match, showing greater specificity (or less flexibility) than monolinguals, at least with consonants.

Finding no evidence for the mispronunciation effect from the monolingual pupillary response is intriguing because it was demonstrated with 30-month-old monolinguals in previous research (i.e., based on their pupil size changes, monolingual children reliably differentiated between correct and mispronounced labels) [9, 10, 11]. The apparent discrepancy of monolingual performance versus the present null-result may be due to inherent differences with respect to manipulation, paradigm, and participant age. It is challenging to compare the study with previous research that explored the effect of featural distance and thus manipulated the number of feature mispronunciations produced by a monolingual speaker [10, 11] (vs. type of mispronunciations produced by a bilingual speaker), using single-picture pupillometry paradigms [9, 10] (vs. the preferential looking paradigm) with 30-month-olds [9, 10, 11] (vs. 24-month-olds). Of these potential reasons for discrepancy, age seems the least likely as younger children have been shown to be sensitive to mispronunciations with other methodologies [2, 3, 4, 5].

It is worth noting that in this study, the attrition rate was high due to calibration error and track loss, barring eye and pupil tracking for 13 children. Once more data are collected, further analyses can be carried out that allow for the restriction of the analysis window to target looks (similarly to [11]). Nevertheless, results at this stage suggest that bilinguals' pupillary response was more sensitive to phonemic contrasts than monolinguals' (a finding consistent with those of [1, 15]).

5. Conclusions

The present study is the first to offer evidence that children's mispronunciation detection skills – as measured by pupillary reactions – are affected by language background. Results show that for bilinguals, mispronounced labels yielded an increase in pupil dilation in comparison to correct labels, whereas for monolinguals, no significant differences in pupil dilation were observed. Furthermore, reliable differences were recorded between mono- and bilinguals at two levels of mispronunciation type. Correct labels were associated with smaller degrees of pupil dilation – while consonant change with larger degrees of pupil dilation – in the bilingual vs. in the monolingual group.

Following past research [9, 10, 11], we interpret changes in the pupillary response as an indicator of resource consumption. As such, bilingual children seem to require less effort to link the correct label to its corresponding representation than monolingual children, suggesting more economical processing. On the other hand, establishing such a link when the label has a consonant change seems to be more demanding for bilingual than monolingual children. Methodologically, this study contributes to the growing body of literature demonstrating pupillometry to be a viable - dynamic and gradient - tool to study early word recognition [9, 10, 11], hence contributing to our understanding of mono- and bilingual developmental trajectories.

6. Acknowledgments

This research was supported by a Tier 1 grant to LS (FY2013-FRC2-009). KT and BH was supported by EMJD 520101-1-2011-1-DE-ERA. We would like to thank Wong Liang Hui and Felicia L.S. Poh for recruiting and testing all of the participants.

7. References

[1] Singh, L., Poh, F. L., and Fu, C. S. (2016). Limits on Monolingualism? A Comparison of Monolingual and Bilingual Infants Abilities to Integrate Lexical Tone in Novel Word Learning. *Frontiers in Psychology, 7:* 667.

[2] Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*(5), 432-443.

[3] Yoshida, K. A., Fennell, C. T., Swingley, D., and Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 8*, 12(3), 412-418.

[4] Swingley, D., and Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Developmental Science, 8*(2), 147-166.

[5] Mani, N., Mills, D. L., and Plunkett, K. (2012). Vowels in early words: an event-related potential study. *Developmental Science, 15*(1), 2-11.

[6] Karatekin, C. (2007). Eye-tracking studies of normative and atypical development. *Developmental Review, 27*(3):283348.

[7] Hochmann, J.-R. and Papeo, L. (2014). The invariance problem in infancy: A pupillometry study. *Psychological Science, 25*(11):2038-46.

[8] Kuipers, J. and Thierry, G. (2011). N400 amplitude reduction correlates with an increase in pupil size. *Frontiers in Human Neuroscience, 5.*

[9] Fritzsche, T., and Höhle, B. (2015). Phonological and lexical mismatch detection in 30-month-olds and adults measured by pupillometry. In The *Proceedings of the ICPhS, (18),* 03-39.

[10] 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*.

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

[12] Fennell, C., and Byers-Heinlein, K. (2014). You sound like Mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. International Journal of Behavioral Development, 38(4), 309-316.

[13] Fennell, C. T., Byers-Heinlein, K., and Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development, 78*(5), 1510-1525.

[14] Mattock, K., Polka, L., Rvachew, S., and Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science, 13*(1), 229-243.

[15] Liu, L., Kager, R. (2016). Perception of Tones by Bilingual Infants Learning Non-Tone Languages. *Bilingualism: Language and Cognition, 1-15.*

[16] Wewalaarachchi, T.D., Wong, L.H. and Singh, L. (2016). *Sensitivity to phonological detail in bilingual and monolingual infants: Evidence from familiar word recognition.* Manuscript submitted for publication.

[17] Boersma, P., and Weenink, D. (2013). *Praat: doing phonetics by computer [Computer program] (Version 5.3.51).* Retrieved from http://www.praat.org/

[18] Bates, D. (2005). Fitting linear mixed models in R. *R news, 5*(1), 27-30.

[19] Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3):255278.

[20] Jaeger, T. F., Graff, P., Croft, W., and Pontillo, D. (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology, 15*(2):281 320.

[21] Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2007). Linear and nonlinear mixed effects models. R-Package-version, 3:57.

[22] R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

[23] Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

[24] Dink, J. W., and Ferguson, B. F. (2015). *eyetrackingR: An R Library for Eye-tracking Data Analysis.* Retrieved from http://www.eyetrackingr.com.