D E S C	12244	WILEY	Dispatch: 6.10.14	CE: Wiley
Journal Code	Manuscript No.		No. of pages: 7	PE: Sharanya

Developmental Science (2014), pp 1–7 DOI: 10.1111/desc.12244

SHORT REPORT

Mechanisms underlying accent accommodation in early word learning: evidence for general expansion

Rachel Schmale,[1,†] Amanda Seidl[2,†] and Alejandrina Cristià[3,†]

1. Department of Psychology, North Park University, USA
2. Department of Speech, Language and Hearing Sciences, Purdue University, USA
3. Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS, DEC-ENS, CNRS, France

Abstract

Previous work reveals that toddlers can accommodate a novel accent after hearing it for only a brief period of time. A common assumption is that children, like adults, cope with nonstandard pronunciations by relying on words they know (e.g. 'this person pronounces sock *as* sack*, therefore by* black *she meant* block*'). In this paper, we assess whether toddlers might additionally use a general expansion strategy, whereby they simply accept non-standard pronunciations when variability is expected. We exposed a group of 24-month-old English-learning toddlers to variability in indexical cues (very diverse voices from native English talkers), and another to variability in social cues (very diverse-looking silent actors); neither group was familiarized with the target novel accent. At test, both groups succeeded in recognizing a novel word when spoken in the novel accent. Thus, even when* no *lexical cues are available, variability can prepare young children for non-standard pronunciations.*

Research highlights

- Exposure to indexical and social variability promotes subsequent cross-accent word learning.
- Greater experience with different talkers may promote better accent accommodation.
- Children can accommodate a foreign accent when exposed to variability on linguistically irrelevant, but socially relevant dimensions.

Introduction

In a multilingual society, young children hear speech from an assortment of different talkers from diverse linguistic backgrounds. For example, an infant raised in the United States will likely be exposed to several different languages, regional and social dialects, and foreign accents. Thus, to become proficient language learners, infants and children must learn to accommodate variation in the speech stream. This variation can occur at the prosodic level (e.g. differences in pitch, duration, and pause) as well as the segmental level (e.g. differences in the production of phonemes and allophones). Some research suggests that early experience with multiple languages and accents may affect children's linguistic and cognitive development. For example, Kovacs and Mehler (2009) found that 7-month-olds growing up bilingual are more flexible when learning statistical patterns than their monolingual peers. More specifically to speech processing, mere exposure to talker variability appears to facilitate children's accommodation of voice and accent in the speech stream (e.g. Houston, 2000; Rost & McMurray, 2009; Van Heugten & Johnson, 2014). Exposure to variable speech may even boost semantic processing. Indeed, Barker and Meyer Turner (in press) found that preschoolers' comprehension of a story was better when the story was read by a foreign-accented speaker, compared to a native speaker. The mechanisms underlying children's ability to cope with speech variation are still unclear. In this paper we ask: *how* do children adapt to non-standard pronunciations (i.e. those that deviate from their native norms)?

Address for correspondence: Rachel Schmale, North Park University, 3225 W Foster Avenue, Box 57, Chicago, IL 60625, USA; e-mail: rschmale@northpark.edu
†All authors contributed equally to this work.

A growing body of recent research has documented young children's adaptation to unfamiliar accents. While infants and toddlers initially show processing costs from accented speech (e.g. Schmale & Seidl, 2009), toddlers can use lexical knowledge to retrieve word forms in unfamiliar accents (Best, Tyler, Gooding, Orlando & Quann, 2009; Mulak, Best, Tyler, Kitamura & Irwin, 2013). For example, 19-month-olds being raised in Connecticut show a preference for high-frequency word forms in both native and unfamiliar (Jamaican English) accents. In addition, White and Aslin (2011) found that American English-learning 18–20-month-olds adapt to a novel, single sound category change (e.g. *block* □ *black*) by generalizing the pronunciation deviation to other highly familiar words. Furthermore, Van Heugten and Johnson (2014) found that Canadian English-learning 15-month-olds successfully recognize words in an unfamiliar accent following brief exposure to the accent in a familiar story. In all of these studies, toddlers can utilize top-down lexical knowledge (i.e. highly familiar words) to inform the adaptation process. This mechanism has been referred to as a *lexically-based expansion strategy* (Schmale, Cristià & Seidl, 2012). Specifically, to employ this strategy, toddlers need clear evidence for a given pronunciation deviation, from which they can then extrapolate to accommodate the accented speech.

Nonetheless, this lexically-based expansion strategy is unlikely to be helpful when toddlers are faced with an unfamiliar foreign accent, which may deviate from their native input on several dimensions. In this case, they would need to hear enough examples of each kind of change to extract the general patterns. Moreover, this strategy is simply unavailable in the absence of top-down lexical information; that is, when toddlers are faced with unfamiliar words in an unfamiliar accent, or when infants are so young that they do not yet have a lexicon. Thus, in these scenarios, the learner must utilize an alternative strategy, a *general expansion* strategy, in which they generally relax their categories to accept a certain degree of deviation from native pronunciation norms (Schmale *et al*., 2012). A recent study by Schmale *et al*. (2012) suggested that toddlers could be using this strategy to cope with accent variation. Specifically, they found that English-learning 24-month-olds could successfully recognize newly learned words in an unfamiliar Spanish accent when the word-learning experiment was preceded by 2 minutes of exposure to Spanish-accented speech. Importantly, this was a task in which toddlers had previously failed when given no previous accent exposure (Schmale, Hollich & Seidl, 2011). This finding suggests that after hearing the unfamiliar Spanish accent, toddlers may have relaxed their categories to accommodate the nonstandard pronunciations that followed, or may have specifically relaxed them only to accommodate Spanish-accented speech. The argument was made that, since the familiarization did not provide toddlers with specific lexical or protolexical bootstraps, then general expansion must have been used. However, since the familiarization *did* include Spanish-accented speech, this previous study cannot demonstrate that a general expansion strategy is *sufficient* to accommodate a novel accent. And yet establishing whether this strategy is part of the toddlers' repertoire is a key theoretical and empirical question. Therefore, in the present work, we explore whether a general expansion strategy alone can support successful accent accommodation.

Experiment

We asked what types of variability would suffice to promote the use of a general expansion strategy in accent accommodation. Replicating the methodology of Schmale *et al*. (2012), we tested monolingual English-learning 24-month-olds on their ability to learn novel words when trained by a native English talker and tested by a Spanish-accented talker. In Schmale *et al*. (2012), toddlers were exposed to the speech of single or multiple native and Spanish-accented talkers before the word-learning experiment, and toddlers only learned the words in the foreign-accent conditions. In the present work, toddlers were assigned to one of two different Exposure conditions. In the *Indexical* exposure condition, toddlers heard 2 minutes of speech produced by four native English talkers with very different voices while watching an unrelated silent cartoon. In the *Social* exposure condition, toddlers watched 2 minutes of video of the same four talkers gesturing, but not speaking, while listening to classical music. Because the talkers differed greatly in age as well as sex, both conditions were exposed to a high degree of variability, but crucially not in terms of an accent. If toddlers require specific evidence on the effects of the accent on familiar words, then both groups should fail. In contrast, if variability along indexical and/or social dimensions is sufficient to trigger general expansion, then one or both groups will succeed at test, when presented with Spanish-accented speech.

Method

Participants

Sixty-five monolingual English-learning 24-month-olds with no history of hearing impairment were randomly

Table 1 *Participant characteristics by exposure condition: Average age (and range), Number of males/females, and Average productive vocabulary size (and range) according to the short form of the MacArthur-Bates Communicative Developmental Inventory: Words and Sentences (Fenson et al., 1994)*

Exposure condition	Age	Sex	Vocabulary size
1 Indexical	23.94 (23.42–24.54)	19/13	48.87 (11–94)*
2 Social	23.90 (23.39–24.74)	18/15	48.06 (14–100)

*A CDI was not available for one child.

assigned to one of two exposure conditions (Indexical: $N = 32$; Social: $N = 33$). Their general characteristics are reported in Table 1. An additional 26 children were not included because of the following reasons: crying, being overly restless, or refusing to sit in the chair to finish the study (12), equipment or experimenter error (4), foreign language exposure (5), and a productive vocabulary size of 5 or less according to the short form of the Fenson *et al*. (1994) MacArthur-Bates Communicative Developmental Inventory: Words and Sentences (5).

Design and stimuli

The experiment was a replication of the design in Schmale *et al*. (2012) and consisted of an exposure phase followed by a word-learning experiment (see Table 2 for an overview of the design).

Participants were randomly assigned to one of two Exposure conditions. In the Indexical exposure condition, all children heard the same four passages used in Schmale *et al*. (2012), but each was produced by four different native-English talkers with diverse voices. These talkers varied on dimensions of both sex and age, including a 10-year-old male, a 21-year-old male, a 21-year-old female, and a 64-year-old female. To reduce attrition, the passages were accompanied by a silent Curious George cartoon video. In the Social exposure condition, all children watched four videos of the same individuals that produced the passages in the Indexical exposure condition, who looked very different. In these videos, the individuals engaged in a series of nonlinguistic gestures (e.g. clapping, blowing kisses, giving thumbs up, saluting, shrugging, writing, playing peek-a-boo, waving, nodding) while maintaining a pleasant facial expression. To mirror the presentation of auditory and visual stimuli together as was used in the Indexical exposure condition, the videos were accompanied by a classical piano piece (Panizza, 2007). None of the talkers in the exposure conditions were used in the subsequent word-learning experiment. The peak amplitude (70 dB SPL) and duration (1 min, 42 sec) of all speech and music files were matched across Exposure conditions. See Figure 1 for examples of the Exposure conditions.

Table 2 *Experimental design (replication of Schmale et al., 2012)*

Exposure Phase
Word Learning Block A
Word Learning Block B
Word Learning Block A
Word Learning Block B

The Word-Learning Experiment was a replication of Schmale *et al*. (2012), whereby toddlers were trained and tested on two novel words by a female native English speaker and a Spanish-accented speaker. The visual stimuli consisted of four unidentifiable glass objects of different colors. The auditory stimuli consisted of four novel words (*choon*, *feem*, *moof*, *neech*) that were recorded within the sentences: 'Do you see a ____? Look, it's a ____! A ____!' Because it is expected that talkers with an accent will deviate from standard pronunciations along linguistically relevant dimensions (such as tenseness, in *neech*), stimuli were matched in linguistically *irrelevant* dimensions (e.g. intensity, pitch, and talkers' voice similarity ratings), but no effort was made to correct for the speaker's non-native accent. Consequently, the speakers varied naturally in all linguistically *relevant* dimensions, including first and second formant frequencies and duration of vowels. See Table 3 for specific acoustic measurements.

All words, objects, and sides were counterbalanced across participants within each exposure condition (Figure 1).

Procedure

In this experiment, toddlers were tested on a version of the preferential looking paradigm (Fagan, 1971), in which the child is seated on their caregiver's lap in front of a large video screen, where video images are projected. A hidden experimenter conducted the experiment while videotaping children's looking patterns to different objects on the screen. Parents were asked to wear sunglasses to prevent them from inadvertently influencing children's behavior. The children were instructed to look at the screen at the start of the experiment.

The experiment began with the Exposure phase, which played continuously for 1 min, 42 sec. The Word-Learning Experiment followed the Exposure phase and consisted of two six-trial word-learning blocks that were

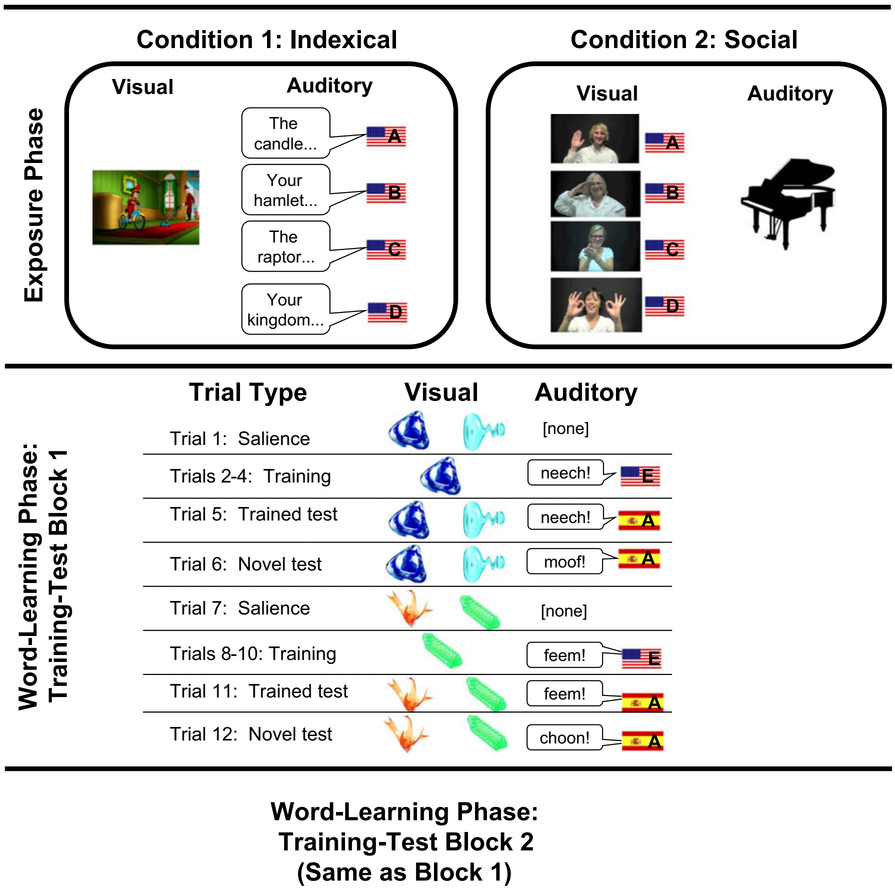

Figure 1 *Toddlers were randomly assigned to one of two Exposure conditions: Indexical or Social. The Exposure conditions differed in visual and auditory stimuli. In the Indexical exposure condition, toddlers listened to voices of four diverse native English actors while watching a silent Curious George video. In the Social exposure condition, toddlers listened to classical music while watching videos of the same four actors engaging in a series of gestures. After Exposure, toddlers were tested on a word-learning task with two repetitions of the same Training-Test block (only one block is shown here). All toddlers, regardless of exposure condition, were trained with the same North Midland American talker and tested with the same Spanish-accented talker. The objects, side of presentation, and labels in the Training-Test block were counterbalanced across toddlers within each Exposure condition.*

presented twice (see Table 2 and Figure 1 for an overview of the design). Since Schmale *et al.* (2012) reported a difference across testing blocks, we added a short delay period between the Exposure phase and the Word-Learning Experiment to allow any effects to stabilize. In the Word-Learning Experiment, each trial began with an attention getter: As soon as the child looked at this attention getter, the experimenter activated the visual and auditory stimuli (Figure 1). The first trial was a silent, Salience trial, in which two objects were presented on the right and left sides of the screen. This trial functioned to reduce the difference in exposure to the two test objects, one of which is presented multiple times during training ('trained' object), while the other only appears at test ('novel' object). The following three trials were Training trials, during which toddlers heard three repetitions of a novel word (e.g. *moof*) by a native female speaker of Midwestern English and the associated object was presented at the center of the screen. The last two trials of the Word-Learning Experiment were Test trials, in which the trained and novel objects were presented on the right and left sides of the screen while a female Spanish-accented speaker provided the labels. In Trained test trials, the label provided was the same as the one provided in training (*moof*). In Novel test trials, a new label was used (e.g. *neech*). This Novel test trial (which tested word learning through mutual exclusivity) functioned as a control for any trained object preference that may develop in the Trained test trial from greater exposure to the trained object in the Training trials. The presentation order of test trials was counterbalanced across participants such that half of the participants

Table 3 *Average acoustic measurements and standard deviations (in italics) of vowels of target words*

Produced by talkers in the Word-Learning Experiment				
	Native talker		Spanish-accented talker	
Vowel	/i:/	/u:/	/i:/	/u:/
F0 (Hz)	378.33	341.17	403.83	379.33
	158.38	*91.96*	*78.13*	*118.58*
F1 (Hz)	503.83	482.83	501.33	467.67
	117.18	*124.67*	*97.97*	*161.6*
F2 (Hz)	2447	1657	2847.17	1301.5
	575.83	*315.8*	*346.12*	*269.81*
Duration (ms)	278.2	249.4	159.97	158.19
	73.98	*42.14*	*17.04*	*62.47*

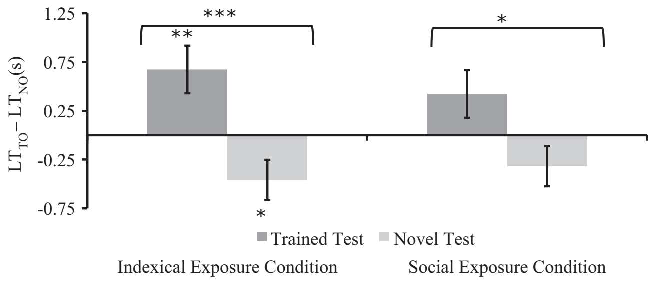

Figure 2 *Mean difference looking times (Trained object minus Novel object, $LT_{TO} - LT_{NO}$) during trials in which the Trained or the Novel labels are provided in the Indexical and Social exposure conditions. Error bars indicate standard error. Stars over bars indicate the significance level in one-sample two-tailed t-tests against zero; asterisks over groups of bars represent the significance level in a RM ANOVA within each exposure condition.*

received the Trained test trial first and half of the participants received the Novel test trial first.

Results

The videos of toddlers' looking patterns were digitized at 30 frames per second and coded offline by a highly trained coder. Looking times to each object in each trial were measured over a 2-second period, starting 200 ms after the onset of the label and analyzed by trial type. If children recognize the newly learned words, they should exhibit longer looking times to the trained object than to the novel object in the Trained test trial, but higher looking times to the novel object than the trained one in the Novel test trial.

Statistical analyses were carried out in R (R Development Core Team, 2009). The dependent measure was infants' total looking time to the trained object minus their total looking time to the novel object, within each of the test trials (Trained and Novel). If toddlers successfully recognize the recently learned words, this average should be reliably above zero for Trained trials (where the label refers to the trained object), and reliably below zero for Novel trials (where the label refers to a novel object). A repeated measures analysis of variance (RM ANOVA) was fit on the difference in looking times averaged across the two blocks, declaring trial type (Trained, Novel) as within- and Exposure condition (Indexical, Social) as between-subjects factors. This analysis revealed a main effect of trial type [$F(1, 63) = 19.02$, $p < .001$]. The effect of trial type emerged because the difference in looking times was significantly higher in Trained trials than it was in Novel trials: One-sample two-tailed *t*-tests showed that this difference was significantly greater than zero for Trained trials [$t(64) = 3.14$, $p = .002$; 95% confidence interval .20 to .89], and significantly below zero for Novel trials [$t(64) = -2.65$, $p = .01$; 95% confidence interval −.68 to −.09]. The RM ANOVA revealed no effect of exposure [$F(1, 63) = .05$] and, most importantly, no interaction [$F(1, 63) = 0.84$, $p > .36$]. This lack of interaction indicates that, although effects were somewhat weaker in the Social exposure condition, infants' performance did not differ significantly across these two conditions. This should be considered when inspecting Figure 2, where, for the purposes of completeness, results are separated by condition. (Data, scripts, and additional analyses are available in the online materials.[1])

The same result ensued in an ANOVA declaring Trial type as a within-subject factor and Exposure as a between-subjects factor and the difference scores averaged across blocks as dependent measure [Trial type $F(1, 54) = 8.1$, $p = .006$; Exposure and the interaction $p > .87$].

[1] Three previous studies have documented that toddlers fail at this word-learning experiment: (1) When they are presented with the task on its own (Schmale *et al*., 2011); (2) When the task is preceded by an Exposure period with the same native English talker used in the word-learning experiment (Schmale *et al*., 2012); and (3) When the task is preceded by an Exposure period spoken by four native English talkers with very similar voices (Schmale *et al*., 2012). Given that we introduced a delay period between Exposure and Word-Learning, we ran a fourth control experiment including this delay, where passages were spoken by a female, native English talker, who was different from the English talker in the Word-Learning Experiment. This group, like the three others before it who did not have a delay phase, failed at test [$F(1, 25) = .04$, $p = .84$]. A RM ANOVA including the two conditions in the main experiment and this control experiment revealed an interaction between Trial type and Experiment [$F(2, 88) = 3.39$, $p = .04$].

Discussion

Previous work has provided compelling evidence that toddlers can use a linguistically informed, sophisticated strategy to accommodate a novel accent, by noticing systematic deviations in the pronunciations of well-known words or word forms. In the present study, we sought to assess whether toddlers can also rely on an unsophisticated strategy: seeing or hearing diverse humans leads them to accept diverse pronunciations. Results suggest that exposure to indexical variability (people with very different voices), as well as social variability (people who look very different), suffices to promote accommodation of an unfamiliar accent in word learning. Therefore, toddlers do not require the presence of lexical information to accommodate unfamiliar accents. This naturally does not negate the importance of the well-established lexically-based strategy. Instead, this study is the first to provide evidence that the *general expansion strategy* is (1) employed by toddlers and (2) sufficient in accent accommodation. This evidence thus provides support for a heretofore-unexplored mechanism underlying adaptation to non-standard pronunciations.

One possible interpretation of these findings is that exposure to multiple talkers who vary in indexical or physical characteristics modulates young children's expectations for what type of speech they are likely to subsequently encounter. For example, if a child hears speech from four talkers who all have very distinct voices, they may relax their categories in anticipation that the fifth talker they hear will be distinct as well. Thus, instead of providing evidence for how children are mapping input to output, as much of the previous work on accent accommodation has done, the present work may have uncovered a mechanism for the detection of an upcoming irregular speech pattern.

While the present study has demonstrated that general expansion is a viable strategy, several open questions remain. First, it is unclear the precise range of variability that is required to trigger the employment of this strategy. Previous work demonstrates that mere multi-talker exposure does not support the use of the general expansion strategy: Toddlers fail in a cross-accent word-learning paradigm when they are exposed to four female talkers with similar voices (Schmale *et al*., 2012; see also footnote 1). It appears that in order for 24-month-olds to exhibit success in this task, they require exposure to some type of nonstandard or highly variable input, such as prior exposure to the accent or to multiple speakers (via either visual or auditory means). Moreover, it could be that general expansion only occurs in the presence of socially relevant stimuli such as speech (as in the Indexical exposure condition) or videos with communicative gestures (as in the Social exposure condition), but based on the current results we cannot be certain of the precise mechanisms that allow for this accommodation. Indeed, accommodation may have even been triggered by the piano music accompanying the videos in the Social exposure condition, an even less socially relevant signal. It will remain to future work to explore the precise range and type of variability sufficient to activate this strategy.

A second open question regarding the prerequisites for the use of the general expansion strategy is the perceiver's level of linguistic expertise in general, or in specific situations. As to the former, since this strategy is not based on a lexicon or even a protolexicon, it should be available from very early on in development. In fact, recent work suggests that infants profit from variable speech even in the first half of the first year. For example, Seidl, Cristià and Onishi (2014) report that 4-month-old infants are better able to learn phonotactic patterns when they are presented with variable talkers (as opposed to a single talker), suggesting a boost in learning long before infants accumulate even a small protolexicon. Furthermore, the general expansion strategy may be blocked in specific situations that the child has already mastered. For instance, it may be used in the present word-learning task, but not during recognition of highly familiar words because the child is more confident of the pronunciation of the latter than the former.

A final avenue for exploration concerns a more detailed description of the general expansion strategy as an accommodation mechanism and its effects on processing. Specifically, while a lexically-based strategy makes very specific predictions regarding the kinds of mismatches that are tolerated in the signal, the general expansion strategy does not. It is then unclear how precisely children perform a lexical search, and whether this renders the process of word recognition less efficient in some way. Studies using a finer design, involving eye-tracking and minimal pair competitors, may be able to shed light on this question.

Thus, the experiment presented here opens an exciting avenue of research, which further informs our understanding of how children accommodate variable speech. We document that toddlers exposed to variable voices, and even to different-looking talkers accompanied by music, become more accepting of non-standard pronunciations. This unsophisticated, but effective strategy suggests that children are well prepared to communicate with those who speak differently.

References

Barker, B.A., & Meyer Turner, L. (in press). Influences of foreign accent on preschoolers' word recognition and story comprehension. *Applied Psycholinguistics.*

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B., & Quann, C.A. (2009). Development of phonological constancy: toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, **20**, 539–542.

Fagan, J. (1971). Infant recognition memory for a series of visual stimuli. *Journal of Experimental Child Psychology*, **11**, 244–250.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59** (5, Serial No. 242).

Houston, D.M. (2000). The role of talker variability in infant word representations. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Kovacs, A.M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences, USA*, **106**, 6556–6560.

Mulak, K.E., Best, C.T., Tyler, M.D., Kitamura, C., & Irwin, J. (2013). Development of phonological constancy: 19-month--olds, but not 15-month-olds, identify familiar words spoken in a non-native regional accent. *Child Development*, **84**, 2064–2078.

Panizza, A. (2007). Suite De Danzas Criollas (Adaggietto Pianissimo, Opus 15). Complete Piano Works by Alberto Ginastera, Vol. II. Buenos Aires: Tradition Recordings.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Version 2.15.1 (2012-06-22).

Rost, G.C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, **12**, 339–349.

Schmale, R., Cristià, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, **15** (6), 732–738.

Schmale, R., Hollich, G.J., & Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, **38**, 1096–1108.

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science*, **12**, 583–601.

Seidl, A., Cristia, A., & Onishi, K. (2014). Talker variation aids infants' phonotactic learning. *Language, Learning, and Development*, **10**, 297–307.

Van Heugten, M., & Johnson, E.K. (2014). Learning to contend with accents in infancy: benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, **143**, 340–350.

White, K.S., & Aslin, R.N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, **14**, 372–384.

Received: 16 September 2013
Accepted: 29 July 2014