Does a Vowel by Any Other Accent Sound the Same … to Toddler Ears?

Catherine T. Best[1,2], Christine Kitamura[1], Sophie Gates[1] and Angela Carpenter[3]
[1]Western Sydney University, Australia
[2]Haskins Laboratories, USA
[3]Wellesley College, USA
{c.best, c.kitamura, s.gates}@westernsydney.edu.au, acarpent@wellesley.edu

Abstract

Research on spoken word recognition in young children has emphasized detection of minimal phonetic contrasts, and offers conflicting evidence about the role of consonants versus vowels. The complementary ability to recognize words across natural phonetic variation, *phonological constancy*, is equally important to language development. Prior studies of phonological constancy found that 15- and 19-month-olds recognize familiar toddler words in an unfamiliar regional accent containing Category Goodness vowel and/or consonant differences from their native accent, or Category Shifting consonant differences. In the present study, Category Shifting vowel differences disrupted word recognition at both ages, supporting different roles for vowels than consonants.

Index Terms: early word recognition, regional accent variation, perceptual assimilation, vowels versus consonants

1. Introduction

Over four decades of research on infants' perceptual attunement to their native language have provided many insights into experiential effects on speech perception over the course of the first year. Infants under 8 months discriminate many non-native consonant contrasts, but show a dramatic decline in discriminating many, though not all, of these contrasts by 9-10 months. The decline in discrimination appears earlier for vowel contrasts, by around 5-6 months [e.g., 1-4]. Conversely, for some native consonant contrasts, discrimination improves over the first year or so [5, 6]. Importantly, however, some non-native contrasts continue to be discriminated well even into adulthood, suggesting that changes in perception of non-native speech distinctions cannot be attributed solely to lack of early exposure to their specific surface-level acoustic-phonetic properties. Rather, as posited by the Perceptual Assimilation Model (PAM) [7, 8], developmental shifts in perception of both native and non-native contrasts must reflect the emerging ability to map varying phonetic details to more abstract native phonological structures, such as spoken words and their component phonological elements. Thus, when a listener perceives the members of a non-native phonetic distinction as equivalent exemplars of the same single native phoneme (Single Category assimilation: SC), discrimination is poor because they perceive the phonetic difference as not phonologically or lexically contrastive in their language. But if they perceive a non-native distinction as corresponding to a native phonological contrast (Two Category assimilation: TC), they continue to discriminate it well. Moreover, they remain sensitive to some within-category phonetic variation if they detect it as a difference in goodness of fit to a single native phoneme, which they discriminate moderately well though not as well as a TC contrast (Category Goodness: CG) [9, 10].

An important theoretical issue not considered in the cross-language perception literature is the role that perceptual assimilation may play in children's growing ability to recognize words (and their component consonants and vowels) across various types of natural phonetic variations that they encounter *within* their native language. Recognizing that critical phonetic differences can convey phonological distinctions between native words, e.g., that /**p**iz/ (***p**eas*) is not the same word as /**k**iz/ (***k**eys*), is obviously crucial to language development. However, of equal or greater importance is that the child also needs to develop the ability to recognize words across the lexically-irrelevant phonetic variations presented by different speakers and regional accents of their language, i.e., to recognize the *phonological constancy* of words [1, 11-13]. Phonological constancy requires perceptual assimilating those types of variations to their common, underlying phonological forms.

Less is known about the emergence of phonological constancy, but research is growing on the topic. The first such study examined 15- and 19-month-olds' recognition of familiar words (i.e., known to toddlers), as indexed by a preference for listening to sets of toddler words over sets of unfamiliar words (i.e., low-frequency adult words), in their own native regional English accent versus in a phonetically-differing regional English accent they had not previously been exposed to. Whereas both age groups showed a listening preference for the familiar toddler words, only the 19-month-olds showed this preference when tested with the unfamiliar accent, indicating that they but not the younger children had achieved phonological constancy [11]. A follow-up eye-tracking study assessed whether the same age groups could identify the familiar toddler words in a visual preference task, by looking more at the named object than the unnamed one (distractor) in pairs of photographs. Both ages showed a reliable looking preference for words in their native accent, but again only the 19-month-olds did so when the words were spoken in the unfamiliar accent [13]. Two subsequent studies found a relationship between children's expressive vocabulary size and their listening preference for familiar toddler words under task conditions with high stimulus variability (more speakers, words and tokens than in [11]). 15- and 17-month-olds with small vocabularies ($\leq$ 25 words) failed to show a reliable preference for toddler words in either accent, whereas 17-month olds with larger vocabularies ($\geq$ 50 words) showed a familiar toddler word preference in their native accent and 19-month-olds with even larger vocabularies ($\geq$ 100 words) preferred familiar toddler words in both their native and the unfamiliar accent [14, 15]. Thus, vocabulary development is linked to recognizing the phonological constancy of words across moderately high phonetic variation, first in the native accent, and later for the greater variations of an accent not previously experienced.

PAM has since been extended to predict assimilation patterns for regional accent variation in the phonetic realizations of vowels versus consonants within the native language, as perceived by both adults [16-18] and toddlers [19-22]. For many native phonemes regional accent variations are assimilated as Native-Like (NL), i.e., the phonetic deviation from the native accent realization is small enough as to be perceptually insignificant. For others, however, the phonetic deviation is more noticeable as a Category Goodness (CG) difference within the matching native phoneme, i.e., perceived as the correct consonant or vowel but also heard as being pronounced differently than in the native accent. In striking distinction, though, some accent differences may transgress native-accent category boundaries, and thus be perceived as a different consonant or vowel than the other-accent speaker intended, i.e., the difference is Category Shifting (CS). CG versus CS cross-accent assimilations are expected to have notably different influences on recognition of spoken words in the other, non-native accent: correctly though perhaps more slowly identified for CG differences, but incorrectly identified, even by adults, for CS differences. Moreover, the developmental differences for discrimination of non-native vowel versus consonant contrasts summarized earlier suggest that these effects may differ for vowel versus consonant deviations from the native accent. Importantly, the relative role of consonants versus vowels in early word learning is under debate. For example, 11-16-month-olds have been reported to rely more on consonants than vowels [23], or to show symmetrical sensitivities to vowels and consonants [24], in recognition of known words, and have been claimed to rely more on vowels [25], or more on consonants [26], in learning new words.

Given that debate and the previous mixed findings, we ran a series of cross-accent word listening preference studies examined the impact of CG versus CS differences from the children's native accent (Australian English: AusE) in an unfamiliar regional accent's vowel pronunciations (JaME: Jamaican Mesolect English) or consonant pronunciations (London "Cockney" English: CknE). In those studies, for the first time, not only 19- but also 15-month-olds showed a familiar toddler word preference in both AusE and the unfamiliar accent, when the word sets showed only CG differences in either the vowels (JaME) or the consonants (CknE). Conversely, when there were CS differences in both consonants and vowels (JaME), neither age recognized the familiar toddler words in the non-AusE accent. Strikingly, however, when only the consonants showed CS differences (CknE), both age groups did generalize the familiar toddler word preference to the unfamiliar accent [19-22]. Together, those findings indicate, firstly, that even by 15 months, toddlers assimilate CG vowel variations, and both CG and CS consonant variations, in unfamiliar accents to their (native-accented) representations of known words. Secondly, in light of that pattern, the failure of both age groups to recognize words across CS differences in both consonants and vowels implies that their difficulty was due specifically to the CS *vowel* differences, not to the CS consonant differences. But this inference was not directly assessed. Moreover, the two ages may well differ in how they respond to CS vowel differences alone. Therefore, the current study tested whether restricting the pronunciation differences to CS vowel deviations (JaME) from the native accent (AusE) would disrupt 15- and/or 19-month-olds' recognition of familiar toddler words.

2. Method

The virtually identical findings for toddlers' visual identification of words via eye-tracking [13] and for their listening preferences in other studies of cross-accent word recognition [11, 14, 15, 19-22] imply that both tasks index lexical recognition. Therefore, we used the listening preference task for comparison to the other CG-CS vowel and consonant studies, using CS vowel-differing words in AusE versus JaME. We again compared children at 13-15 months ("15 month-olds"), i.e., the early word-learning period (< 25 word expressive vocabulary), and 18-20 months ("19-month-olds"), who have typically reached a 50+ word vocabulary ('vocabulary spurt').

2.1. Participants

Two sets of participants successfully completed the listening preference test in both target accents. The younger group ('15 months') had 32 children (M_{age} = 14.05 mo, range = 13.28 – 15.22 mo; 17 females), as did the older group ('19 months': M_{age} = 19.19 mo, range = 18.51 – 20.22 mo; 17 females). All were full-term at birth, healthy on the test day, lacked familial speech/language disorders, and received little to no exposure to other languages or non-AusE accents including JaME.

30 additional children at the younger age, and 36 at the older age, were tested but excluded due to fussing/crying (n = 52), falling asleep (n = 2), climbing out of the parent's lap during testing (n = 1), parental interference (n = 6) or withdrawal from testing (n = 1), or technical problems (n = 4). This rejection rate (~50%) is typical of 1-2 year olds across a range of tasks, commensurate with their general behavioral tendencies.

2.2. Stimulus materials

The target items were multiple audio tokens of words produced by several speakers of each comparison accent, for each of the two main word types. The listening preference task also uses static visual stimulus displays, such that participants could control audio stimulus presentations by fixating their gaze on the display.

2.2.1. Visual fixation displays

The visual fixation displays were colored checkerboards against a white background, with a central circular "swirl" to attract infants' attention to the center of the screen. Two different checkerboard colors (magenta, blue) were used for each child's AusE vs. JaME preference tests, with the color assignment to each test counter-balanced across participants.

2.2.2. Spoken word audio stimuli

The listening preference task presents separate, alternating trials containing sets of words known to toddlers vs. sets of low frequency adult words they are highly unlikely to have ever heard (see Table 1). Each word set (Familiar [toddler]; Unfamiliar [adult]) comprised 8 monosyllabic and 8 bisyllabic target words, carefully selected such that: a) Familiar words occur in >50% of AusE expressive vocabularies at 13-15 months [27] and/or appear often in toddler picture-books, whereas Unfamiliar words occur ≤ 2/million in standard and Australian English lexical databases [28, 29]; b) each word contained one stressed vowel that differed in a Category-Shifting way between AusE and JaME, i.e., adult AusE listeners hear the JaME vowel as a categorically different one than the speaker intended. All other phonemes in each target word displayed only NL or CG differences between AusE-JaME.

We recorded three female native speakers each of AusE (from western Sydney, Australia) and of JaME (from St Catherine and St. James parishes, Jamaica). We paired speakers across the two accents to assure similar voice quality, F0 mean and range across the AusE and JaME stimulus sets (one JaME speaker's F0 was raised ~10 Hz via Praat resynthesis to better match her AusE pair). Each speaker produced multiple tokens of each target word, using a Shure SM10A headset microphone connected to a Sony PCMM1 portable DAT recorder (44.1 kHz sampling rate). The printed targets were shown in quasi-random order on a laptop screen. For the task, two tokens per target word were selected from each speaker, resulting in a total of 192 tokens for each regional accent (16 words x 2 word sets x 2 repetitions x 3 speakers). Thus, a child was highly unlikely to hear any token more than once in the tests.

Table 1. *Target words for each word set in each accent.*

Single syllable		Two syllable	
Familiar (toddler)	**Unfamiliar (adult)**	**Familiar (toddler)**	**Unfamiliar (adult)**
ball	shawl	baby	frailty
bear	mare	doggy	blobby
boat	dose	flower	doubter
bus	shun	grandma	vanguard
door	gore	lolly	fauna
duck	muck	mummy	putty
nose	foes	paper	taper
socks	knocks	water	spotty

2.3. Procedure

We used a word-type listening preference task [30, 31] in which a series of alternating trials plays out a set of familiar (toddler) words versus a set of unfamiliar (adult) words, for as long as the child fixates on the checkerboard display directly in front of them. When the child looks away for $\geq$ 2 seconds, the trial ends and the checkerboard flashes on/off until the child looks back, at which time the display stabilizes again and the next trial begins. A trained observer monitors child fixations via a hidden low-light video camera below the checkerboard. The child sits in the parent's lap (who is instructed not to point or interfere, and listens to vocal music over Sennheiser HD650 circumaural headphones to mask test stimuli). A significant listening preference (higher fixation times) for the familiar word trials relative to the unfamiliar word trials is taken to index recognition of the familiar toddler words.

Each participant completed two preference tests of eight trials per test (four trials per word set, in alternating trials). In one test all words were spoken in the native AusE accent; in the other they were all spoken in the unfamiliar JaME accent (test order counterbalanced across participants), as in previous cross-accent word recognition studies [11, 14, 15, 19-21].

3. Results

Total fixation times were summed across familiar word trials, and separately across unfamiliar word trials. These data were submitted to a 3-way Analysis of Variance (ANOVA) on the factors age group (15 vs 19 months) x word set (familiar vs unfamiliar) x accent (AusE/native vs. JaME). (An initial 4-way ANOVA found no significant effects of accent order).

A significant main effect of word type, $F_{(1,62)}$ = 14.030, p < 0.0001, indicates an overall listening preference for familiar toddler words over unfamiliar adult words. The main effect for accent, $F_{(1,62)}$ = 4.927, p = 0.03, revealed significantly greater overall fixation during the AusE test than the JaME test. However, these were qualified by a significant word set x accent interaction, $F_{(1,62)}$ = 9.742, p = 0.003, showing that the familiar word preference was reliable only for AusE, not for JaME (see Figure 1). This held true for both ages: the main effect of age was not significant, nor did it interact with word set or accent.

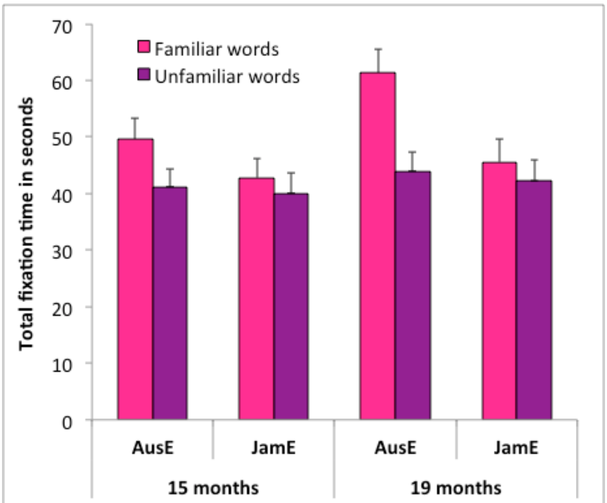

Figure 1: *Total listening times to familiar toddler vs. unfamiliar adult word sets spoken in the native accent (AusE) vs. the unfamiliar accent (JaME), for younger (15 months) and older toddlers (19 months). Error bars are standard errors of the means (s.e.m.).*

4. Discussion

The current findings are compatible with the picture of results from prior investigations of toddlers' ability to recognize familiar toddler words spoken in an unfamiliar accent when the differences from the native accent (AusE) were restricted to either a CG or a CS difference in vowels and/or consonants [19-22]. Here, we found that word recognition was disrupted for the unfamiliar JaME accent in both the 15- and 19-month groups when the JaME pronunciation showed a single CS vowel difference from their native AusE accent. By comparison, the previous studies had found word recognition to remain intact for an unfamiliar accent that differs from native AusE in CG vowel differences alone or in either CG *or* CS consonant differences, but that CS differences in vowels *and* consonants disrupts cross-accent recognition of familiar toddler words. The latter finding, then, appears to have been due solely to the CS vowel differences, and not to the CS consonant differences. Analogously, AusE-speaking adults show marked differences in how they perceptually assimilate vowel versus consonant differences of other English regional accents relative to AusE [16-18]. These cross-accent findings thus extend, and offer some challenges to, previous reports on differential roles for vowels and consonant in young children's learning and recognition of spoken words [23, 25, 26, *cf* 24].

Note, however, that this research has thus far only examined CG and CS vowel versus consonant differences among regional accents of English. In English, vowel pronunciation differences provide the primary source of regional accent variation; consonant differences are much more restricted in English. There are languages in which regional variation instead involves more consonant than vowel variation, for example as a result of variable consonantal lenition processes, such as Spanish. Examining these effects across accents of such languages will be important to teasing out the basis for differential tolerance of the two types of pronunciation variations.

5. Conclusions

We speculate that perceptual assimilation operates not only for non-native speech perception but also within the listeners' native language. It reflects the necessity to handle phonological abstraction across natural, systematic phonetic variation within the language, and is evident in the early development of spoken word recognition once the child has achieved phonological constancy. Furthermore, this developmental achievement appears to proceed differently for accommodating Category Goodness (CG) differences and Category Shifting (CS) differences between the native accent and other unfamiliar regional accents, and in particular for CS vowel versus consonant differences. But these effects have only been investigated in English, in which regional accents differ more in vowel than consonant pronunciations. Research on languages with more consonantal than vowel regional variation are needed to evaluate the extent to which, and conditions under which, vowel versus consonant effects in spoken word recognition reflect universal versus language-specific principles.

6. Acknowledgements

Research supported by Australian Research Council grants DP0772441 and DP130104237.

7. References

[1] Best, C. T. "Devil or angel in the details? Complementary principles of phonetic variation provide the key to phonological structure", in J. Romero and M. Riera [Eds] Sounds, representations and methodologies: Essays on the phonetics-phonology interface, 3-31. John Benjamins, 2015.

[2] Best, C. T. "Speech perception in infants: Propagating the effects of language experience", in E. M. Fernandez & H. S. Cairns [Eds], Handbook of psycholinguistics. Wiley, in press.

[3] Best , C. T. "Learning to perceive the sound pattern of English", in C. Rovee-Collier and L. Lipsitt [Eds], Advances in infancy research, 217-304, Ablex, 1994.

[4] Maurer, D. and Werker, J. F. "Perceptual narrowing during infancy: A comparison of language and faces", Developmental Psychobiology, 56(2):154-178, 2014.

[5] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. and Iverson, P. "Infants show a facilitation effect for native language phonetic perception between 6 and 12 months", Developmental Science, 9:F13–F21, 2006.

[6] Polka, L., Colantonio, C. and Sundara, M. "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern", J. Acou. Soc. Am., 109:2190-2201, 2001.

[7] Best, C. T. "A direct realist view of cross-language speech perception", in W. Strange [Ed], Speech perception and linguistic experience, 171–206, York Press, 1995.

[8] Best, C. T. and Tyler, M. D. Nonnative and second-language speech perception: Commonalities and complementarities", in M. Munro and O-S. Bohn [Eds], Second language speech learning, 13-34, John Benjamins, 2007.

[9] Best, C. T., McRoberts, G. W. and Sithole, N. M. "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants", J. Exp. Psych.: Human Perc. Perf., 14:45-60, 1988.

[10] Best, C. T., McRoberts, G. W. and Goodell, E. "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology", J. Acou. Soc. Am., 1097:775-794, 2001.

[11] Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B. and Quann, C.A. "Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words", Psych. Science, 20:539-542, 2009.

[12] Watson, T. L., Robbins, R. A. and Best, C. T. "Development of face and speech perception: An integrated approach", Developmental Psychobiology, 56:1454-1481, 2014.

[13] Mulak, K. and Best, C. T. "Development of word recognition across speakers and accents", in L. Gogate and G. Hollich [Eds], Theoretical and computational models of word learning, 242–269, IGI Global-Robotics, 2013.

[14] Best, C. T., Tyler, M. D. Kitamura, C., Notley, A. and Bundgaard-Nielsen, R. "Phonetic specificity of early words? Australian toddlers' perception of Australian versus Jamaican English pronunciations", Presented at International Conference on Infant Studies, Vancouver, March 2008.

[15] Best, C. T. Tyler, M. D., Kitamura, C. and Bundgaard-Nielsen, R. Vocabulary size at 17 months and the emergence of phonological constancy in word recognition across native and nonnative dialects. Presented at International Conference on Infant Studies, Baltimore, March 2010.

[16] Best, C. T., Shaw, J., Mulak, K., Docherty, G., Evans, B., Foulkes, P. Hay, J., Al-Tamimi, J., Mair, K., and Wood, S. Perceiving and adapting to regional accent differences among vowel subsystems. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.

[17] Best, C. T., Shaw, J., Mulak, K., Docherty, G., Evans, B., Foulkes, P. Hay, J., Al-Tamimi, J., Mair, K., and Wood, S. "From Newcastle MOUTH to Aussie ears: Australians' perceptual assimilation and adaptation for Newcastle UK vowels", *Proceedings of Interspeech*. Dresden, 2015.

[18] Shaw, J., Best, C. T., Mulak, K., Docherty, G., Evans, B., Foulkes, P. Hay, J., Al-Tamimi, J., Mair, K., Peek, M. and Wood, S. "Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants", Presented at Australasian Speech Science and Technology, Christchurch, December, 2014.

[19] Best, C. T., Kitamura, C., Pal, M. and Dwyer, A. "Young toddlers recognize non-native Jamaican-accented words differing only in 'category-goodness' from native-accent vowel pronunciations", presented at International Conference on Infant Studies, Minneapolis, June 2012.

[20] Best, C. T., Gates, S., Kitamura, C., Docherty, G., Pinet, M. and Evans, B. G. "Young Australian toddlers recognize Cockney-accented words involving only 'category-goodness' differences in consonant pronunciations from Australian English", Presented at International Conference on Infant Studies, Berlin, July 2014.

[21] Best, C. T. and Kitamura, C. The role of perceptual assimilation in early development of recognition of words spoken in native vs unfamiliar regional accents", Presented at International Conference on Infant Studies, Berlin, July 2014).

[22] Best, C. T., Gates, S., Kitamura, C., Docherty, G. and Evans, B. "Category-shifting consonant differences between English accents do not interfere with familiar word recognition at 14 or 19 months", Presented at International Conference in Infant Studies. New Orleans, May 2016.

[23] Hochmann, J-R., Benavides-Varela, S., Nespor, M. and Mehler, J. "Consonants and vowels: Different roles in early language acquisition", Dev. Sci., 14(6):1445-1458, 2011.

[24] Mani, N., and Plunkett, K. "Phonological specificity of vowels and consonants in early lexical representations", J. Mem. Lang., 57(2):252-272, 2007.

[25] Pons, F., and Toro, J. M. "Structural generalizations over consonants and vowels in 11-month-old infants", Cognition, 116(3):361-367, 2010.

[26] Havy, M., and Nazzi, T. "Better processing of consonantal over vocalic information in word learning at 16 months of age", Infancy, 14(4):439-456, 2009).

[27] Kalashnikova, M., Schwarz, I. C. and Burnham, D. "OZI: Australian English Communicative Development Inventory", First Language, 0142723716648846, 2016.

[28] Baayen, R. H., Piepenbrock, R., & Van Rijn, H. "The CELEX lexical data base, release 2 [CD-ROM]", Linguistic Data Consortium, 1995. Available online: http://www.ldc.upenn.edu/

[29] Dennis, S. "The Sydney Morning Herald Word Database", Noetica: Open Forum, 1995. http://psy.uq.edu.au/cogpsych/noetica/.

[30] Hallé, P. A., and de Boysson-Bardies, B. "Emergence of an early receptive lexicon: Infants' recognition of words", Infant Behavior and Development, 17(2):119-129, 1994.

[31] Hallé, P. A., & de Boysson-Bardies, B. "The format of representation of recognized words in infants' early receptive lexicon", Infant Behavior and Development, 19(4):463-481, 1996.