REPRODUCING SINGLETONS AND FAKE GEMINATES

Melissa A. Redford & Grace E. Oh

University of Oregon, USA
redford@uoregon.edu; goh@uoregon.edu

ABSTRACT

Thirty-two native American-English speaking children and adults reliably reproduced a non-contrastive segmental length distinction encoded in a set of monomorphemic nonce words. However, participants were unable to explicitly identify the contrast in a same/different judgment task. The results support the view that grammatically irrelevant, lexically specific temporal patterns can be learned and are represented in the lexicon.

Keywords: implicit phonetic imitation, consonant duration, relative timing, fake geminates

1. INTRODUCTION

Fake geminates are phonetically long segments that arise when morpheme concatenation results in a sequence of two identical segments. In recent work, we found that English word-internal fake geminates are longer than matched singletons in absolute and relative terms, but that English word boundary fake geminates are only longer than matched singletons in absolute terms [8]. These results were taken to indicate that word-internal fake geminates are represented as a single long consonant, whereas word boundary fake geminates are represented as two identical consonants crossing a morpheme boundary.

The idea that English speakers might represent segmental length follows from the well-vetted assumption that heteromorphemic words have holistic lexical representations (e.g., [1, 4, 10]) and from the assumption that these representations encode patterns of relative timing [1, 2]. Of course, if lexically specific relative timing patterns are represented, then they must be learned when words are learned. Learning such patterns is plausible given the literature on implicit phonetic imitation, which indicates that speakers reproduce grammatically irrelevant acoustic details without being explicitly aware of these details (e.g., [6, 7, 11]).

Then again, previous studies on phonetic imitation have principally focused on global shifts in production along one or more dimensions rather than on the imitation of a specific phonetic contrast. For example, Neilsen [7] found a global shift in VOT durations relative to a baseline measure following exposure to word lists with longer-than-average VOTs. This type of shift might be explained as an adjustment to the phonetic grammar rather than to specific lexical items. By contrast, the hypothesis that English speakers represent both long and short versions of the same consonant requires that timing be represented on a lexeme-by-lexeme. If this is correct, then it follows that speakers should be able to imitate lexically specific timing patterns without additional linguistic or social cues to their existence.

The primary goal of the current study was to determine whether speakers are able to reproduce lexically specific timing patterns that encode a non-contrastive consonantal duration difference. A secondary goal was to determine whether accurate imitation of the contrast implies its explicit perceptual identification. Explicit identification of the contrast in the presence of accurate imitation could imply that perceptually based exemplar representations guide production. If speakers reproduce a contrast that they are unable to explicitly identify, then the guiding representations are unlikely to be grounded in perception (i.e., acoustically specified).

A final goal of the study was to investigate whether children are better able to reproduce and/or hear a non-contrastive consonantal duration difference than adults. Age-dependent differences in performance were anticipated because children are presumed to be learning words at a faster rate than adults and to have less entrenched a priori expectations about English sound patterns.

2. METHODS

2.1. Participants

Participants were 16 children between the ages of 5;5 and 7;4 years (mean = 6;5) and 16 college aged adults. All participants were native Standard American English speakers. The child and adult participants were free of speech and hearing problems as determined by parental report and a pure-tone hearing screen for the children and by self-report for the adults.

2.2. Materials

Twenty tri-syllabic CVCVCV(C) nonce word stimuli were designed with a lax vowel in the syllable 1 (V1), a tense or diphthongized vowel in syllable 2 (V2), and a nasal consonantal onset in syllable 2 (C2), which was labial in half of the words and alveolar in the other half. All other Cs were oral sounds, final Cs were liquids, and the final V was /i/: e.g., /dɪmoʊbəl/, /ʃɪnæɾɚ/ /gɛmagi/. Target words were written in IPA and produced with stress on syllable 2 and child-directed prosody by a native English speaking adult female in the following context: *This is a X*. *That is a Y*. *Can you say X*? *Can you say Y*? where X and Y represented a random and exhaustive pairing of two of the 20 nonce words.

Although uttered together, the phrases were subsequently segmented at the IP boundaries and the single best productions of each target word in its statement and question context were selected. Once selected, all audio files were copied and the target words altered to produce 2 sets of identical phrases containing nonce words with short and long nasal consonants. Specifically, the nasal consonant in the target word was lengthened 1.6 times in one copy and 1.9 times in another copy to produce target words with long consonants that were otherwise identical to the variants with short consonants. The condition with short and long nasals that differed by a factor of 1.6 will be referred to as the *smaller difference* condition, and those that differed by a factor of 1.9 as the *larger difference* condition. Accurate reproduction and perceptual identification was deemed more likely in the larger difference condition than in the smaller one.

All altered and unaltered files were copied again and the target words excised from their statement contexts (i.e., *This is a X*. *That is a Y*.). The isolated words were used as stimuli in the speech perception experiment.

2.3. Tasks

All participants first completed the speech production and then the speech perception task. 8 children and 8 adults were assigned to each difference condition. The mean age of children was 6;5 years (SD = 8.8 mos.) in the smaller difference condition and 6;4 years (SD = 8.3 mos.) in the larger difference condition.

2.3.1. Production

The 40 stimuli in each condition were divided in half so that participants only ever produced 20 unique target words. Half of the 20 words had labial nasals and half had alveolar nasals. More importantly, half of the 20 words had short nasals and the other half had long nasals. The purpose of dividing the stimuli in this way was to ensure that the durational contrast remained sub-phonemic: no participant was ever presented with two word forms that differed only in nasal duration.

Productions were elicited in a picture-naming task to encourage participants to treat target words as monomorphemic forms. Specifically, the tester told a participant that s/he was to learn the new names we had given 20 unique Pokémon creatures. Participants provided 2 repetitions of each target word. During a single trial, the tester placed 2 cards face up on a table, played the first stimulus phrase while pointing to one card (*This is a X*), then indicated to the participant to repeat the name, then played the second stimulus phrase while pointing to the other card (*That is a Y*), and indicated to the participant to repeat the name. The same sequence was followed for the next phrasal stimuli (*Can you say X?* repetition #2 elicited, *Can you say Y?* repetition #2 elicited). X and Y were random and exhaustive pairings of the 20 target words.

Participants' speech was digitally recorded using a Marantz PMD660 and Shure ULXS4 standard wireless receiver and lavaliere microphone, which was attached to a hat or to a headband that participants were given to wear.

2.3.2. Perception

After the speech production task, participants engaged in several unrelated tasks before completing a same/different perception task. The tester played pairings of the altered and unaltered target words and asked the participant to determine whether the paired words were exactly the same or slightly different from one another. The testers monitored participants' attention throughout the task, and continually encouraged participants to listen carefully for any differences that might be present. They also typically broke the task up into several shorter segments for children, who tended to lose focus more quickly than the adults.

2.4. Measurement and analyses

The spoken phrases were displayed in Praat as oscillograms and spectrograms. Acoustic duration of the first vowel and nasal consonant were measured to obtain absolute and relative duration measures. Nasal-vowel boundaries were identified in the waveform by a drop in acoustic energy, a drop in F1, the significant diminishment of F2 and near absence of higher formants. The absolute duration of the preceding vowel and subsequent nasal was calculated based on the segmentation procedure. The relative duration of nasals was calculated using the measure that best distinguishes singleton and geminates in languages with phonemic duration; namely, as the proportion of total consonant duration to preceding vowel duration (C:V1; [6, 9]).

The analyses were based on the mean value from the two repetitions. The effects of condition (smaller v. larger), group (child v. adult), target length (short v. long), and consonant POA (labial v. alveolar) on absolute and relative duration were assessed in a mixed effects model. Participant and word were entered as random factors in the model.

Same/different judgments, which were coded simply as correct (1) or incorrect (0), were analyzed in logistic regression model. Condition, group, and stimulus type (same v. different) were entered as predictors in the model. Participant and word did not contribute to the explanatory power of the model and so were removed.

3. RESULTS

Both children and adults produced very reliable short and long nasal consonants by producing lexically specific differences in the relative timing of V1 and the nasal. Overall, V1 was longer and nasals were shorter when the target nasal was short. Conversely, V1 was shorter and nasals were longer when the target nasal was long. Although absolute and relative durations varied across groups and the difference conditions, these effects did not interact significantly with target length. Unlike in production, the durational difference between short and long targets had an effect on perception. In spite of this effect, participants only ever performed at or below chance levels when judging stimuli in the same/different task. The details of these results are given below.

3.1. Production

Fixed effect tests (Type III) indicated significant effects of target length [$F = 29.24$, $p < .001$], group [$F = 712.70$, $p < .001$], and difference condition [$F = 34.84$, $p < .001$] on V1 duration as well as significant 2- and 3-way interactions between these factors, e.g., [target × group × condition, $F = 5.23$, $p = .023$]. Children produced larger vowel duration differences before short and long consonants than adults, and these differences were especially pronounced in the larger difference condition.

Significant effects of target length [$F = 87.02$, $p < .001$], consonant [$F = 16.63$, $p < .001$], group [$F = 1067.73$, $p < .001$], and difference [$F = 23.83$, $p < .001$] on consonant length were also found. Although there was a significant interaction between group and condition on consonant length, the effect of condition did not interact with target length; only consonant POA and group interacted with target length to affect consonant duration [$F = 3.98$, $p = .046$]. The difference between short and long variants was larger for labials than for nasals, and children produced somewhat larger duration differences between short and long consonants than adults. This latter result is shown in Table 1, which provides the mean absolute duration values for child and adult productions of V1 and the nasal consonant.

Table 1: The mean duration in milliseconds of V1 and the nasal are collapsed across difference conditions and shown for the two groups and target lengths. The standard deviations are given in parentheses.

	Child		Adult	
Measure	short	long	short	long
V1	110.99 (36.03)	96.21 (44.96)	47.00 (8.99)	39.98 (9.78)
Nasal	164.92 (44.46)	205.43 (64.19)	73.84 (1.62)	92.05 (2.43)

Only target length and consonant POA had significant effects on the ratio of C to V1 duration [target, $F = 313.35$, $p < .001$; consonant, $F = 5.19$, $p = .023$]. There was also a significant group by condition interaction [$F = 8.27$, $p = .004$], but the effect of target length was stable across consonants, groups, and conditions.

Figure 1: The relative durations of target long and short consonants are shown for children and adults in the two difference conditions. The difference encoded by the stimuli in each condition is shown for comparison.

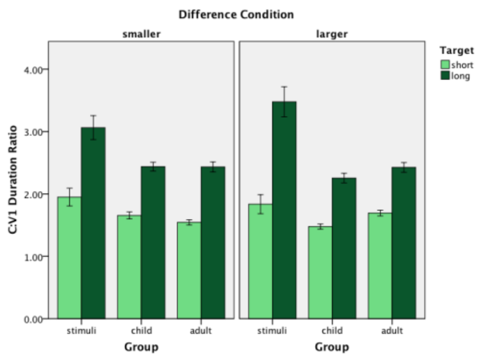

Figure 1 also shows that the duration differences produced by children and adults were substantial, but smaller than that which was encoded by the stimuli. Contrary to our a priori expectation, adults were as good as children at reproducing consonantal duration differences.

3.2. Perception

Children and adults performed poorly in the same/different task in that both groups were biased towards a same response whether or not the paired stimuli were in fact the same (see Table 2). This response bias was affected by condition [β =.56, SE = .26, p = .03], suggesting that listeners were at least attending to the stimuli. Nevertheless, listener performance was still at or below chance when judging stimuli with nasal durations that differed by a factor of 1.9.

Table 2: The percentage of child and adult same and different judgments are shown for the same and different stimulus pairings in the two difference conditions.

		Smaller Difference		Larger Difference	
		Same	diff-erent	Same	diff-erent
Child Judgment	Same	**77.8**	22.2	**71.9**	28.1
	different	75.4	**24.4**	70.9	**29.1**
Adult Judgment	Same	**82.5**	17.5	**86.9**	13.3
	different	69.4	**30.6**	50	**50**

4. GENERAL DISCUSSION

The principal finding of the current study is that speakers are able to reproduce grammatically irrelevant, lexically specific timing patterns that encode a non-contrastive consonantal length distinction. Assuming that accurate reproduction is the first step in word learning and in lexical representation, this result lends credence to the hypothesis that, like true geminates in languages with a phonemic length contrast, the word-internal fake geminates of English may be represented as a single long consonant rather than as a sequence of identical consonants straddling a morpheme boundary.

In spite of their ability to accurately reproduce the non-contrastive length difference, speakers were unable to accurately identify paired stimuli that differed only in nasal length. This result is consistent with other studies on phonetic imitation, particularly those using a shadowing technique, as these seem to suggest that imitation follows from the implicit perception of acoustic detail rather than from changes to abstract perceptual representations.

Finally, the result that children and adults were equally able to reproduce grammatically irrelevant phonetic detail was surprising given that adults are certainly less flexible than children when it comes to learning new sound patterns (viz. persistent non-native accents in adult second language learners). Then again, the imitation task in this study did not stretch the participating adults' sound pattern repertoire: English-speaking adults in this study certainly already produce long and short consonants in their own speech given that the English lexicon has a number of word-internal fake geminates.

5. ACKNOWLEDGMENTS

We are grateful to Nina Rinaldi for getting us going on this project and to the 2010–2011 Speech & Language lab testing team for help with data collection. This work was supported by NIH grant 5R01HD061458.

6. REFERENCES

[1] Browman, C.P., Goldstein, L. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.
[2] Browman, C.P., Goldstein, L. 1992. Articulatory phonology: an overview. *Phonetica* 49, 155-180.
[3] Butterworth, B. 1983. Lexical representation. In Butterworth, B. (ed.), *Language Production.* London: Academic Press, 2, 257-294.
[4] Bybee, J. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
[5] Goldinger, S.D. 1998 Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-279.
[6] Idemaru, K., Guion, S. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of International Phonetic Association* 38, 167-186.
[7] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics.* 39, 132-142.
[8] Oh, G., Redford, M.A. Resubmitted. *The Production and Representation of Fake Geminates in English.*
[9] Pickett, R., Blumstein, S., Burton, M. 1999. Effects of speaking rate on the singleton/geminate contrast in Italian. *Phonetica* 56, 135-157.
[10] Schreuder, R., Baayen, R.H. 1995. Modeling morphological processing. In Feldman, L. (ed.), *Morphological Aspects of Language Processing*. Hillsdale, NJ: Erlbaum, 131-154.
[11] Shockley, K., Sabadini, L., Fowler, C.A. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66, 422-429.