University of Pennsylvania Working Papers in Linguistics

Volume 13 2007
Issue 2 *Selected Papers from NWAV 35*

Article 16

10-1-2007

Are "Covert" /ɹ/ Allophones Really Indistinguishable?

Alina Twist
University of Arizona

Adam Baker
University of Arizona

Jeff Mielke
University of Ottawa

Diana Archangeli
University of Arizona

Are "Covert" /ɹ/ Allophones Really Indistinguishable?

Are ‘covert’ /ɹ/ allophones really indistinguishable?*

Alina Twist, Adam Baker, Jeff Mielke and Diana Archangeli

1 Introduction

The American English phoneme /ɹ/ has several different articulatory realizations, which have been well documented in the literature (Delattre and Freeman, 1968; Tiede et al., 2004; Mielke et al., 2006). A taxonomy of /ɹ/ shapes is shown in Figure 1. Mielke et al. (2006) showed that individual speakers can use more than one articulation, and do so in systematic ways: classic allophony. Mielke et al. also showed, however, that different speakers have different conditioning environments in which each allophone appears.

In many cases of allophonic variation, the different articulations are perceived (consciously or unconsciously) by the listener. The perceptibility of the allophonic pattern allows the pattern to be conventionalized in a speech community (e.g., the allophony of aspirated and unaspirated stops in English). /ɹ/ allophony appears to be different in two respects: there is no conventionalization among speakers of American English, and the allophonic pattern does not appear to be perceptible. Small acoustic differences between different /ɹ/ productions have been documented (cf. Delattre and Freeman 1968), but no research exists to indicate where different /ɹ/ allophones are perceptually distinct. The purpose of the present study is to determine whether speakers can perceive differences between different /ɹ/ allophones.

2 Methods

Two experiments were carried out to test whether or not listeners are able to distinguish between different articulations of /ɹ/, the Segment Experiment and the Whole Word Experiment.

2.1 Stimuli

The data collected by Mielke et al. (2006, Forthcoming) were used as stimuli in the perceptual experiment. Specific data collection procedures are described

*Thanks to Brad Story for providing the glottal source functions. The work reported here was supported in part by the College of Social and Behavioral Sciences at the University of Arizona, and James S. McDonnell Foundation grant #220020045 BBMB. And a special thanks to Mark Liberman for encouraging us to attempt this study.

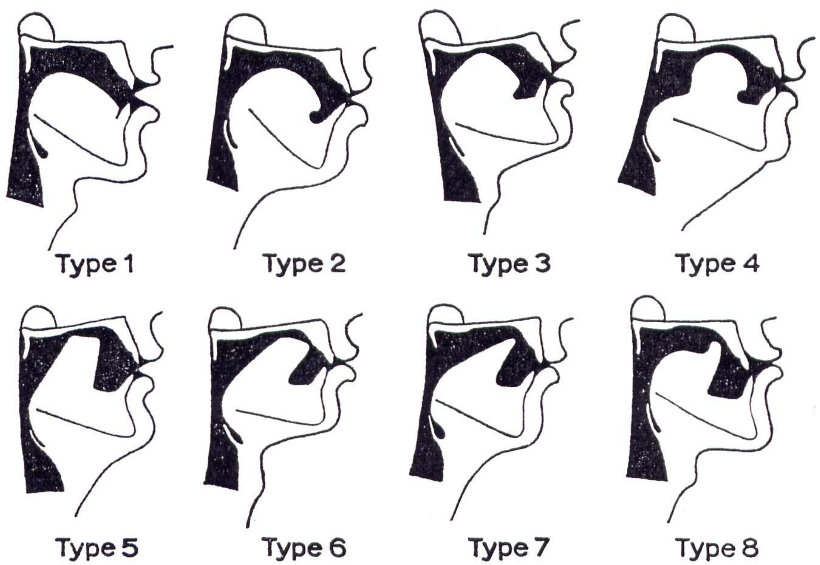

Figure 1: The Delattre and Freeman /ɹ/ taxonomy: The images are flipped so that the face points to the right, not the left, in order to conform to the standard presentation of tongue images in ultrasound language research.

by Mielke et al. (Forthcoming). In brief, the stimuli consisted of monosyllabic words, which represented /ɹ/ in a number of phonetic contexts. Each was produced in the frame, "Please say________again." The data consisted of ultrasound and acoustic recordings of many words containing /ɹ/, both in pre- and postvocalic position. Each /ɹ/ allophone was classified according to the Delattre and Freeman (1968) taxonomy of tongue shapes. Trained phoneticians added further acoustic annotation: identifying the boundaries of each word, and the center of each /ɹ/.

Two sets of stimuli were created. The "word" stimuli were whole words, extracted from the frame sentence. The "segment" stimuli were resynthesized stimuli intended to provide the acoustic information only from a single time point, the center of the /ɹ/ production. The word stimuli preserved phonetic naturalness by presenting the words in context. The segment stimuli contained only non-dynamic acoustic information, with controlled contrastive pitch and length. This eliminated potentially confounding phonetic variables from the stimuli.

To create the segment stimuli, the LPC spectrum was calculated for the center of each /ɹ/. Up to 10 formants were extracted between 0 and 5 kHz, with a window length of 25 ms. Formants collected for each token formed a filter which was applied to a glottal source function. The glottal source used was ~ 0.700 s long. Two source functions were available, one with a peak pitch of ~ 188 Hz, another with a peak pitch of ~ 136 Hz. The higher-pitched source was used for spectra from female speakers, while the lower-pitched source was used for spectra from male speakers.

2.2 Subjects

Participants in the experiment were undergraduate students at the University of Arizona who received either course credit or a candy bar for their participation. Fourteen monolingual native speakers of English and eleven native speakers of Mandarin completed the experiments. Since the Mandarin sound system includes segments that must have a retroflex articulation, speakers of this language might be expected to perceive articulatory differences more accurately than English speakers.

Due to a paucity of Mandarin-speaking subjects, the subject pool also included two native speakers of Mandarin who were post-graduates with extensive linguistic knowledge. All Mandarin speakers were also speakers of English, and several had command of additional languages. Due to equipment malfunction, data from one of the Mandarin speakers in the Whole Word Experiment was lost, leaving ten for analysis.

2.3 Procedure

The experimental protocol was identical in the Word and Segment experiments. Participants heard each item once binaurally over headphones in a sound-attenuated booth. Each item consisted of a series of four stimuli. The endpoints (stimuli 1 and 4) were identical. One of the middle two stimuli differed from the endpoint stimuli in some respect (described below). Subjects were instructed to pick the stimulus (2 or 3) that was unlike the endpoints by pressing the appropriate button on a response box. This design was selected because the sounds to be discriminated are very similar, and this design has been successfully used to investigate pairs of sounds that have minimal acoustic differences, such as incomplete neutralization (Dinnsen and Charles-Luce, 1984; Charles-Luce, 1985; Port and O'Dell, 1985; Slowiaczek and Dinnsen, 1985; Warner et al., 2004). If a response was not given within 1500 ms of the offset of the last stimulus in a series, a non-response was recorded and the next item was presented. Reaction time (RT) and accuracy were recorded by E-Prime (Schneider et al., 2002). Each participant completed both experiments: the Segment Experiment followed by the Whole Word Experiment. The two experiments differed in the type of stimuli presented and the presentation format.

2.3.1 Segment Experiment

The Segment Experiment manipulated stimuli with respect to two factors: word position and articulation of /ɹ/. The items from both conditions were intermingled as part of the same experiment. Items were presented randomly in 10 blocks of 10 items each. Forty-eight items contained midpoint stimuli that differed in *speaker* and *word position* (pre- or post-vocalic). The articulation type (bunched/retroflex) was constant across the stimuli in these items. There were an equal number of items with each /ɹ/ articulation serving as endpoints. Stimuli came from different speakers, and were from different words.

	Endpoint	Match	Mismatch (correct answer)	Endpoint
(1)	Speaker 1	Speaker2	Speaker 3	Speaker 1
	rope	*road*	*core*	*rope*
	bunched	bunched	bunched	bunched

Fifty-two items contained midpoint stimuli that different in /ɹ/ *articulation type*. There were an equal number of items with each articulation of /ɹ/ serving as endpoints. Stimuli for each item were all from the same speaker, and from the same word.

	Endpoint	Match	Mismatch (correct answer)	Endpoint
(2)	Speaker 1	Speaker1	Speaker 1	Speaker 1
	core	*core*	*core*	*core*
	bunched	bunched	retroflexed	bunched

2.3.2 Word Experiment

For the Whole Word Experiment, items were presented randomly in 4 blocks of thirteen items each. Stimuli were analogous to the articulation condition of the Segment Experiment. The speaker and word were constant across all stimuli in each item, but the midpoint stimuli differed in /ɹ/ *articulation type*. There were an equal number of items with each articulation of /ɹ/ serving as endpoints. (The schema is identical to (2) above; it differed only in that the stimuli were whole words.)

2.4 Data Analysis

The data were analyzed by-subjects (F_1) and by-items (F_2), using a series of repeated measures ANOVAs, with language (American English, Mandarin Chinese) and either articulation type (retroflex, bunched) or word position (pre-vocalic, post-vocalic) as the two dependent factors. Articulation type and word position were within-subjects factors and language was a between-subjects factor. In the by-items analyses, articulation type and word position were between-items factors and language was a within-items factor.

As each item elicited one of three responses—correct, incorrect, or no response—three separate ANOVAs were run, using the percentage of each type of response as the independent factor. This approach allowed comparison of differences in the response type means for each condition.

3 Results

The Segment Experiment: Word Position Condition (Figure 2). The percentage of incorrect responses was not affected by either word position ($F_1(1,23) = 1.0$, $p > .05$; $F_2(1,46) = 1.6$, $p > .05$) or language ($F_1(1,23) = 1.0$, $p > .05$; $F_2(1,46) < 1$). There was no interaction of these factors ($F_1(1,23) = 1.63$, $p > .05$; $F_2(1,46) < 1$).

The percentage of non-responses was higher in the postvocalic position than the pre-vocalic position ($F_1(1,23) = 52.6$, $p < .05$; $F_2(1,46) = 18.9$, $p < .05$). Mandarin speakers showed slightly lower percentages of non-responses to all stimuli: This effect was significant in the by-items analysis ($F_2(1,46) = 39.8$, $p < .05$) but not in the by-subjects analysis ($F_1(1,23) = 2.8$, $p > .05$). There was no interaction of the factors, position and language ($F_1(1,23) < 1$; $F_2(1,46) = 2.9$, $p > .05$).

Participants were able to correctly match pre-vocalic stimuli more often than post-vocalic stimuli ($F_1(1,23) = 37.9$, $p < .01$; $F_2(1,46) = 13.3$, $p < .01$). The effect of language was significant in the by-items analysis ($F_2(1,46) = 8.3$, $p < .01$) but not in the by-subjects analysis ($F_1(1,23) < 1$). Likewise, the interaction of the two factors, position and language, was

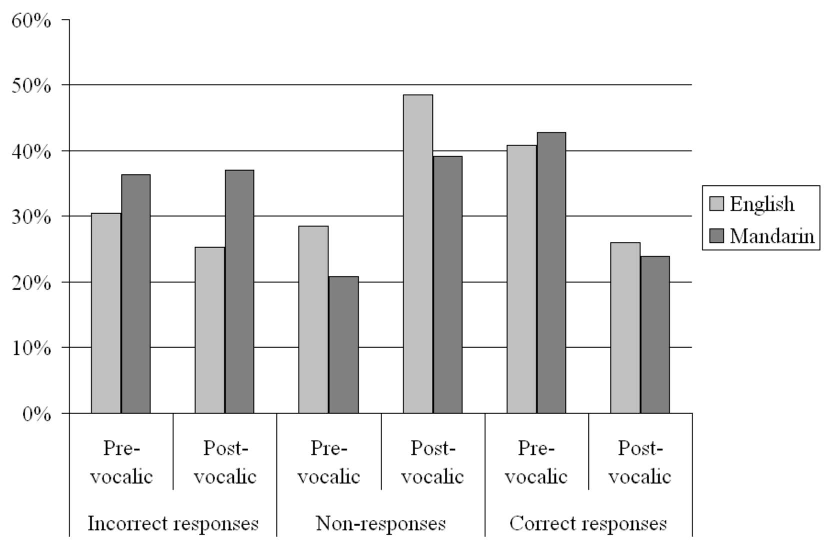

Figure 2: Percentages of response types to pre- and post-vocalic segments.

significant only in the by-items analysis ($F_1(1, 23) < 1$; $F_2(1, 46) = 4.5$, $p < .05$). An examination of the simple effects showed that responses of both English and Mandarin speaking participants were affected by word position (English: $F_2(1, 46) = 13.1$, $p < .05$; Mandarin: $F_2(1, 46) = 5.6$, $p < .05$). Mandarin speakers were better than English speakers at matching pre-vocalic stimuli, but worse at matching the post-vocalic segments.

The Segment Experiment: Articulation Type Condition (Figure 3). The percentage of incorrect responses was not affected by either articulation ($F_1(1, 23) = 4$, $p > .05$; $F_2(1, 50) = 1.2$, $p > .05$) or language ($F_1(1, 23) = 1.6$, $p > .05$; $F_2(1, 50) < 1$). There was no interaction of these factors ($F_1(1, 23) = 1.4$, $p > .05$; $F_2(1, 50) < 1$).

The percentage of non-responses was not significantly affected by articulation ($F_1(1, 23) = 1.2$, $p > .05$; $F_2(1, 50) < 1$). The effect of language was significant in the by-items analysis ($F_2(1, 50) = 15.0$, $p < .01$) but not in the by-subjects analysis ($F_1(1, 23) < 1$). As in response to the word position items, Mandarin speakers gave fewer non-responses. There was no interaction

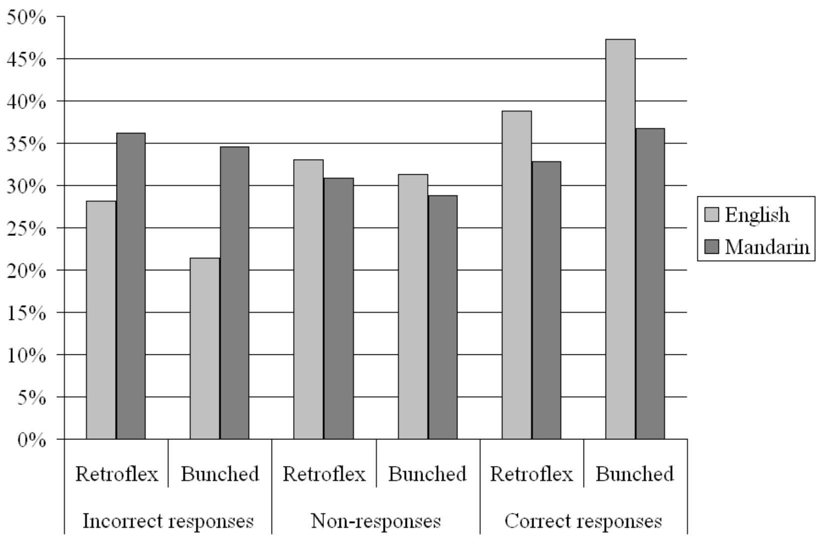

Figure 3: Percentages of response types to bunched and retroflex segments.

of the factors ($F_1(1, 23) < 1$; $F_2(1, 50) < 1$).

For correct responses, articulation had a significant effect only in the by-subjects analysis ($F_1(1, 23) = 6.6$, $p < .05$; $F_2(1, 50) < 1$), with bunched stimuli eliciting more correct responses than retroflex stimuli. Mandarin speakers showed a lower percentage of correct responses overall: This effect was significant in the by-items analysis ($F_2(1, 50) = 50.0$, $p < .01$) but not in the by-subjects analysis ($F_1(1, 23) = 2.3$, $p > .05$). There was no interaction of the factors ($F_1(1, 23) < 1$; $F_2(1, 50) = 1.9$, $p > .05$).

The Whole Word Experiment (Figure 4). The percentage of incorrect responses was not affected by either articulation ($F_1(1, 22) = 2.7$, $p > .05$; $F_2(1, 50) < 1$) or language ($F_1(1, 22) < 1$; $F_2(1, 50) = 3.7$, $p > .05$). There was a significant interaction of the factors only in the by-items analysis ($F_1(1, 22) = 4.3$, $p > .05$; $F_2(1, 50) = 11.6$, $p < .05$). An examination of the simple effects yielded no significant effects.

The percentage of non-responses was not significantly affected by articu-

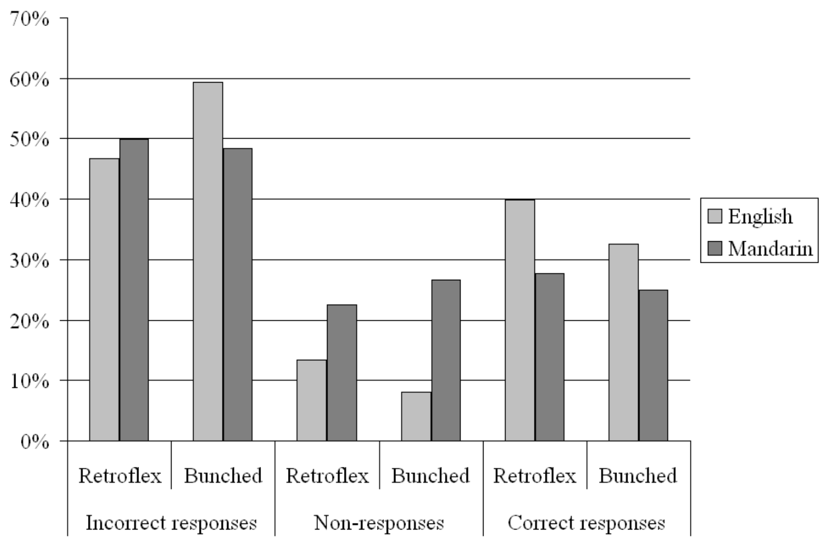

Figure 4: Percentages of response types to bunched and retroflex articulations in whole words.

lation ($F_1(1, 22) < 1$; $F_2(1, 50) < 1$). The effect of language was significant in the by-items analysis ($F_2(1, 50) = 65.3$, $p < .01$) but not in the by-subjects analysis ($F_1(1, 22) = 3.8$, $p > .05$). Unlike their performance in the Segment Experiment, English speakers were less likely to fail to respond to stimuli than Mandarin speakers. There was no interaction of the factors ($F_1(1, 22) = 3.9$, $p > .05$; $F_2(1, 50) = 3.1$, $p > .05$).

For correct responses, retroflex articulations were matched correctly more often than bunched stimuli. This effect was significant only in the by-subjects analysis ($F_1(1, 22) = 4.9$, $p < .05$; $F_2(1, 50) < 1$). As in the Segment Experiment, Mandarin speakers were less likely to respond correctly ($F_1(1, 22) = 5.2$, $p < .05$; $F_2(1, 50) = 30.0$, $p < .01$). There was no interaction of the factors ($F_1(1, 22) = 1.1$, $p > .05$; $F_2(1, 50) = 2.6$, $p > .05$).

4 Discussion

One of the challenges in designing a study to examine the perceptibility of very small acoustic differences is the interpretation of a null result. Are different /ɹ/ allophones actually imperceptible, or was the power of the experiment too low? In this study, the pre- and post-vocalic segment stimuli provided a baseline to determine whether subjects could respond appropriately to the types of stimuli used. The significant differences in response rates to segments in the word position condition indicate that participants were able to discern differences in the stimuli based on acoustic information. This indicates that the methods employed are reasonable for the perception task.

Variation in /ɹ/ production strategy produced less consistent results. Speakers were able to make use of articulatory information to identify mismatched segments at least some of the time, but overall accuracy is fairly poor. Use of whole word stimuli improved the rate of definitive responses (the percent of non-responses to whole words was less than that to segments), but did not seem to improve accuracy.

The results presented here show that there is some effect of different articulations, indicating that speakers are at best weakly aware of variations in /ɹ/ production. This may give speakers license to use various articulation types in their own speech, since the decision comes without social consequence. Even if the listener "gets it right" 35-40% of the time, the rest of the time the articulation of /ɹ/ is either misperceived or not classified at all. Thus, it seems improbable that this perceptual information could contribute to a homogenization of /ɹ/ allophony patterns in a speech community.

References

Charles-Luce, Jan. 1985. Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* 13:309–324.

Delattre, Pierre, and Donald C. Freeman. 1968. A dialect study of American R's by X-ray motion picture. *Language* 44.

Dinnsen, Daniel, and Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12:49–60.

Mielke, Jeff, Adam Baker, and Diana Archangeli. 2006. Forever young: Inaudible /ɹ/ allophony resists conventionalization. Talk given at 80th Meeting of the Linguistic Society of America.

Mielke, Jeff, Adam Baker, and Diana B. Archangeli. Forthcoming. Covert /ɹ/ allophony in English: Variation in a socially uninhibited sound pattern.

In *Variation, Detail, & Representation: LabPhon 10*, ed. Barbara Kühnert. Berlin: Mouton de Gruyter.

Port, Robert, and Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13:455–471.

Schneider, Walter, Amy Eschman, and Anthony Zuccolotto. 2002. *E-Prime Reference Guide*. Psychology Software Tools, Inc., Pittsburgh.

Slowiaczek, Louisa A., and Daniel Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13:325–341.

Tiede, Mark K., Suzanne E. Boyce, Christy K. Holland, and K. Ann Choe. 2004. A new taxonomy of American English /r/ using MRI and ultrasound. Poster presented at the ASA.

Warner, Natasha, Allard Jongman, Joan Sereno, and Rachel Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32:251–274.

Alina Twist, Adam Baker, Diana Archangeli
Department of Linguistics
University of Arizona
Tucson, AZ 85721
{*atwist,tabaker,dba*}*@u.arizona.edu*

Jeff Mielke
Department of Linguistics
PO Box 210028
Ottawa, Ontario
K1N 5N6
Canada
jmielke@uOttawa.ca