Published in final edited form as:

J Mem Lang. 2009 October 1; 61(3): 412–422. doi:10.1016/j.jml.2009.07.003.

Perceptual learning of co-articulation in speech

Cynthia M. Connine and Laura M. Darnieder

Binghamton University

Abstract

Four experiments investigated the novel issue of learning to accommodate the coarticulated nature of speech. Experiment 1 established a co-articulatory mismatch effect for a set of vowel-consonant (VC) syllables (reaction times were faster for co-articulation matching than for mismatching stimuli). A rhyme judgment training task on words (Experiment 2) or VC stimuli (Experiment 3) with mismatching information was followed by a phoneme monitoring task on a set of VC stimuli; training and test stimuli contained physically identical (same condition) or new (different condition) mismatching coarticulatory information (along with a set containing matching coarticulatory information). A third group received no training. A coarticulatory mismatch effect was found without training but not when the same mismatching tokens were used at training and test. Both word (Experiment 2) and syllable (Experiment 3) training stimuli eliminated the mismatch effect; overall reaction times were somewhat slower when the training stimuli were words. Perceptual learning generalized to new tokens only when the acoustic manifestation of the critical co-articulatory information in the training stimuli was sufficiently large (Experiments 3 and 4). The results are discussed in terms of speech processing and perceptual learning in speech perception.

Listeners are confronted with a dizzying array of variation in the language they hear. The variation can take the form of alternative pronunciations of words from both familiar and novel dialects, speech categories that differ from a listener's native language, within talker variability and contextually conditioned segmental variation. The range of variability found in speech would require a system capable of keen sensitivity to detailed information and at the same time able to make rapid adjustments to those details. The present research addresses the novel issue of learning to accommodate the coarticulated nature of speech. Coarticulation is a pervasive aspect of spoken language and refers to a production process in which a speaker adjusts their production of a segment to accommodate the surrounding segments. A consequence of the coarticulated nature of speech is that the acoustic information relevant to a given segment is distributed over adjacent (and sometimes non-adjacent) segments. Listeners must learn what information is relevant for a given segment and integrate the distributed sources of information to arrive at a percept.

The contextual variation in co-articulated segments is due in part to general production constraints generated by individual vocal tract configurations (Fowler & Salzman, 1993) but other work indicates that there is language-specific variation (Strange, Weber, Levy, Shafiro & Nishi, 2002; Maddieson & Emmory, 1985; Sole, 1992). Given that languages and speakers vary in the way in which co-articulatory variation is instantiated in production, a listener must

© 2009 Elsevier Inc. All rights reserved.

Address correspondence to Cynthia Connine, Binghamton University, State University of New York Department of Psychology, Binghamton, N.Y. 13901. Phone: 607.777.2286. Fax: 607.777.4890. connine@binghamton.edu..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

learn the general parameters of co-articulatory variability from the language environment. A classic example of the impact of vowel-consonant coarticulatory dynamics is that the production of the vowel in 'mish' anticipates the 'sh' (/ʃ/) such that it differs from a production of the same vowel in the word 'miss'. It is possible to disrupt the typical course of coarticulatory events by cross splicing — a production of a vowel-consonant sequence ending in 'sh' can be edited such that the 'sh' is replaced with a 's' taken from another vowel-consonant (VC) sequence. The resultant sequence consists of a token in which the vowel formants moving into the final consonant are mismatching (e.g., combining the initial consonant-vowel from 'mish' to the final consonant of 'miss' creates a token of 'miss' with mismatching co-articulatory information). The mismatching information is subcategorical, that is, it does not result in a shift in the perceived category of the final consonant — the 's' in the above example is perceived clearly as such in both the matching and mismatching tokens. Despite the subcategorical nature of the mismatching information, Whalen (1984; see also Streeter & Nigro, 1979) showed that listener judgements are slowed in mismatching relative to matching tokens.

Coarticulation is among the reasons that specification of invariant acoustic properties for phonetic segments has been elusive. The related question of how adult listeners process coarticulated speech has been a focus of research in speech perception for decades (see Nygaard & Pisoni, 1995 for a review). A number of accounts have been proposed concerning how listeners integrate information spread across more than one segment. The gestural approach assumes that speech involves direct perception of vocal tract configurations (Galantucci, Fowler & Turvey, 2006). On this view, listeners parse gestural properties gleaned from the acoustic signal into articulatory dimensions. The feature parsing approach describes a process in which co-articulated cues are grouped and assigned to segments (Gow, 2001, 2003). Central to this view is that anticipatory co-articulatory properties offer the listener a window into upcoming speech and can thus facilitate speech processing. In a related approach, simulations of the TRACE model of speech perception have shown that coarticulatory information in a vowel can begin to activate the appropriate consonant in an anticipatory fashion (Elman & McClelland, 1986). A third approach assumes that coarticulatory information (in the form of abstract features) is mapped directly onto lexical representations (Marslen-Wilson & Warren, 1994 but see Dahan, Magnuson, Tanenhaus & Hogan, 2001; Gaskell, Quinlan, Tamminen & Cleland, 2008). The assumption that lexical representations encode speech variability is shared with episodic models where individual instances of detailed speech events are lexically stored (Goldinger, 1998). Although early episodic models made no claims about abstract units, more recent discussions combine both abstract and episodic properties of lexical representations (Goldinger, 2007, see also Connine, 2004; O'Reilly & Norman, 2002).

The novel question addressed by the present research concerns perceptual learning of co-articulation: how do listeners learn a new vowel-consonant co-articulatory combination? One central theoretical issue is a specification of the information sources that are brought to bear in learning and subsequent effective use of newly learned co-articulatory parameters. A recent approach to the question of how listeners adapt to variability in speech categories suggests some general principles governing perceptual learning. In their seminal paper, Norris, McQueen and Cutler (2003) showed that listeners can learn to adjust their phoneme categories to incorporate an ambiguous phoneme. Listeners were exposed to a speech token containing an ambiguous segment (e.g., ambiguous between /s/ and /f/) in words normally ending in /f/ or to words normally ending in /s/. Subsequent categorization on a /s/ - /f/ speech continuum reflected learning: the group exposed to /f/ bias words in training showed more /f/ responses than listeners exposed to /s/ bias words. The primary conclusions of Norris et al. included 1) the results reflect modification of abstract phoneme categories 2) category modification provides a processing solution for listeners faced with variability in speaker accent. Although Norris et al. characterized speech category tuning as requiring lexical knowledge (see also

Davis, Johnsrude, Hervais-Adelman, Taylor & McGettigan, 2005), other evidence suggests that non-lexical information sources are sufficient. Bertelson, Vroomen and de Gelder (2003) capitalized on the McGurk effect, a phenomenon in which visual presentation of a speaker articulating a syllable presented simultaneously with an auditory spoken syllable results in a percept reflecting the integration of the visual and auditory events (McGurk & McDonald, 1976). In an ingenious experiment that used the McGurk paradigm in a contextual manipulation like that of Norris et al. Bertelson et al. showed that listeners modify speech categories based on visual speech articulatory information. Other recent work has focused on sub-lexical sources of segmental constraint available from phonotactic knowledge. Cutler, McQueen, Butterfield and Norris (2008) used training syllables that promoted interpretation of an ambiguous segment via phonotactic constraints (a phoneme sequence that is legal for one interpretation of the ambiguous segment). Subsequent labeling performance on a speech continuum showed responses consistent with speech category returning.

Thus, perceptual learning of speech categories capitalizes on both lexical and sublexical language specific constraints. Norris et al. argue that processes that support perceptual learning without lexical knowledge must come into play for infants without (or with limited) vocabularies. Acquisition of co-articulatory dynamics would appear to fit well with this characterization as the relevant domain for learning is segmental identity. Listeners integrate co-articulatory information in a vowel with a consonant because they have learned that the coarticulation co-occurs with a particular segment. Given novel co-articulatory dynamics, constraints available at the segmental level can be brought to bear to learn the novel combination and acquisition of new coarticulatory dynamics may occur based solely on VC information. (This assumes that a listener has an existing speech category for the segments.) These considerations do not rule out the possibility that other levels of representation are brought to bear in acquiring novel co-articulatory VC configurations but nevertheless suggest that segmental identity may be a sufficient knowledge source.

The second issue we address is the scope of learning: does learning garnered from exposure to a token set containing new co-articulatory dynamics generalize to a completely new token set for a given speaker? Investigations that have examined transfer of speech category tuning from one token set to another have found both generalization and specificity. Using the paradigm pioneered by Norris et al. (2003) Eisner and McQueen (2005) showed that listeners trained on one talker do not generalize to a new talker. Within talkers, perceptual learning generalizes for speech categories specified by temporal (voicing) but not for spectral (place) dimensions (Kraljic & Samuel, 2005) and across phonemes with similar properties (e.g., modification of a voicing category generalizes to another voicing category defined by a new place of articulation, Kraljic & Samuel, 2006). Kraljic, Brennan and Samuel (2008) showed that variability attributed to an idiolect promotes category modification. In a somewhat different domain, Maye, Aslin and Tanenhaus (2008) showed that listeners can generalize knowledge about a new dialect to a novel set of words. Connine and Pinnow (2006, Pinnow & Connine, submitted) showed that learning an infrequent variant form (schwa vowel deletion) for a set of words facilitates processing of novel words produced as that infrequent variant form. Here, we focus on the nature of the acoustic cues signaling co-articulation in supporting learning and generalization across tokens for a given speaker. Identifying the parameters that control generalization of co-articulation learning to new tokens will contribute to specification of the nature of the learning as well as the constraints that govern the learning process.

In each of the experiments, tokens with novel co-articulatory dynamics were created by cross splicing two productions; in the coarticulatory mismatch condition, a vowel produced before a 'sh' sound was combined with a following 's' sound. In the control coarticulatory match condition, a vowel produced before an 's' sound was combined with the 's' from a second production of the same vowel+'s' sequence. In all stimuli, word/nonword pairs were selected

such that the mismatching co-articulation was embedded in a word carrier but the coarticulation was consistent with a nonword (see Marslen-Wilson & Warren, 1994). Experiment 1 established a co-articulatory mismatch effect (slower phoneme monitoring reaction time to mismatching than to matching stimuli) with a set of VC (vowel-consonant) stimuli. In subsequent experiments, listeners were given experience with mismatch tokens in a training phase and then performed the phoneme monitoring task on the VC stimuli from Experiment 1. In each experiment, mitigation of the mismatch effect is an indication that listeners encoded information about the mismatching co-articulatory information during training and used that knowledge in the phoneme monitoring task. Subsequent experiments investigated whether experience with the mismatch tokens in a lexical and/or segmental context mitigates the mismatch effect. Experiments 3 and 4 focus on the role of the acoustic parameters of the learned and novel stimuli in affording learning and generalization.

Experiment 1

Method

Participants—22 undergraduates at Binghamton University participated in the experiment to fulfill a class requirement. All participants were native speakers of American English and reported no hearing problems.

Materials and design—A set of 16 one syllable words ending in a 's' phoneme was selected the majority of which (12 stimuli or 75%) became a nonword when produced with a 'sh' (see Appendix A, Same training stimuli). The stimuli had an inventory of nine vowels. Each 'sh' and 's' ending token was recorded three times. Mismatching tokens were created by cross-splicing the initial CV of a production ending in 'sh' to the final frication of the production ending in 's'. Matching tokens were created by cross-splicing the initial CV of one word production with the final 's' frication of the second word production. Each token was edited at a zero-crossing and the cross-spliced stimuli had no audible disfluencies when combined. The frication duration and vocalic portions of the match and mismatch tokens did not differ (frication: 257.5ms and 255.8ms; vowel: 186.1 and 188.3ms, match and mismatch respectively, both $F < 1$).

For the phoneme monitoring task, each cross-spliced matching and mismatching word stimulus was edited to remove the initial segments leaving the remainder (final vowel-consonant) of the stimulus intact. An additional set of 32 one syllable words was recorded as fillers that did not contain an 's' sound and then edited to remove the initial segments to create a nonword VC. In some stimuli, the resulting VC combination was itself a word (e.g. PLUS-> US); for these stimuli, a nonword CVC token was created by recording a token with a new onset that created a word when combined with the VC. The new onset was spliced into the onset of the mismatch and match tokens to create a nonword CVC (VC combinations that were letter names were not categorized as a word and therefore not altered). All stimuli were recorded by a female native American English speaker, digitized (44.4 kHz, 16 bit resolution) and stored directly onto a computer for later editing.

Procedure—Participants, run in groups of three or fewer, heard both the match and mismatch token for each stimulus along with the 32 fillers. Stimuli were presented over closed ear headphones in a sound damped room. A different random order of stimulus presentation was used for each participant group. Participants were instructed to listen for the 's' sound on a given trial and to press an appropriately marked key (Target/No Target). Instructions emphasized both speech and accuracy. A practice session was run prior to the experiment.

Results and Discussion

Reaction time was measured from fricative onset and responses longer than 1200 ms were eliminated as outliers (less than 1% of the data). Matching stimuli were responded to 20 ms faster than mismatching stimuli (428 vs 448; $CI \pm 17.7$, $F(1,21) = 5.43$, $Mse = 802$, $p < .029$; $F(1,15) = 7.3$, $Mse = 3383$, $p < .01$; $\min F'(1,36) = 3.11$, $p < .08$). Accuracy rates for the match compared with the mismatch condition (97% vs 95%) were not significantly different ($F(1,21) = 1.66$, $p = .21$; $F(2 < 1)$).

Experiment 1 replicated the previously reported advantage for coarticulatory matching relative to mismatching stimuli for nonwords (Whalen, 1991). The subsequent experiments investigated whether the coarticulatory mismatch effect can be modified by exposure to mismatching stimuli. In Experiment 2, listeners performed rhyme judgments on words with mismatching information in a training phase. The rationale of the rhyme judgement task was to focus listeners on the critical VC portion of the stimulus tokens. In a subsequent test phase, listeners performed a phoneme monitoring task on the set of vowel-consonant (VC) stimuli with matching and mismatching coarticulatory information used in Experiment 1. Table 1 outlines the experiment sequence for the Same and Different Exposure groups. The VC mismatching stimuli heard at test were either physically identical (same exposure group) or a new set of words with physically different stimuli (different exposure group) from those tokens heard during training. A third group received no prior exposure and served as a replication of Experiment 1.

Experiment 2

Method

Participants—60 Binghamton University undergraduate students participated in the experiment (20, 22 and 18 participants in the same, different and control conditions respectively) for course credit. All subjects were native speakers of English and reported no hearing difficulties.

Materials—The set of VC stimuli from Experiment 1 was used for the phoneme monitoring task. The same exposure group training stimuli consisted of the mismatching word tokens created in Experiment 1. The different exposure group training stimuli consisted of a new set of 16 one-syllable words ending in 's'; in the majority of items (14 or 87%), replacing the 's' with 'sh' formed a nonword (see Appendix A). The different exposure training stimuli had an inventory of nine vowels, five of which overlapped with the same exposure training stimuli. The different exposure group stimuli were recorded, digitized and cross-spliced to create a set of mismatching tokens using the procedure from Experiment 1; the new set of stimuli had mismatching coarticulatory information that was physically different from that in the set of stimuli presented during test.

Procedure and design—For the same and different exposure sessions, a set of ten rhyming and ten non-rhyming word pairs were created by pairing stimuli from the set of 16 mismatching tokens. Across the exposure list, each rhyming and non-rhyming pair was presented four times for a total of 80 trials (40 rhyme and 40 non-rhyme trials). Each stimulus was presented eight times, across rhyme and no rhyme trials. The stimuli used in the same exposure group training session consisted of the mismatching word tokens created in Experiment 1. Four additional mismatch fillers were added to permit equal numbers of rhyme/no rhyme trials and an equal number of presentations per stimulus token across trial type.

The stimuli used in the different exposure group training session consisted of the new set of one syllable CVC mismatching tokens. The composition of the different training list was

identical to the same training list: each stimulus was presented eight times, four times as a rhyme pair and four times as a non-rhyming pair. The training rhyme judgement task for the same and different exposure groups proceeded identically: participants were asked to listen carefully to each stimulus pair and to indicate whether the pair rhymed or not by pressing an appropriately marked response box. Accuracy was emphasized. A different randomized presentation was used for each participant group.

In the phoneme monitoring test phase of the experiment, all participants heard the set of matching and mismatching VC syllables used in Experiment 1. Participants were instructed to listen for the 's' sound on each trial and to press an appropriately marked key (Target/No Target). All three groups received identical instructions that emphasized speed and accuracy. A different randomized presentation was used for each participant group. A practice session was run prior to the training and test phases.

Results

Figure 1 shows phoneme monitoring reaction time as a function of training group and match/mismatch condition. Reaction times longer than 1200 ms were eliminated as outliers (less than 1% of the data). As is evident from Figure 1, the mismatch effect remained for the different exposure group but was mitigated for the same exposure group.

Planned comparisons showed a coarticulatory mismatch effect for the group with no training (29 ms CI±17.9, $F(1,17) = 6.55$, $p < .02$; $F(2,15) = 19.9$, $p < .001$; $\text{MinF}'(1,27) = 4.86$, $p < .04$). The size of the mismatch effect was roughly comparable to that found in Experiment 1 (20 ms). A mismatch effect was also found for the different exposure group (24ms CI±15.5, $F(1,21) = 5.66$, $p < .02$; $F(2,15) = 6.2$, $p < .02$; $\text{MinF}'(1,35) = 2.95$, $p = .09$). In contrast, the same exposure group showed no mismatch effect with phoneme monitoring reaction times identical for match and mismatch tokens (0ms CI±17.9, F and $F_2 < 1$). An ANOVA with carrier type (match vs mismatch) and training condition (same, different, control) on reaction time showed an interaction of carrier type and training condition that was marginally significant across subjects ($F(2,57) = 2.66$, $\text{MSe} = 927$, $p < .07$; $F(2,30) = 14.4$, $\text{MSe} = 1005$, $p < .001$; $\text{minF}'(2,75) = 2.24$, $p < .11$). The main effect of match/mismatch was significant ($F(1,57) = 10.4$, $\text{MSe} = 52815$, $p < .002$; $F(2,15) = 16$, $\text{MSe} = 779$, $p < .001$; $\text{MinF}'(1,60) = 6.3$, $p < .015$) and the main effect of group was significant across items ($F(2,57) = 1.8$, $p = .2$; $F(2,30) = 8$, $\text{MSe} = 968$, $p < .001$; $\text{MinF}'(2,78) = 1.46$, $p = .23$).

The control condition had identical accuracy rates for the match and mismatch conditions (98.6%). Accuracy rates for the same exposure group were slightly higher in the match compared to the mismatch condition (99% vs 97.5%) and the different exposure group was slightly less accurate in the mismatch condition (99.5% vs 98%). Both effects were quite small and only the same exposure group approached significance (same: 2% CI±1.25, $F(1,19) = 3.54$, $p < .07$; $F(2,15) = 3.66$, $p < .07$; $\text{MinF}'(1,34) = 1.79$, $p = .19$; different: .5% CI±1.5, $F(1,21) = 1.57$, $p = .20$; $F(2,15) = 2.1$, $p < .16$; $\text{MinF}'(1,36) = .89$, $p = .35$). The small accuracy advantage for the same exposure group suggests that perceptual learning did not completely mitigate the mismatch effect. An ANOVA with carrier type (match vs mismatch) and training condition (same, different, control) on accuracy showed no significant effects.

The results for the same training group suggest that listeners learned to accommodate the new co-articulatory configuration in the mismatching tokens. Exposure to the mismatching information during training dramatically influenced the way in which listeners utilized co-articulatory information rendering it comparable to matching co-articulatory information. There are, however, some indications from Experiment 2 that the learning via a lexical source may not completely neutralize co-articulatory mismatch effects. Although the mismatch effect in reaction time disappeared in the same exposure group, a trend in the data suggest that learning

may have been incomplete. There was a tendency for less accurate responding in the mismatch vs the match condition for the same exposure group. The accuracy effect suggests that the perceptual learning (as evidenced in the reaction time data) is not complete when given experience with words during training.

Unexpectedly, the learning effect for the same exposure group appeared to come at somewhat of a processing cost (although the main effect of learning group was only marginally significant across items and not significant across participants in the omnibus ANOVA): the data trend suggests that the mitigated mismatching effect occurred in the context of slower responses, especially in the match condition. The increased overall reaction times suggests that lexically based learning tended to increase processing time on subsequent encounters of any co-articulation mapping involving /s/. This is a surprising result because lexical context during training seemed to interfere with processing both matching and mismatching stimuli although this may be related to the switch in carrier stimulus between training and test or simply reflect a slower group of participants.

The robust mismatch effect for the different exposure group suggests that learning was specific to the stimuli encountered during training and does not generalize to new instances of coarticulatory mismatching information --- modulation of co-articulatory integration enabled by the training phase is restricted to those VC sequences encountered in training. Lexical information was clearly effective for perceptual learning of co-articulatory information but is lexical context necessary? Experiment 3 addressed this issue directly by presenting VC mismatching tokens during training and test, that is, Experiment 3 was identical to Experiment 2 except that listeners were trained on mismatching VC stimuli (instead of words). In addition, only same and different training groups were used.

Experiment 3

Method

Subjects—50 Binghamton University undergraduate students participated in the experiment (25 participants in each condition) in order to receive course credit. All subjects were native speakers of English and reported no hearing difficulties.

Materials and design—The materials were identical to Experiment 2 except that the stimuli presented during the different training condition were edited to create VC sequences. As in Experiment 1, VC combinations that were words were further edited by splicing a consonant onto the onset of the VC sequence to create a nonword (VC combinations that were letter names were not altered). In the same exposure group, the identical stimuli were presented during the rhyme judgment and phoneme monitoring phase. The composition of the same and different exposure group training lists were identical to Experiment 2. In the different exposure group, the mismatching word tokens used from Experiment 2 were edited to create VC sequences.

Procedure—The procedure was identical to Experiment 2.

Results and Discussion

Figure 2 shows phoneme monitoring reaction time as a function of training group and match/mismatch condition. A similar pattern of results was found as in Experiment 2 -- a mismatch effect was found for the different (24ms CI±10.6; $F(1,24) = 14.5$, $p < .001$; $F(1,15) = 4.4$, $p = .052$; $\text{MinF}'(1,24) = 3.37$, $p = .07$) but not the same exposure group (1ms CI±15.4, F_1 & $F_2 < 1$). An ANOVA with carrier type (match/mismatch) and training condition (same/different) showed a main effect of carrier and an interaction that were marginal across items (carrier: $F(1,48) = 5.3$, $\text{MSe} = 1030$, $F(1,15) = 3.74$, $\text{MSe} = 313$, $p < .07$, $\text{minF}'(1,38) = 2.19$, $p < .14$;

carrier X training interaction: $F(1,48) = 4.17$, $p < .05$, $F(1,15) = 3.68$, $p < .07$, $\min F'(1,43) = 1.95$, $p < .16$).

A post-hoc analysis of reaction time in the same training group condition examined responses separately for early (first half) and late (second half) list positions. Match and mismatch response times were nearly identical across early and late list positions for both match (431ms vs 425ms) and mismatch (423ms vs 424ms) carriers. An ANOVA with early/late list position and carrier type as variables showed no significant effects (all $F < 1$). Accuracy rates in the match and mismatch conditions did not differ (match vs mismatch, respectively; same exposure group: 97.4% vs 96.7%, $.7 \text{ CI} \pm 1.3$, $F_1 \text{ \& } F_2 < 1$; different exposure group: 97.3% vs 97.8%, $.5\%$, $\text{CI} \pm 1.6$, $F_1 \text{ \& } F_2 < 1$). An ANOVA with carrier type and group as variables showed no significant effects (all $F < 1$).

The dramatic mitigation of the mismatch effect found in Experiment 2 using word training stimuli is now demonstrated using VC training stimuli. The same exposure group no longer showed any hint of a processing cost with both the match and mismatch conditions as fast as the different exposure group match condition.

Experiment 3 replicated the data pattern of Experiment 2 - providing training on a set of mismatching stimuli eliminates the mismatch effect but only if physically identical mismatching training/target combinations are presented. In contrast to Experiment 2, the mitigation of the mismatch effect was found without any evidence of a slowdown in reaction times. Similar to perceptual retuning of speech categories, modifications in processing coarticulatory information does not require lexical context.

In both Experiments 2 and 3, exposure to the mismatching tokens did not generalize to a new set of mismatching stimuli created from a new set of words; this finding suggests that the nature of the learning was highly specific and confined to the particular tokens that underlie the adjustment to co-articulation processing. Before elaborating the theoretical relevance of this finding, it is important to establish the acoustic parameters of the learning and test stimuli because previous investigations have reported inconsistencies concerning the impact of coarticulation in fricatives. Investigations using fricatives with co-articulatory mis-matching information have not consistently found a mismatch effect. Marslen-Wilson and Warren (1994) did not find a mismatch effect in lexical decision for cross-spliced match and mismatch word carriers (counterbalanced across participants) using fricatives, but did find a mismatch effect for plosives (see also McQueen, Norris & Cutler, 1999; Warren & Marslen-Wilson, 1987). Whalen (1984, 1991) found a mismatch effect using lexical decision but he used a somewhat different design: matching tokens were presented intact (unspliced) and listeners heard both match and mismatch tokens of a given stimulus. These considerations motivated an analysis of the formant structure of the mismatching tokens (and their matching counterparts) in the training and test stimuli to assess the acoustic properties of the coarticulatory information. Table 2 displays formant frequency values (F_2 and F_3) for a 50-ms window of the vowel prior to the onset of frication for the mismatching tokens used as the different training stimuli. For reference, the formant frequencies for the matching tokens are displayed along with the values for the same training stimuli and their matching counterparts (used in phoneme monitoring). The lower panel shows the difference between the formant values for the matching and mismatching stimuli for F_2 and F_3 .

In both token sets, the difference between F_2 and F_3 is in the direction consistent with a co-articulatory influence in the vowel as a consequence of the fricative (Whalen, 1981). However, as is clear from Table 2, the co-articulation influences in production were smaller in the different training stimuli compared with the same training stimuli. Thus, the physical match of the test and training stimuli was confounded with the relative degree of coarticulatory

mismatching information. In order to determine whether the smaller coarticulatory mismatching information in the test stimuli contributed to the apparent ineffectiveness of transfer learning in Experiment 3, a new experiment was conducted. In Experiment 4, a new set of training stimuli for the different exposure condition were created: the different exposure stimuli differed physically from the test stimuli but contained a comparably large co-articulatory influence in the vowel formants. A second condition, the same exposure condition, replicated that condition from Experiment 3.

Experiment 4

Method

Subjects—A total of 46 Binghamton University undergraduate students participated in the experiment (24 in the different and 22 in the same training groups) for course credit. All subjects were native speakers of English and reported no hearing difficulties.

Materials and design—The same training materials were identical to those used in the same condition in Experiment 3; the mismatching VC sequences presented during the rhyme judgment and phoneme monitoring phase were physically identical. For the different exposure group, a new set of mismatching tokens were created for presentation during the training session using the stimulus pairs from Experiment 3 (see Appendix A). Word and nonword pairs were recorded four times each and tokens were selected that exhibited the requisite coarticulatory effects in F2 and F3 (see Table 3). Two word/nonword pairs that did not show the requisite coarticulatory difference for any of the tokens (mice/mish and place/plash) were replaced by two new pairs (goose/goosh and moose/moosh) (see Appendix A). The different exposure group stimuli had 9 different vowels and had six vowels that overlapped with the same exposure group stimuli. As in Experiment 3, each word and nonword pair were cross spliced and edited to create a VC token.

The composition of the same and different exposure group training lists were identical to Experiment 3 (80 trials with equal number of rhyme and no rhyme trials, and mismatching stimuli were distributed equally across rhyme and no rhyme pairs).

Procedure—The procedure was identical to Experiment 3.

Results and Discussion

Figure 3 shows phoneme monitoring reaction time as a function of training group and match/mismatch condition. Match and mismatch conditions did not show a mismatch effect for the different (8ms CI $\pm$ 11.5, F1 & F2 < 1) or same (6ms CI $\pm$ 13.4; F1 & F2 < 1) exposure conditions. An ANOVA with carrier type (match/mismatch) and training condition (same/different) showed no main effects or interactions (F < 1). Accuracy rates between match and mismatch stimuli were comparable for both groups (same exposure group: 95% vs 96%, 1%, CI $\pm$ 1.7; different exposure group: 97% vs 96%, 1%, CI $\pm$ 1.4; match vs mismatch, respectively; F's < 1). An ANOVA with carrier type and group as variables showed no significant effects across subjects (all F's < 1). The main effect of group approached significance in the item analysis (F2(1,15) = 4.2, MSe = 123, p < .06) with slightly higher accuracy in the different exposure group (97.2% vs 95.8%).

The combined results of Experiment 3 and 4 suggest that new patterns of coarticulation can be learned for co-articulatory events and that the learning can generalize to new tokens. Critically, however, the stimuli must contain a sufficiently large acoustic manifestation of the co-articulatory information to afford learning and subsequent transfer.

General Discussion

Four experiments demonstrated a novel phenomenon in perceptual learning - listeners learn to accommodate new coarticulatory cues when processing speech. Experiment 1 replicated the mismatching effect for a set of VC syllables: in the absence of training listeners are slower to detect a target phoneme in carriers with mismatching compared to carriers with matching coarticulatory information. Experiment 2 pre-exposed listeners to words with mismatching information and established that lexical context can be used to encode the new combination of coarticulatory cues; experience with mismatching information mitigates the coarticulatory mismatch effect in phoneme monitoring. A second group exposed to a different set of word carriers with mismatching information continued to show a mismatch effect of comparable size to that found for listeners with no mismatch training. Experiment 3 used syllables containing mismatching information as training stimuli and demonstrated that syllable context is sufficient to learn the new co-articulatory/segmental pairing. Similar to lexical context, the mitigation of the mismatch effect did not extend from training to new test tokens. Experiment 4 used different training stimuli with a larger acoustic manifestation of the coarticulatory mismatch information than the training stimuli from Experiment 3. The large coarticulatory mismatch training stimuli mitigated the mismatch effect. This suggests that the acoustic properties in the training stimuli were sufficient to trigger learning of the novel coarticulatory parameters in training and this knowledge was brought to bear in processing novel tokens. Experiment 4 also replicated the co-articulation learning effect for syllables given physically identical mismatching tokens used at training and test.

The results support two major conclusions: learning new configurations of coarticulation can be accomplished using segmental information and learning occurs and is expressed (via generalization to novel stimuli) when the novel co-articulation configuration in the learned stimuli contains sufficiently large co-articulatory cues. That listeners encode the coarticulatory dynamics of vowel-consonant sequences is remarkable in light of the fact that the mismatching coarticulatory stimuli violate the typical co-articulatory dynamics. Three specific accounts for processing co-articulatory information in speech were considered in the introduction: the TRACE model, which assumes a facilitory role for coarticulation by partial activation of an upcoming segment, the related feature parsing approach, which assumes that coarticulatory information serves as a preview for and is ascribed to particular segments, and the gestural model, which assumes that gestural properties from the acoustic signal are parsed into articulatory dimensions. These three approaches are discussed in turn.

Both TRACE and the feature parsing views assume that co-articulatory features are used to anticipate upcoming segments (and accounts for the co-articulatory mismatch effect). When encountering a novel mapping between co-articulatory features and segments, and in particular, mappings that violate previously established feature-segment relationships, listeners may simply learn to ignore the novel feature information. In this case, learning would interfere with the normally facilitory aspect of co-articulation - that is, new configurations of co-articulation that are inconsistent with previously established feature/segment relationships inhibit the anticipatory role of features. However, the data do not appear to support this — exposure to mismatching co-articulatory information in training serves to facilitate processing relative to performance on mismatching tokens without training, and is comparable to processing the matching tokens. A second possibility is that learning serves to provide a new mapping between features and segments using the same mechanism underlying the facilitation found for matching tokens. Acquisition of new feature-segment mappings provides the kind of flexibility required to apply the acquired mappings to new tokens. Incorporating new mappings into current co-articulatory knowledge would afford the anticipatory aspect of coarticulation, one of a number of mechanisms that afford fast and efficient speech processing. Early versions of the TRACE model (Elman & McClelland, 1986) explicitly incorporated processing of co-

articulation: a back-propagation algorithm associates the co-articulatory events with a segment and provides a means for co-articulatory events to anticipate upcoming segments. In TRACE, variability in the speech signal introduced by co-articulatory events essentially provides a preview for upcoming segments. More recent developments of the TRACE model use a Hebbian learning algorithm to model a number of current findings in perceptual learning of speech categories (Mirman, McClelland, & Holt, 2006). How might TRACE accommodate perceptual learning of coarticulation? In principle, the learning mechanism that accounts for lexically guided retuning of speech categories could extend to a phoneme-guided re-mapping of feature-to-phoneme associations. Other approaches to speech perception such as the gestural model have no specific mechanism for learning but presumably, learning would involve changes to the way in which the gestural properties are parsed into articulatory dimensions. New gestural properties would reflect new vocal tract configurations and these would be incorporated into parsing of those articulatory dimensions. Generalization of the mismatched co-articulatory properties would require a listener to modify knowledge about what vocal tracts typically do. Such a view would have to account for learning impossible gestural properties, that is, gestural properties that no vocal tract would make; altering gestural knowledge to accommodate an impossible speech event would require a highly adaptable gestural parsing mechanism. Demonstrations such as the McGurk effect show that cross modality impossible events can be accommodated (a visible vocal tract gesture is integrated with an incompatible auditory event) and within a gestural framework, suggests that parsing of gestural properties can be quite malleable. Similarly, it may be the case that the gestures extracted from the mismatching VC sequences during training may fall within the range of events that can alter the way in which gestural properties are parsed.

Transfer of co-articulation learning to a novel set of speech inputs situates the learning process at a pre-lexical level, a level at which coarticulation and segmental information are integrated (see also Hervais-Adelman, Davis, Johnsrude & Carlyon, 2008). This contrasts with a purely episodic account in which the novel co-articulatory dynamics are represented in the context of individual lexical representations. The extent of the violation (as indicated by differences in F2 and F3 formant frequencies for the matching and mismatching tokens) determines the generalizability with more extreme values producing the generalization. This suggests that a large deviation in the typical co-articulation information is taken as a more robust and reliable cue than a less extreme event in signaling a novel feature-to-segment mapping. For perceptual learning, generalization is simply a by-product of encoding the novel mappings between features and segments. On this view, the trigger for learning is that the coarticulation in the exposure stimuli is sufficiently large to alter a listener's expectations (sufficiently deviant from their current state of co-articulation knowledge). An alternative account for the results of Experiments 3 and 4 (the sensitivity to the range of coarticulation for generalization) places the trigger for learning (and generalization) in terms of the overall compatibility of the learned and novel speech. This account places learning and generalization at the intersection of the manifestation of the learned co-articulatory parameters relative to the newly encountered stimuli. The lack of generalization in Experiment 3 is attributed to the overall differences in the co-articulatory mismatch information between the training and test stimuli; this account assumes that learning occurred but was not expressed when new stimuli were encountered because the parameters of learning were not applicable to the novel stimuli. Experiment 4 showed generalization because the set of novel stimuli had co-articulatory parameters that were consistent with what was learned during training. A prerequisite for learning is that the co-articulatory events are encoded; in our experiments, this is expressed behaviorally by the co-articulatory mismatch effect in phoneme monitoring. Did the mismatching stimuli used in training in Experiment 3 contain a sufficiently large manifestation of co-articulatory information to be learned? In order to determine whether a mismatch effect was in force with the different mismatch stimuli used in Experiment 3, a phoneme monitoring experiment was conducted that included the mismatch training stimuli (and the match counterparts) from

Experiment 3 ($N = 8$); a small (5ms) but not significant mismatch effect was found (423ms vs 428ms, match and mismatch, respectively; $F < 1$). This suggests that the co-articulatory mismatching information was insufficiently large to alter established routines for integrating co-articulatory and segmental properties. Thus, there are limits to perceptual learning - listeners will adjust speech processing only when there is sufficient reason based on the speech input (Bedford, 1995).

As with modification of speech categories, learning new configurations of coarticulatory information did not require lexical input. However, lexical representations may play a more subtle role in acquiring co-articulatory mappings. Co-articulation has been found to play a functional role in the activation and inhibition of lexical competitors (Dahan et al. 2001) and co-articulatory features tend to become lexicalized (Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, 1999; see also Klatt, 1986). Transient activation of a lexical competitor from the to-be-acquired novel co-articulation may complicate acquisition of the new feature-segment mappings. The presence/absence of a lexical competitor consistent with the /ʃ/ coarticulatory information was not explicitly manipulated but the majority of the stimuli did not have a lexical competitor. The impact of a lexical competitor that is consistent with the mismatching co-articulatory information on perceptual learning remains an issue for further research.

It remains to be determined how co-articulatory learning is accomplished in the context of other kinds of novel speech (language, dialect, segment or speaker mappings) but critically, testing for generalization will require novel speech that is sufficiently novel relative to previously encountered speech to alter the process of feature integration. The distinction between learning co-articulatory information and re-tuning phoneme categories is clear but how does co-articulatory learning compare with other types of learning in speech? Perceptual learning has been demonstrated for time compressed speech of one's own (Dupoux & Green, 1997) and a foreign language (Pallier, Sebastian-Galles, Dupoux, Christophe & Mehler, 1998). Less than one minute of exposure to foreign accented speech reduces errors in a recognition task for both a familiar (Spanish) and unfamiliar (Chinese) foreign accent (Clarke & Garrett, 2004). Gow and Im (2004) showed that the process of grouping co-articulatory cues can occur for speakers processing a foreign (unknown) language. A modified vowel category can be learned and generalizes (Maye, Aslin & Tanenhaus, 2008; see also Floccia, Goslin, Girard, & Konopczynski, 2006; Evans & Iverson, 2004; Dahan, Drucker & Scarborough, 2008). In the domain of intra-speaker variability, a large number of studies have shown that familiarity with a speaker's voice increases processing efficiency (Nygaard & Pisoni, 1998, Nygaard, Summers & Pisoni, 1994). Processing under challenging listening conditions can require modification of mappings involved in speech perception but a characterization of novel speech events in terms of the relevant dimension to be learned (novel speech categories and/or co-articulation mappings) is not entirely straightforward. Novel dialects may involve re-tuned speech categories (as in Maye et al.) but novel dialects, speakers and languages can also involve new coarticulation mappings. There are a number of possibilities for how such learning might occur but it is likely that learning novel category structures are informed by coarticulatory events and vice versa. Acquiring a new language or dialect that contains categories tuned differently from those in the native language may require synchronization between acquisition of the speech category and acquisition of co-articulatory mappings. Given the important role that co-articulatory cues play in efficient and fast processing, the difficulty listeners have in processing a new dialect or accent may be because the co-articulatory cues typically available for efficient and fast processing (both within and between words) have not yet been acquired. The co-requisite process of acquiring new speech categories will further hamper this process of integrating co-articulatory cues into a percept. For infants, the speech categories that appear to be available at birth may provide a boot strapping mechanism for mapping the co-articulatory events on to the existing category structures.

Speech perception, perhaps more so than any other aspect of language processing, must have evolved to adjust with great speed and efficiency to variations in the input. Coarticulation is a source of information that listeners can exploit in on-line processing of spoken language in a way that can speed up analysis of upcoming information. The rapid acquisition and generalization of new coarticulatory configurations suggests that listeners are highly attuned to the acoustic landscape of novel speech events.

Acknowledgments

The research was supported by NIH grant R01DC02134. Additional support was provided by the Center for Cognitive and Psycholinguistic Science. The authors thank Albrecht Inhoff and Eleni Pinnow for their helpful comments.

Appendix A

Same training stimuli. Note: The final VC of each mismatching token was presented for the phoneme monitoring task in Experiments 1-4.

boss

close

crass

cross

dress

grass

gross

guess

kiss

lease

less

loose

miss

plus

press

twice

Different training stimuli (Experiments 2-4). Note: Stimuli mice and place (*) were substituted with goose and moose in Experiment 4.

bless

bus

choice

face

glass

house

juice

mess

niece

mouse

nice

pass

rice

voice

mice*

place*

Experiment 4 additional stimuli

goose

moose

References

- Bedford F. Constraints on perceptual learning: Objects and dimensions. *Cognition* 1995;54:253–297. [PubMed: 7720360]
- Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science* 2003;14:592–597. [PubMed: 14629691]
- Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* 2004;116:3647–3658. [PubMed: 15658715]
- Connine CM. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychological Bulletin and Review* 2004;11(6):1084–1089.
- Connine CM, Pinnow E. Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review* 2006;23:235–245.
- Cutler, A.; McQueen, KM.; Butterfield, S.; Norris, D. Prelexically-driven perceptual retuning of phoneme boundaries; *Proceedings of INTERSPEECH 2008*; Brisbane, Australia. 2008;
- Dahan D, Drucker SJ, Scarborough RA. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 2008;108:710–718. [PubMed: 18653175]
- Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes* 2001;16:507–534.

- Davis MH, Johnsruide IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General* 2005;134:222–241. [PubMed: 15869347]
- Dupoux E, Green K. Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23(3):914–927. [PubMed: 9180050]
- Elman, J.; McClelland, J. Exploiting lawful variability in the speech wave. In: Perkell, J.; Klatt, D., editors. *Invariance and Variability in Speech Processes*. MIT Press; 1986. p. 360–380.
- Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 2005;67:224–238. [PubMed: 15971687]
- Evans BG, Iverson P. Vowel normalization for accent: An investigating of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America* 2004;115:352–361. [PubMed: 14759027]
- Floccia C, Goslin J, Girard F, Konopczynski G. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception & Performance* 2006;32:1276–1293. [PubMed: 17002537]
- Fowler CA, Saltzman E. Coordination and coarticulation in speech production. *Language and Speech* 1993;36:171–195. [PubMed: 8277807]
- Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 2006;13:361–377. [PubMed: 17048719]
- Gaskell MG, Quinlan PT, Tamminen JT, Cleland AA. The nature of phoneme representation in spoken word recognition. *Journal of Experimental Psychology: General*. In press.
- Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 1998;105:251–279. [PubMed: 9577239]
- Goldinger, S. A complementary-systems approach to abstract and episodic speech perception; *Proceedings of the 16th International Congress of Phonetic Sciences*; Saarbrücken. 2007;
- Gow DW. Assimilation and Anticipation in Continuous Spoken Word Recognition. *Journal of Memory and Language* 2001;45:133–159.
- Gow DW. Feature Parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 2003;65:575–590. [PubMed: 12812280]
- Gow DW, Im A. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 2004;51:279–296.
- Hervais-Adelman A, Davis MH, Johnsruide IS, Carlyon RP. Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance* 2008;34:450–474.
- Klatt, DH. Problem of variability in speech recognition and in models of speech perception. In: Perkell, J.; Klatt, D., editors. *Invariance and Variability in Speech Processes*. MIT Press; 1986. p. 360–380.
- Kraljic T, Brennan S, Samuel AG. Accommodating variation: Dialects, idiolects and speech processing. *Cognition* 2008;107(1):51–81.
- Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 2005;51:141–178. [PubMed: 16095588]
- Kraljic T, Samuel AG. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 2006;13:262–268. [PubMed: 16892992]
- Maddieson I, Emmorey K. Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic differences and coarticulation. *Phonetica* 1985;42:163–174. [PubMed: 3842771]
- Marslen-Wilson W, Warren P. Levels of perceptual representation and process in lexical access. *Psychological Review* 1994;101:653–675. [PubMed: 7984710]
- Maye J, Aslin R, Tanenhaus MT. The wecked wetch of the wast: Lexical adaptation of a novel accent. *Cognitive Science*. 2008
- McQueen JM, Norris D, Cutler A. Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance* 1999;25:1363–1389.

- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 1976;264(5588):746–748. [PubMed: 1012311]
- Mirman D, McClelland JL, Holt L. An interactive Hebbian account of lexically guided tuning of speech. *Psychonomic Bulletin and Review* 2006;13:958–965. [PubMed: 17484419]
- Norris D, McQueen J, Cutler A. Perceptual learning in speech. *Cognitive Psychology* 2003;47:204–238. [PubMed: 12948518]
- Nygaard, LC.; Pisoni, DB. Speech Perception: New Directions in research and theory. In: Miller, JL.; Eimas, PD., editors. *Speech Language and Communication*. Academic Press; 1995.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception and Psychophysics* 1998;60(3):355–376. [PubMed: 9599989]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker contingent process. *Psychological Science* 1994;5(1):42–46.
- O'Reilly RC, Norman KA. Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *TRENDS in Cognitive Sciences* 2002;6:505–510. [PubMed: 12475710]
- Pallier C, Sebastian-Galles N, Dupoux E, Cristophe A, Mehler J. Perceptual adjustments to time-compressed speech: a cross-linguistic study. *Memory and Cognition* 1998;26(4):844–851.
- Pinnow, E.; Connine, CM. Learning phonological variants in spoken word recognition. Submitted
- Sole M-J. Phonetic and phonological processes: The case of nasalization. *Language and Speech* 1992;35:29–43.
- Strange W, Weber A, Levy E, Shapiro V, Nishi K. Within and across language acoustic variability of vowels spoken in different phonetic and prosodic contexts: American English, North German, and Parisian French. *Journal of the Acoustical Society of America* 2002;112:2384.
- Streeter LA, Nigro GN. The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America* 1979;65:1533–1541. [PubMed: 489823]
- Warren P, Marslen-Wilson W. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics* 1987;41(3):262–75. [PubMed: 3575084]
- Whalen DH. Effects of vocalic formant transitions and vowel quality on the English [s]-[.] boundary. *Journal of the Acoustical Society of America* 1981;69(10):275–282. [PubMed: 7217525]
- Whalen DH. Subcategorical phonetic mismatches slow phonetic judgements. *Perception & Psychophysics* 1984;35:49–64. [PubMed: 6709474]
- Whalen DH. Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics* 1991;50:351–360. [PubMed: 1758766]

Figure 1. Mean phoneme monitoring reaction time as a function of training condition (control, same, different) and stimulus type (match, mismatch). Experiment 2.

Figure 2. Mean phoneme monitoring reaction time as a function of training condition (same, different) and stimulus type (match, mismatch). Experiment 3.

Figure 3. Mean phoneme monitoring reaction time as a function of training condition (same vs different) and stimulus type (match, mismatch). Experiment 4.

Table 1

Example stimuli for the Rhyme/No Rhyme judgement task and the Phoneme Monitoring task for the Same Exposure and Different Exposure Groups in Experiment 2. Physically identical tokens that are repeated across the Rhyme/No Rhyme judgement and phoneme monitoring tasks are indicated by underline

		Rhyme Training Task Coarticulation		Phoneme Monitoring Test Coarticulation	
		Mismatch	Mismatch	Match	Mismatch
Same Exposure Group					
Rhyme trial		<u>PRESS</u>	<u>DRESS</u>	ESS	<u>ESS</u>
		<u>MISS</u>	<u>KISS</u>	ISS	<u>ISS</u>
No-rhyme trial		<u>LOOSE</u>	<u>CLOSE</u>	OOSE	<u>OOSE</u>
		<u>GROSS</u>	<u>CROSS</u>	OSS	<u>OSS</u>
Different Exposure Group					
Rhyme trial		BLESS	MESS	ESS	ESS
		CHOICE	VOICE	ISS	ISS
No-rhyme trial		JUICE	NICE	OOSE	OOSE
		HOUSE	FACE	OSS	OSS

Note: Onsets were excised for the rhyme training tokens for Experiments 3 and 4.

Table 2

Formant frequency for mismatching (/ʏ/) and matching (/s/) vowels used in Experiments 2 and 3 as a function of formant (F2/F3) and training stimulus type (same or different). Frequency values (Hz) for positions immediately (0ms), 25ms (-25ms) and 50ms (-50) prior to onset of frication are shown. Frequency differences between mismatching and matching vowels are indicated in the bottom panel

		SAME				DIFFERENT			
		-50	-25	0		-50	-25	0	
F2	/ʏ/	1800	1810	1888	1953	1972	2025		
	/s/	1756	1736	1798	1885	1916	1964		
F3	/ʏ/	2781	2754	2807	2849	2847	2871		
	/s/	2825	2835	2927	2885	2871	2903		

		/ʏ/ - /s/ Frequency Difference							
		SAME				DIFFERENT			
		-50	-25	0		-50	-25	0	
F2		44	73	90	68	56	61		
F3		-44	-80	-120	-36	-24	-32		

Table 3

Average formant frequency (Hz) for F2 and F3 for mismatching vowels (/ʃ/) used in the different training condition along with the matching (/s/) vowel counterpart in Experiment 4. Values for three positions immediately (0ms), 25ms prior (-25ms) and 50ms (-50) prior to onset of frication are shown. Frequency differences between mismatching and matching vowels are indicated in the bottom panel

		DIFFERENT		
		-50	-25	0
F2	/ʃ/	1919	1994	2132
	/s/	1926	1930	2004
F3	/ʃ/	2826	2821	2885
	/s/	2926	2918	2989

		DIFFERENT		
		/ʃ/ - /s/ Frequency Difference		
		-50	-25	0
F2		-7	64	128
F3		-100	-97	-104