

## Reassignment of the flap allophone in rapid dialect adaptation

Running title: Allophone reassignment

Katy Carlson<sup>a,\*</sup>, James German<sup>b</sup>, and Janet Pierrehumbert<sup>b</sup>

<sup>a</sup>Morehead State University Department of English, Foreign Languages and Philosophy, 103 Combs, 150 University Boulevard, Morehead, KY 40351, USA

<sup>b</sup>Northwestern University Department of Linguistics, 2016 Sheridan Road, Evanston, IL 60208-4090, USA

\*Corresponding author. Tel.: +1-606-783-2782; fax: +1-606-783-9112

E-mail addresses: [k.carlson@morehead-st.edu](mailto:k.carlson@morehead-st.edu), [j-german@northwestern.edu](mailto:j-german@northwestern.edu),  
[jbp@northwestern.edu](mailto:jbp@northwestern.edu)

## Abstract

In an experiment spanning a week, American English speakers imitated a Glaswegian (Scottish) English speaker. The target sounds were /t/ and /r/, as the Glaswegian speaker aspirated word-medial /t/ but pronounced /r/ as a flap. This experiment therefore explored whether speakers could learn to reassign a sound they already produce (the flap) to a new phoneme and to new phonetic contexts. Speakers appeared to learn systematically, as they could generalize to words which they had never heard the Glaswegian pronounce. There was a mix of categorical learning, with the allophone simply switching to a new use, and parametric approximations of the “new” sound. The phonetic context was clearly important, as flaps were produced less successfully when word-initial. And although there was variety in success rates, most speakers learned to produce a flap for /r/ at least some of the time and retained this learning over a week’s time. These effects are most easily explained in a hybrid of neo-generative and exemplar models of speech perception and production.

Keywords: flap, allophone, dialect, rhotic, learning, generalization, lexicon

## **Acknowledgments**

This work was supported by a grant from the James S. McDonnell Foundation. We are extremely grateful to our Glaswegian speaker, Alistair J. McGowan, for the many hours he spent helping us to create the speech stimuli for the experiment. We'd also like to thank Matt Bauer for supplying illustrative ultrasound pictures of the relevant phone contrasts for use in the figures.

## 1. Introduction

Given the extreme precociousness of first language (L1) phonological acquisition, and the difficulty of acquiring a fully native second language (L2) accent as an adult learner, it was previously assumed by some that a critical period for phonological acquisition made adult dialect adaptation almost impossible. A number of recent sociophonetic studies have targeted this important issue, and their results provide evidence of dialect adaptation in adult speakers under natural conditions. Munro et al. (1999) found that Canadian health care professionals who had taken up residency in Birmingham, Alabama partially acquired an American accent, as evidenced by perceptual ratings of extemporaneous speech samples. Harrington et al. (2000a, 2000b), drawing on 40 years of recorded Christmas broadcasts of Queen Elizabeth II, used objective acoustic measures to show that Her Majesty's pronunciation has been modified in some respects towards the rising Southern British (Estuary) patterns. A post-hoc study by Sankoff (2004) of a longitudinal set of recordings made for the British documentary series *Seven Up* also found dialect adaptation by two speakers who moved from Yorkshire and Liverpool to southern England and elsewhere as adults. Such sociophonetic studies provide provocative evidence of late plasticity in the phonological and phonetic system. However, sociophonetic studies cannot by their nature provide diagnostic evidence about the cognitive architecture responsible for this adaptation. Being uncontrolled, extemporaneous speech samples simply do not support the kind of in-depth statistical analysis that would be possible for data from controlled experiments.

In this paper, we report the results of a controlled experiment on dialect learning. In a 20-minute (48-sentence) training session with no feedback, 24 Northwestern undergraduates undertook to imitate Glaswegian English to the best of their ability. The targets of interest were

/t/ and /r/, though the subjects were not told this. For /t/, we were interested in the allophone that appears intervocally under falling stress (as in the word *pretty*). This is a flap in American English, but is aspirated in the speech sample of our Glaswegian speaker that was used to construct the experimental materials. The Glaswegian /r/, in contrast, is a rhotic tap in all positions, very similar to the medial flap in American English *butter*. The training phase was immediately followed by a test for generalization to novel lexical items. Subjects also returned to the lab a week later, where they were tested for retention of the Glaswegian pattern. The retention testing had three components: the original training set, the original generalization set, and a new generalization set.

Our goal in designing this experiment was to address three related issues. These issues are suggested by prior work on second language learning and on learning of individual speaker traits. Specifically:

- 1) Lexical versus systematic learning: To what extent do subjects learn general phonological or phonetic patterns, which transfer from words in the training materials to new words?
- 2) Categorical versus parametric learning: To what extent do learners succeed by exploiting phone categories which they already know from their L1 (or D1, native dialect)? To what extent do they succeed by forming new categories over the parametric phonetic space?
- 3) Positional constraints: If existing categories can be reused at all, are they confined to their original D1 context? Or can they be reused in a different context? In short, can D1 positional variants, or allophones, be promoted to independent contrastive status in D2?

The literature on L2 learning has emphasized systematic phonological and phonetic learning. Dialect learning (D2 learning) resembles L2 learning in that it involves competition between the L1 or D1 phonological system and the novel system. Prior results in L2 learning

indicate that a speaker's success in learning an L2 speech segment depends on the exact nature of its relationship to segments in the L1 inventory. Research on this issue, going back several decades, has been dominated by examination of the L2 phoneme inventory compared to that of L1. Two of the most widely known models, Best's Perceptual Assimilation Model (Best et al., 2001) and Flege's Speech Learning Model (1995), share a number of key assumptions about how the L1 phoneme inventory comes into play during L2 exposure. If an L2 phoneme is phonetically equivalent to an L1 phoneme, it will be processed using the L1 code, hence it will be successfully perceived and produced. If it is phonetically similar to an L1 phoneme, but not equivalent, strong interference is expected: the L2 sound is perceptually assimilated to the L1 phoneme, and hence it is very difficult for the learner to improve beyond his/her initial rapid but partial success with the phoneme. If it is extremely distinct from all L1 phonemes (as Zulu clicks are for English speakers), then there is much less interference, and the phoneme is a candidate for the kind of parametric learning involved in new category formation. Success by adults in such learning would be indicative of phonetic plasticity, and failure would be indicative of a critical period for phonetics.

As discussed in Strange (1995), the experimental paradigms employed in work on acquisition of L2 phoneme inventories generally explore only a particular positional variant of the target phonemes (for example: a novel consonant contrast in stressed, word-initial position). It is unclear whether the cognitive units involved are phonemes in the classical sense (which retain their identity across variations in context), or less abstract, allophonic, units. She reviews results on the acquisition of the /r/-/l/ distinction by Japanese learners of English (Mochizuki, 1981; Logan et al., 1991). Though this contrast is generally difficult for Japanese learners, it is much more difficult in some contexts than others, indicating that allophonic units are probably

the relevant level of description. Whalen et al. (1997) studied the [p] vs. [p<sup>h</sup>] allophones of English and found that speakers could imitate these sub-phonemic differences even if they could not reliably distinguish them. Polka (1991) explored in detail the ability of English listeners to perceive the dental-retroflex distinction in Hindi as a function of the laryngeal features of the target phoneme pair. Hindi has a four-way distinction amongst plain voiced, plain unvoiced, breathy-voiced, and aspirated-unvoiced stops, which combines with a dental-retroflex place distinction to yield a total of eight distinct obstruent coronal stops. Because English /t/ and /d/ have dentalized contextual variants, different consonant pairs in the set of eight Hindi consonants were predicted to have different degrees of support from the English allophonic system. Polka did find that perception of the dental-retroflex contrast was significantly above chance for three of the four consonant pairs, suggesting that the English phonetic system in some way supports perception of this contrast. Her specific hypotheses about the Hindi pairs were not supported, possibly because she relied on a literature review of English allophony which did not, apparently, encompass the stop allophones of [θ] and [ð] observed in emphatic productions and in many dialects. The study we report here supports a more secure interpretation of the allophonic variation, because we obtained baseline recordings in order to estimate the D1 allophonic statistics on a subject-by-subject basis.

Two recent studies by psycholinguists use artificial language learning tasks to explore the malleability of the coding system in perception. Maye et al. (submitted) used a speech synthesizer to create an artificial dialect of English with categorical lowering of target vowels. For example, the substitution of /ε/ into the word *witch* yields the form *wetch*, a nonword in the base dialect. Maye et al. found that subjects exposed to the novel dialect significantly increased their endorsement of modified forms as words in a lexical decision task. The effect generalized

to new words. It did not generalize to other vowel substitutions than were made in the training phase. Since endorsement of unmodified words was not reduced, the results point to an architecture in which the relation of the phonological code to the lexicon can be systematically augmented in response to novel speech patterns. Parametric learning is not implicated in Maye et al.'s experiment, though, because the stimulus materials were created by categorical substitution of phonemes. Peperkamp and Dupoux (in press) also investigate the malleability of the system at the categorical level. They use an artificial language learning paradigm to explore a categorical feature neutralization for certain consonants. In their materials, voicing was contextually predictable for stops but not for fricatives, or vice versa in the counterbalancing condition. Their series of experiments also manipulated the degree of semantic support for learning the phonological patterns of the language. Subjects were tested using a picture-pointing task. When word-learning was semantically supported, learning of the phonological constraint was very efficient and generalized to new words.

Results such as those of Maye et al. and Peperkamp and Dupoux suggest a neo-generative architecture, similar to Figure 1. The production system, following the broad lines of Levelt (1980), is involved in retrieving word forms from the lexicon, assembling the phonological code for the word forms in their phrasal context, and computing the phonetic implementation of the assembled phonological representation. The perception side is portrayed as analogous in the figure; the acoustic phonetic signal is phonologically parsed, and the phonological parse serves to access the lexicon. Systematic effects of the type that Maye et al. and Peperkamp and Dupoux have demonstrated do not involve any modification of the units in the coding level; the adaptation resides in the relationship of these units to the lexicon, with Maye et al.'s experiment involving the subjects' existing lexica, and Peperkamp and Dupoux's

experiment involving novel lexical items in a putatively novel language. Comparing Maye et al.'s and Peperkamp and Dupoux's results to Polka's tends to support Strange's suggestion that the relevant units at the coding level may be positional variants of phonemes rather than phonemes themselves. On this interpretation, Peperkamp and Dupoux's subjects succeeded in learning equivalence classes of these positional variants.

-----Insert Figure 1 about here-----

A neo-generative architecture, such as Figure 1, readily captures categorical, across-the-board effects. That is, if the phonological coding level is systematically modified in production, then this modification will be reflected in the phonetic realizations of all words. No words—whether in the training set or not, whether frequent or rare—will have any privileged status with respect to the new coding pattern. If the coding system is modified in perception, it will likewise affect all words equally. The architecture is also consistent with certain word-by-word effects. We all know that some words have more than one pronunciation. If the subjects in an experiment simply memorized the new pronunciations for the words in the training materials as categorical alternatives, then the model would capture this fact by listing multiple word-forms for these specific words in the lexicon. A mixed situation, in which words used in training materials show an effect most reliably, but the effect also generalizes to new forms, can also be described by assuming that subjects update their coding systems through statistical generalizations over known examples, as suggested in Pierrehumbert (2003). If we assume Bayesian updating (e.g., modifying prior probabilities in the light of new statistical evidence), then the grammar statistics will lag the lexical statistics until the learning is complete. This is exactly what Maye et al. (submitted) and Peperkamp and Dupoux (in press) report. Given the brief training and variable outcomes in these studies, the claim that the experiments ended before the learning was complete

is plainly justified.

A different architecture has been proposed by many researchers working on voice recognition and social identity, such as Goldinger (1998) and Johnson (in press). Dialect recognition is similar to voice recognition, because an idiolect can be viewed as a one-person dialect. Recognizing a dialect means recognizing something about the speaker's social identity, just as recognizing gender or sexual orientation does. Learning to produce a dialect means learning to project a particular social identity, and modern sociophonetic theory indeed explores dialect learning in the context of social identity construction (Mendoza-Denton et al., 2003). Experiments on speech processing in relation to individual speakers and social identity have revealed some surprising interactions, which are problematic for the neo-generative architecture in its basic form. Such effects include: Shifts of category boundaries as a function of gender and gender typicality (Johnson, in press); effects of speaker identity on word recall (Goldinger, 1996; Goldinger et al., 1991; Palmeri et al., 1993; *inter alia*); effects of speaker identity on novel word recognition (Nygaard et al., 1994); and unconscious imitation effects, which are more significant for low frequency words than for high frequency words (Goldinger, 1998).

The discovery of such effects has fueled the rise of exemplar-based models of speech perception. These models presuppose that experiences of speech are stored in memory in very considerable detail. Each memory can be indexed in multiple ways; for example, a memory of the utterance [berbi] can be indexed as an example of the word *baby*, as an example of my mom's speech, and as an example of a female voice. In the simplest exemplar models (e.g., Hintzman's (1986) MINERVA, Johnson's (1997) XMOD), phonological structure emerges epiphenomenally from the similarity space defined by the remembered experiences. Exemplar models readily capture interactions between social variables and lexical access. They also

capture findings that novel patterns are more robustly applied to the exact materials used in the training than to novel materials.

The simplest exemplar models, such as MINERVA and XMOD, encounter difficulties in explaining the extreme reliability of lexical access under changes in speech rate or prosodic position. If lexical access is attempted from the parametric representations of entire words, the alignment of the speech signal with the stored representations can be problematic. Reduction of a segment or syllable early in the word can induce misalignment of the entire remainder of the word with the stored representations. This can lead to a poor match even if phonologically aligning word subparts in the optimal way would have yielded a very good match. This problem is noticeable in calculations using XMOD presented in Baker (2004). Clearly, this same problem is compounded when word recognition in connected speech is considered. A further issue for exemplar models is the mechanism for speech production. Pierrehumbert (2001) takes as a point of departure the idea that production targets are picked by random selection of the exemplar space for the word. Goldinger (1998), taking a position reminiscent of direct realists, proposes that the combined effect of all exemplars activated by a lexical choice creates a production plan. But both positions are regrettably vague about how novel words can be produced. Productions of novel words do not average the properties of all similar real words. If they did, [bɪɑg] would average *bog*, *blog*, *frog*, *broad*, *brought*, and so forth, leading to a hybridized sonorant in the onset and a hybridized obstruent in final position. Instead, productions of [bɪɑg] begin with the [bɪ] of *brought* or *broad*, and end as in *frog*.

Such issues have led to the development of hybridized models, with some such models already reviewed in Goldinger (1998). Pierrehumbert (2002) adopts the neo-generative claim that production of all words involves programming a categorical phonological representation,

and that executing this plan is the only way to produce speech. (See Levelt, 1980). This means that lexical representations of individual words include both a phonological parse, needed to compute alignment and sequencing in speech processing, and a phonetic trace, needed to capture the individual speaker and sociostylistic effects which led to the original rise of exemplar models. She suggests that specific words or social situations can influence phonetic realizations by biasing the selection of phonetic exemplars used as realization targets for phonological plans. Since these biases are within-category, they are expected to yield secondary effects.<sup>1</sup> This hybrid model, then, supports four different mechanisms altogether for imitating a new accent:

- 1) Learning new phonetic categories. Such learning is predicted to be possible only after high levels of exposure and practice, and is subject to strong interference from existing categories.
- 2) Learning situationally-appropriate biases within existing categories.
- 3) Learning alternative pronunciations for known words, encoded using existing categories.
- 4) Learning generalizations about these alternative pronunciations, encoded as generalizations about phonological representation.

The existence of this model, and others like it, means that the comparison between exemplar models and neo-generative models is not dichotomous. Rather, one can define a theoretical spectrum of models, ranging from pure exemplar models (such as the Hintzman's (1986) MINERVA model, which guided Goldinger 1998) to neo-generative models such as Levelt (1980). In MINERVA, phonology is epiphenomenal from phonetic learning and from clustering and similarity at the phonetic level. In neo-generative models, the phonological grammar is primary, and due to critical period effects, phonetic learning may be impossible in adults. This is in contrast to the predictions of exemplar models, which predict phonetic learning after significant exposure to be possible at any age (Hall & Boomershine, 2006). The issues

which guided our experimental design allow us to locate the cognitive system with respect to this spectrum of models. Insofar as we find fast, systematic, categorical learning, we need key features of the neo-generative models. If positional constraints are important, this provides evidence for the units needed at the phonological coding level. In contrast, pure exemplar models, with their epiphenomenal phonology deriving from a less abstract description of speech, do not provide for the same degree of plasticity in the phonological encoding. But key features of exemplar models can capture detailed phonetic learning, as well as lexical gang effects in generalization.

## 2. Background

### *2.1. American English flapping and /r/*

Most sources agree that post-stress intervocalic /t/ is quite frequently realized as a flap in American English, at least in conversational speech. Zue and Laferriere (1979) conducted a production study on /t/ and /d/ in various environments which found flapping of /t/ occurring in 99% of post-stress intervocalic cases. Fisher and Hirsh (1976) found more variable results in their production study, with subjects ranging from 36% to 97% flap production. One suspects that some of these subjects were speaking in a more careful and less conversational manner than others, or possibly spoke different dialects. Perhaps the most definitive results were presented by Patterson and Connine (2001), who examined two corpora of conversational speech. Their overall finding was that 94% of post-stress intervocalic /t/'s were flapped, though they also found lower levels of flapping in low-frequency and morphologically complex words. This work thus suggests that the flap [ɾ] is the most common reflex of /t/ in this particular environment, having been found both in experimental settings and in natural conversation. There is no

tendency for the flap to occur as an allophone of /r/ in American English, either intervocalically or elsewhere.

The normal realization of /r/ in American English is a voiced alveolar approximant [ɹ] which can vary by speaker in the degree of retroflexion. It tends to lower the third formant of any vowel which precedes it (Ladefoged, 1993).

## *2.2. Glaswegian English and our speaker*

The dialect which American English speakers were to adapt to was Glaswegian Standard English. The Lowland Scots language variety, also spoken in and around Glasgow, differs from American Standard English in lexicon and grammar as well as pronunciation (Chirrey, 1999). However, our experiment only involved Glaswegian pronunciation. Descriptions of Scottish accents in Edinburgh (Chirrey, 1999) and Glasgow (Stuart-Smith, 1999) mention a range of realizations of the rhotic consonant, including [r] and [ɹ], but our speaker used a flap or tap articulation exclusively. The phoneme /t/ was primarily realized with aspiration by our speaker. In initial recordings, a glottal stop also occurred in medial positions, but this was infrequent and seemed to be in free variation with the aspirated /t/. For the experimental recordings, the very few utterances with a glottal stop for /t/ were discarded and only aspirated productions were used.

There are a number of other differences between Glaswegian and American English in addition to the /r/ and /t/ realizations. Many of the vowels differ, for example, with the characteristic American [æ] replaced by a vowel much further back in the mouth. Additionally, Glaswegian English has different prosodic patterns, some of which were imitated by subjects (German, 2004).

The Glaswegian English speaker for this experiment was a native Glaswegian who had lived in Scotland up until he came to America for graduate study. At the time of this experiment, he was engaged in graduate study in Chicago, where he had lived for 2 years. He had a strong Scottish personal identity, including active involvement in Scottish political and cultural groups. His retention of his native dialect was very marked and when speaking fast, he could be quite unintelligible to American ears. The experimental sentences were given to the speaker on a script in a pseudo-randomized order similar to that in which they were heard. He read each sentence aloud at a moderate rate of speech.

### 3. Methods

#### 3.1. Stimuli

There were four conditions of sounds under investigation, with /t/ and /r/ each appearing in strong (word-initial) and weak (word-medial) positions (Fougeron & Keating, 1997; Pierrehumbert & Talkin, 1992). A total of 192 sentences were created, 48 of each type, with the constraint that no allophone of /r/ or /t/ appeared anywhere except in the target word of the appropriate condition. The target words were always sentence final, so as to be both prosodically prominent and easy to remember for participants. Sample items are shown in (1):

- (1) /t/, strong position: He gave away his only token.  
/t/, weak position: The damp wind made him all sweaty.  
/r/, strong position: All the family's belongings lay beneath the rubble.  
/r/, weak position: The boy swallowed mud because he was curious.

The items were grouped into four blocks, each containing twelve items of each type for a total of 48 per block. Items within each block were pseudo-randomized such that no two consecutive sentences were from the same condition. The four blocks of items were rotated through the usage

conditions in a counterbalanced order to avoid extraneous lexical effects. All of the blocks of items were recorded by the Scottish English speaker and put on CD. An additional group of three 12-item blocks was created and recorded. These blocks contained only non-target items, so the sentences had no /r/ or /t/ allophones in them at all (e.g., *A display of the dig can be seen in the lobby*). All of the items in the experiment are listed in Appendices 1-2.

### 3.2. Procedure

Each participant produced all four blocks of items in some usage condition, and the blocks were counterbalanced to appear equally often in each condition. One block was produced as a Baseline. Before a participant heard any Glaswegian English recordings, they were asked to read a block of items in a normal conversational style. This set served as an example of the participant's normal production of /r/ and /t/. Another set of items was used as the Training block. The participant would listen to the Scottish English speaker producing each sentence in this block while following along on the written script, stop the CD, and then imitate the sentence into the microphone. This Training session was repeated once with the same procedure immediately after its first iteration. The two Training sessions together took under 20 minutes to complete, on average. The final task in the first week was the Generalization1 block. The participant was given the script of this set of items, which they had not previously seen nor heard the Glaswegian English speaker produce, and asked to continue imitating the accent.

Each participant returned to the lab a week after their first session. In this session, three blocks of items were recorded: the Training set again, the Generalization1 set again, and a new Generalization2 block. The order of these three block types was counterbalanced so that each was recorded first, second, or third by an equal number of participants. Before each of the blocks, participants refreshed their memory of the accent using one of the Non-target blocks of

items. They would listen to the Scottish English speaker and imitate him, as in the first Training sessions, except that these 12-item blocks did not contain any /t/ or /r/ sounds. Participants did not hear the speaker produce any of the target items from the Training or Generalization blocks during Week 2. The full set of recordings is summarized in Table 1.

-----Insert Table 1 about here-----

The recordings were made using a Shure SM 81 microphone connected through an Ariel Proport, an Earthworks preamp, and an Apogee PSX 100 A/D into a Macintosh G4 computer running ProTools. The microphone and participants were located inside a sound-attenuated recording booth. The recordings were saved as 22050 Hz mono sound files and burned onto CDs.

### *3.3. Participants*

The participants in this study were 24 undergraduate students at Northwestern University enrolled in lower-division linguistics classes. They received course credit for their participation. Data from bilingual and non-native participants was excluded from analysis, as was that of students who were unable to return for the second session. The students ranged in age from 19 to 38, and their average age was 22. All but three of the participants had studied at least one foreign language, and twelve of them had studied Spanish. Eight of the participants were male.

### *3.4. Data Analysis*

Each of the recorded sound files was analyzed by one of the first two authors. Coders listened to the final word of each sentence while examining the waveform and spectrogram using Praat. The initial coding transcribed the phonetic result, while a second level of analysis categorized the utterance as a successful imitation of Scottish English (a recruited allophone), an American English result (showing non-adaptation), or another sound (a phonetic innovation). The observed

productions of /t/ included aspirated (the target pronunciation), unaspirated, flapped, and a very few other sounds. In cases where the speaker was clearly aiming at a different target sound, as in the fairly common mispronunciation of the initial segment of *Thames* as [θ], the data were excluded as being irrelevant to the question of accent imitation. There were few ambiguous segments in the pronunciations of /t/, since the stop closure of even an unaspirated [t] is quite distinct from a flap visually and auditorily. Some participants produced shorter periods of aspiration during the imitation portions of the experiment than in their baseline, especially for intervocalic /t/, though this result has not yet been systematically analyzed. The productions of /r/ included flaps (the target), rolled and trilled [r]s, glottalized /r/s, and American [ɹ]s. Some participants produced an interesting retroflex palato-alveolar fricative [ʒ], and occasionally an [l]- or [w]-like sound. Judgments of flap success were based both on acoustic criteria and the visual characteristics of a flap which distinguish it from an American [ɹ].

#### 4. Results

The results of this experiment are shown first in Figure 2 and Figure 3, which display the percentage of completely successful outcomes for /r/ and /t/ respectively. This shows how often participants produced the Glaswegian target allophones.

-----Insert Figures 2-3 about here-----

Unsurprisingly, participants had no trouble producing aspirated /t/ in the strong position (initial) condition, as this is the preferred allophone in American English as well as Glaswegian English. Their performance was very close to perfect for /t/ in strong positions in all usage conditions. They were almost as good at producing aspirated /t/ in weak positions, where American English usually has a flap but allows unaspirated or aspirated /t/ in the formal register. In fact, two of the

participants produced 25-30% unaspirated or aspirated /t/s in the Baseline weak position /t/ condition, though all participants fluently produced flaps in this position.

The data from /t/ in strong position served as a control showing the accuracy of participants in producing the expected allophones of unchanged targets. The data with /t/ in weak position illustrated whether speakers could suppress the usual allophone of a phoneme (the flap) in favor of an allophone typically found in a different environment (aspirated /t/), and their level of success here was also high. The small difference between the strong and weak position /t/ conditions proved to be fully significant in planned contrasts only by items. The within-subjects strong position /t/ vs. weak position /t/ contrasts were significant after the Bonferroni correction for only the Generalization1 block ( $t(23) = 3.88, p < .01$ ) and the Generalization2 block ( $t(23) = 3.15, p < .05$ ). The between-items contrasts were significant after Bonferroni correction for most blocks with sufficient t-initial variance to test: Training2 ( $t(94) = 3.14, p < .05$ ), Generalization1 ( $t(94) = 5.16, p < .01$ ), Week2-First ( $t(94) = 3.66, p < .01$ ), and Week2-Second ( $t(94) = 5.80, p < .01$ ). In any case, average performance for both /t/ conditions was above 90%, so the difference between them was very slight.

The flapped /r/s were clearly more difficult for the participants, with the average percentages for /r/ in strong position all below 50% and below 70% for /r/ in weak position. There was variation in performance, too, with some individual subjects who achieved 100% performance on /r/ conditions as early as the Training2 block, and others whose success rate never rose above 25% flaps in any /r/ condition. This may be related to the participants' innate ability to mimic, which has been shown to affect the degree of foreign accent (Flege et al., 1999; Piske et al., 2001; Purcell & Suter, 1980; Thompson, 1991). The rest of the statistics below will focus on the /r/ conditions as being of most interest and variability.

The two first-week Training blocks were examined to see whether participants improved their imitation with additional exposure to the Scottish speaker. An ANOVA on /r/ performance in strong and weak positions in Training1 vs. Training2 was conducted; the factor of r-position was within-subjects but between-items, while the block factor was within-subjects and within-items. There was a significant main effect of r-position, with better performance for /r/ in weak position than in strong position ( $F(1, 23) = 28.48, p < .001$ ;  $F(1, 94) = 51.74, p < .001$ ). There was also a significant main effect of Training block, such that participants' performance improved in Training2 relative to Training1 ( $F(1, 23) = 13.94, p = .001$ ;  $F(1, 94) = 20.79, p < .001$ ). The interaction between these factors was marginal ( $p$ 's above .05), with a trend towards lesser improvement for /r/ in strong position. In general, then, participants used the extra practice in imitation and exposure to the speaker to improve their rate of flapping for /r/, though performance on words with /r/ in weak position was better than for words with /r/ in strong position from the very start.

In order to look for effects of time and practice on lexical items, an ANOVA containing Training2, Generalization1, Training3, and Generalization1R was conducted on the strong position vs. weak position /r/ items. This analysis examined performance on two particular blocks of items in two successive weeks, with the blocks differing in whether participants had heard the Scottish English speaker produce them or not. There was a significant effect of r-position, with /r/ in weak position always better than /r/ in strong position ( $F(1, 23) = 25.64, p < .001$ ;  $F(1, 94) = 74.51, p < .001$ ). There was a significant to marginal main effect of time, with a small performance drop between the first and second week's sessions ( $F(1, 23) = 7.13, p = .014$ ;  $F(1, 94) = 41.09, p < .001$ ). This was easily significant by items but weak by subjects, perhaps because several of the 24 participants actually showed improvement in their Week 2

recordings. There was a significant main effect of block, since the Training block showed higher levels of success than the Generalization1 block in both weeks ( $F(1, 23) = 11.79, p < .005$ ;  $F(1, 94) = 10.43, p < .005$ ). Still, the difference between the block types was small, showing that speakers had acquired a systematic pattern that they could generalize beyond the particular lexical items they had been trained on. Finally, there was a significant interaction between r-position and time, since the data for /r/ in weak position showed a larger performance difference between weeks than the strong position /r/ data did ( $F(1, 23) = 13.34, p < .005$ ;  $F(1, 94) = 17.51, p < .001$ ). No other interactions approached significance.

While the time effect was significant, not only did some subjects show better performance at the second week, but the overall effect of time was small. In short, speakers were able to retain the dialect adaptations they had learned during the first week's training. Figures 4 and 5 show the correlations between speakers' Generalization1 performance (during Week 1) and the average of their Week 2 performance, for items with /r/ in strong and weak positions. For /r/ in strong position, the correlation was  $R^2 = .7529$ ; for /r/ in weak position,  $R^2 = .8021$ . These correlations show that the similar success rates in Weeks 1 and 2 are not artifacts of averaging, but do reflect individual speakers' long-term retention of the dialect.

-----Insert Figures 4-5 about here-----

Because of counterbalancing, different subjects encountered different block types in week two in different orders. Figures 6-7 summarize the data by block type and block order respectively.

-----Insert Figures 6-7 about here-----

By conditions, the Training3 blocks appear to show better performance than the repeated Generalization1R items, but the Generalization2 results were intermediate where one would

expect worse performance if practice with lexical items were a primary factor. On the other hand, performance increased through the First through Third blocks in the second week for /r/ conditions. An ANOVA on the Training2, Generalization1R, and Generalization2 data by blocks showed the expected main effect of r-position ( $F(1, 23) = 14.38, p < .005$ ;  $F(1, 94) = 65.54, p < .001$ ) but marginal to nonsignificant effects of condition ( $p$ 's  $> .08$ ) and no significant interaction. The ANOVA on the data by order (First, Second, and Third) showed both a significant main effect of r-position ( $F(1, 23) = 14.38, p < .005$ ;  $F(1, 94) = 67.02, p < .001$ ) and a significant main effect of order ( $F(2, 46) = 6.44, p < .005$ ;  $F(2, 188) = 15.45, p < .001$ ). There were no interactions, but the test for linearity (polynomial order 1) was also significant ( $F(1, 23) = 9.51, p < .01$ ;  $F(1, 94) = 24.62, p < .001$ ). It seems, then, that participants improved their dialect adaptation during the course of the Week2 recording session, with the last condition produced being their strongest.

All of these statistics have showed a strong effect of the strong vs. weak position of /r/. There is one caveat to these results, however. Although all of the target words for /r/ in weak positions had /r/ in an intervocalic position after a stressed syllable, there was a minority of strong position /r/ words (15 out of 48) in which /r/ followed a consonant, as the preceding word was consonant-final. Since the usual environment for flap in American English is intervocalic, the small proportion of items with non-intervocalic /r/ in strong position creates a potential confound for the effect of strong versus weak position alone. Figure 8 shows the percentages of success for the intervocalic vs. non-intervocalic items with /r/ in strong position, compared to the items with /r/ in weak position.

-----Insert Figure 8 about here-----

Indeed, the intervocalic set of the strong position /r/ items showed higher percentages of flapping than the non-intervocalic items in all of the usage conditions except the Baseline. The difference between the intervocalic and non-intervocalic items was significant in within-subjects and between-items ANOVAs including the Training2, Generalization1, and Week 2 blocks ( $F(1, 23) = 10.22, p < .005$ ;  $F(1, 46) = 9.66, p < .005$ ). The difference between blocks was also significant. However, similar ANOVAs on the items with /r/ in weak position vs. only the intervocalic strong position /r/ items showed that there was still a fully significant main effect of prosodic position ( $F(1, 23) = 19.38, p < .001$ ;  $F(1, 79) = 61.34, p < .001$ ). Thus the advantage for /r/ in weak position persists even when compared to the subset of items with /r/ in strong position which were maximally similar.

While the reasonable success of subjects in producing medial flapped r's suggests that they were able to map this existing allophone to a new phoneme, the remapping did not always guarantee the target results. In particular, in addition to completely non-adapted American responses, most subjects also produced phonetic innovations. Figures 9-10 show the percentage of flap recruitment and of additional innovated (non-American) outcomes for both /r/ conditions (as the success in the /t/ conditions meant that there were very few innovative or non-adapted responses).

-----Insert Figures 9-10 about here-----

The proportion of partially unsuccessful (innovated) trials was highest for the condition in which the Scottish target was most difficult to achieve, namely the /r/s in strong position, and lowest for the /t/ conditions. The condition with /r/ in weak position had an intermediate level of innovation. Looking at the rate of phonetic innovations by subjects, we found that most subjects who produced phonetic innovations also produced successful flaps, rather than particular

speakers producing only these non-target sounds and not the Glaswegian targets. The intervocalic vs. non-intervocalic strong position items were also examined. The rate of phonetic innovations for the non-intervocalic strong position /r/s equaled or exceeded the rate of phonetic innovations for the intervocalic strong position items in most cases. This is reasonable, since the non-intervocalic strong items were the ones with the lowest rate of successful flap recruitment. These figures do not show another interesting phonetic outcome found in the non-intervocalic strong position /r/ data: the epenthesis of a vowel. While this was not a common result, six of the speakers used this strategy at least once during the experiment in order to place the /r/ in an intervocalic context.

The phonetic innovation data suggests that subjects intended to produce something other than an American /r/ in the partially successful trials for /r/ in strong position, but failed to fully execute either the phonetic coding or the articulatory maneuvers necessary for a flap. This highlights a possible connection between the innovation data and the intervocalic/non-intervocalic data in Figure 8, suggesting that American English speakers are simply only practiced at articulating a flap sound in a medial and intervocalic position, and thus have difficulty producing it in other environments. The result is that they produce a number of intermediate sounds, including a retroflex palato-alveolar fricative, a uvular sound, and a trill (the latter being especially common among subjects who had taken Spanish). This explanation is supported by data from Munson (2001) on error rates in the production of phonological patterns as a function of frequency. He found that infrequent sequences of sounds were more likely to be produced slowly or incorrectly than frequent sequences, even though all of the sequences did occur in grammatical English words. It would not be surprising, then, for speakers to have difficulty producing the flap in a word-initial context (especially a post-consonantal one).

## 5. Discussion

The dominant effect in our study was that the majority of speakers were able to exploit their knowledge of the [r] variant of /t/ in D1 for realizing /r/ in a new dialect. In addition, speakers were able to extend the [t<sup>h</sup>] positional variant of /t/ to weak prosodic positions. This effect was systematic to the extent that the ability generalized strongly to words not in the training materials. In that sense, our main finding represents the production counterpart to the perception results of Maye et al. (submitted) and Peperkamp and Dupoux (in press), and like the results of those studies, it is consistent with the predictions of a neo-generative model.

Our results also show effects that are not predicted by a neo-generative model. For one thing, subjects performed slightly better on Training and previously Generalized items than on new items after an interval of one week. Lexical effects of this type suggest that word representations include information about contextually and socially indexed variants in addition to the abstract phonological coding needed for the systematic effects typified by our main finding. At a minimum, these representations must include information about allophonic variants. In the case of our study, the repetition of the word and sentence context on practiced items provided a retrieval bias toward the targeted (D2) allophone above and beyond the bias contributed by social indexing based on the identity of the Glaswegian speaker. At most, the representations include even finer phonetic detail as suggested by the results of Johnson (in press) and Goldinger (1998). The fact that both systematic and lexical effects were present in our results simultaneously is therefore best captured by a hybrid model in which both mechanisms act together.

The L2 learning models in Best et al. (2001) and Flege (1995) predict that speakers will form a new perceptual category only when the phonetic target in L2 cannot be identified with a

preexisting L1 target. In addition to recruiting [ɾ] for the realization of /r/, subjects in our study realized /r/ with sounds not found in American English. Several subjects, for example, produced /r/ with some variant of a retroflex alveolar fricative [ʒ]. Other sounds that were observed include variants of [ɻ], [r], and [ʒ]. Figure 11 compares the waveforms of subjects' productions of [r], [ʒ], and [ɻ] in the same lexical context (*new road*). Presumably, the exploration of the phonetic space exhibited by subjects in our study is a consequence of the same kind of category innovation discussed by Best et al. and Flege. That is, some subjects may have been unable or unwilling to perceptually identify the Glaswegian tap with the American [r] due to the slight phonetic dissimilarity of those sounds. As a result, they encoded exemplars of /r/ with a novel allophonic category. When those subjects targeted /r/ for production, they accessed the new perceptual category, since it was strongly indexed with both the phoneme /r/ and the social identity of the speaker. However, since they had no experience producing this new category, they were forced to explore the phonetic space in an attempt to approximate it. Neogenerative models do not prohibit new category formation, but they also do not provide for it. A hybrid model, on the other hand, does provide for category formation, since in that model the emergence of categories follows from the representation of experience. The fact that subjects were only partially successful at innovation (i.e., they produced [ʒ] instead of [r]) is a reflection of the fact that these attempts were incipient. After all, hybrid models predict successful imitation to be possible only after much higher levels of exposure and practice.

-----Insert Figure 11 about here-----

The fact that subjects in our study were able to use [ɾ] in word-initial, prosodically strong positions, where it does not occur in their native dialect, indicates that sub-phonemic variants are the relevant level of coding. To see why, consider that if the adaptation to D2 involved only

modification of the relation of the phonological code (phonemes) to the lexicon, then we would expect recruited phonemes to obey the same prosodic conditioning that they do in D1. Thus, if /t/ were being substituted across the board for /r/, /r/ would be realized as [r] in weak position and [t<sup>h</sup>] in strong position. Further support for sub-phonemic encoding derives from the fact that speakers were able to replace an allophone from one prosodic context (the flap) with an allophone from another prosodic context ([t<sup>h</sup>]) within the same phoneme category (/t/). This is not predicted by a model that only permits realignments at the level of phonemic encoding. Both results converge with Polka's (1991) suggestion that listeners are able to use their implicit knowledge of English allophony as a resource for processing non-native phonemic contrasts.

The fact that speakers in our study substituted [r] for [ɹ] and [t<sup>h</sup>] for [r], but did not realize /r/ as [t<sup>h</sup>], is further evidence that an intermediate level of representation is involved. The logical structure of this claim can be clarified by returning to one of the early works of generative phonology. Kiparsky (1968) distinguished “feeding orders” from “counterfeeding orders.” For the substitutions involved in our study, a feeding order would be represented as in (2).

- (2) [1]    ɹ → r  
       [2]    r → t<sup>h</sup>

Ordering the substitutions in this way creates the potential for remapping /r/ to [t<sup>h</sup>] via [ɹ] and then [r], since the output of [1] is equivalent to the input of [2]. The interaction that was actually observed is captured by the opposite, or “counterfeeding,” order in (3).

- (3) [2]    r → t<sup>h</sup>  
       [1]    ɹ → r

Representing a counterfeeding effect of this type in any model requires a level of representation that is intermediate to the lexical representation and the production target. The total picture is therefore illustrated by Figure 12.

-----Insert Figure 12 about here-----

Though our results indicate that subjects were able to reassign [ɹ] to /r/ in prosodically weak as well as prosodically strong contexts, their performance was better in weak positions where [ɹ] typically occurs in D1. This may be partly due to a difference in articulatory difficulty between the two contexts, since the airflow required to produce [ɹ] was reduced in strong positions, where /r/ was preceded by an unstressed vowel or consonant, relative to weak positions, where /r/ was intervocalic and preceded by a primary stressed vowel. Many languages including Glaswegian English, however, use rhotic flaps or taps in such initial positions without difficulty, so this effect is most likely a secondary one. The discrepancy is perhaps best accounted for in terms of the speakers' experience. Motor patterns, such as the articulation of a flap, are learned in context and learned more robustly with a large number of examples. Speakers of American English have experience producing [ɹ] in medial, falling stress, intervocalic contexts across a large number of words, whereas their experience producing [ɹ] in other contexts is very limited. The generalization that is most readily available to them, therefore, is for producing [ɹ] in prosodically weak, medial positions. Edwards et al. (2004) showed that children's repetition accuracy of phoneme sequences in non-words was correlated most strongly with the frequency of the sequence in the lexicon, thereby demonstrating the importance of sequential practice in a variety of cases.

The fact that innovation represents a much larger proportion of all substitutions in strong position is interesting for another reason. It suggests that recruitment of preexisting categories is

preferred whenever possible. Only when a D2 category cannot be identified with a preexisting one (see discussion of Best et al. (2001) and Flege (1995) above), or when implementation or articulation of a preexisting category is inhibited, does a speaker begin to explore the phonetic space. This reinforces Pierrehumbert's (2002) claim that the categorical behavior of the perception and production system is basic to the architecture, and that exemplar effects play a secondary role.

In our experiment, it was not practical to carefully control for the amount and type of language experience that subjects brought with them to the trials. It would have been impossible to determine, for example, whether a given subject had ever heard Glaswegian English, perhaps even unknowingly, in their lifetime. It would have been even less practical to rule out any subject who had prior experience or practice either with a different dialect of English that includes similar phonological features (e.g., Southern British with regard to /t/), or with an entirely different language that has similar phonetic categories in similar phonological contexts (e.g., Spanish with regard to /r/). What we do know, and what was verified by our Baseline condition, is that all subjects were native, first-language speakers of a dialect of English in which the relevant features of our study are not present. Furthermore, we know that there were no native speakers of Glaswegian English in our study. In fact, informal exit interviews suggest that most of our subjects could not identify the dialect they heard as a variety of English spoken in Scotland, and several could not even narrow its origin to the British Isles.

To give one example of the diversity of experience level among our subjects, one subject's productions sounded remarkably like those of a native speaker of an Indian variety of English, both on the training and generalization tasks. Her subject questionnaire, however, indicates that she is a native speaker of American English, and her baseline recordings confirm

this. In an exit interview, the subject reported having had significant contact with the India-born mother of a childhood friend, whom she had learned to imitate through repeated practice. In the subject's attempt to access the experimentally targeted dialect, the social index associated with the experientially robust Indian English features provided a strong competitor for the relatively weakly reinforced social index associated with the experimental dialect. The result is that she was quite successful both at producing [ɾ] for /r/ and at producing [tʰ] in weak position, in addition to other features more typical of Indian English than Glaswegian English. In keeping with our theoretical claims, then, this subject's behavior was both systematic and categorical, yet it displays the contextual and social biasing effects that are indicative of an exemplar component to the model.

Whatever the maximum level of speech experience was that our subjects brought to the experiment, any success they demonstrated in the tasks required one of two abilities. Either they replaced a preexisting category with a new one which they were able to generate parametrically, or they activated a preexisting category in a novel lexical and social context. Either way, the learning was systematic to the extent that it applied to both familiar and unfamiliar word and sentence contexts, and it was long-term, since it persisted over a period of one week.<sup>2</sup>

Comparing our results to Pierrehumbert's (2002) hybrid model then, we find support for the relevance of all four proposed mechanisms. To the extent that subjects in our experiment succeeded at replacing [ɾ] with the flap from their native dialect or from another language, they were able to modify their pronunciation of known words using preexisting categories, while those who succeeded by learning a novel articulation of /r/ demonstrated the ability to form new phonetic categories parametrically through exposure and practice. In both cases, subjects encoded these new pronunciations as generalized phonological principles. All subjects were able

to produce [t<sup>h</sup>] in weak contexts. Some subjects accomplished this by learning to produce a preexisting category in novel lexical contexts, while those whose native dialect includes [t<sup>h</sup>] in weak position demonstrated an ability to learn the situational bias associated with that pattern. In sum, our results show that systematic effects strongly dominate the learning mechanism, though exemplar effects play a secondary role to the extent that speakers showed weak word- and context-specific effects, as well as the ability to form new categories and learn socially and contextually indexed biases.

## 6. Conclusions

The ultimate question is what these results suggest about the speech production system. We suggest that the dominant effects, which show systematic transfer of an existing allophone to a new phoneme, accord best with neo-generative models such as those discussed in Maye et al. and Peperkamp and Dupoux (in press). However, there are small lexical effects which recall the effects found by exemplar theorists (Goldinger, 1998, 2000; Johnson, in press,). Therefore, the total picture might be captured best in a hybrid model (Pierrehumbert, 2002).

## Acknowledgements

We would like to acknowledge the support of a grant from the James S. McDonnell Foundation and an Andrew W. Mellon Postdoctoral Fellowship at Northwestern University.

## References

- Baese, M., & Goldrick, M. (2006). Lexical effects on phonetic variation independent of phonotactics. Poster presented at the Tenth Conference on Laboratory Phonology, Paris, June 29-July 1, 2006.
- Baker, K. (2004). Auditory classification of regular and irregular pseudoverbs. Paper presented at McWOP 10, Northwestern University, Evanston, IL. Oct. 29, 2005.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109(2), 775-794.
- Chirrey, D. (1999). Edinburgh: Descriptive material. In P. Folkes & G. Doherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 223-229). London: Arnold.
- Edwards, J., Beckman, M. and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47, 421-436.
- Fisher, W. M., & Hirsh, I. J. (1976). Intervocalic flapping in English. In *CLS 12-1*, 183-198. Chicago: Chicago Linguistic Society.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E., Yeni-Komshian, G., & Liu, H. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728-3740.

- German, J. (2004). Continuous or gradient? What accent imitation can reveal about two dimensions of tune. Unpublished manuscript, Northwestern University.
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 152-162.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166-1183.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279.
- Goldinger, S. D. (2000). The role of perceptual episodes in lexical processing. In A. Cutler, J. M. McQueen, and R. Zondervan (Eds.), *Proceedings of SWAP (Spoken Word Access Processes)* (pp. 155-159). Nijmegen: Max Planck Institute for Psycholinguistics.
- Goldrick, M., & Blumstein, S. (in press). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*.
- Hall, K. C., & Boomershine, A. (2006). Life, the critical period: An exemplar-based model for language learning. Paper presented at the Tenth Conference on Laboratory Phonology, Paris, June 29-July 1, 2006.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000a). Does the Queen speak the Queen's English? *Nature*, *408*, 927-928.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000b). Monophthongal vowel changes in

- Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, 30, 63-78.
- Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-166). San Diego: Academic Press.
- Johnson, K. (in press). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*.
- Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, 46, 295-349.
- Kiparsky, P. (1968). Linguistic universals and linguistic change. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 170-202). New York: Holt, Rinehart and Winston.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874-86.
- Levelt, W. J. M. (1980). *Speaking*. Cambridge, MA: MIT Press.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (submitted). The Weckud Wetch of the West: Lexical adaptation to a novel accent.
- Mendoza-Denton, N., Hay, J., & Jannedy, S. (2003). Probabilistic sociolinguistics: Beyond variable rules. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 97-138). Cambridge, MA: MIT Press.
- Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, 9, 283-303.
- Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: A perceptual study

- of dialect acquisition in adults. *Journal of Phonetics*, 27, 385-403.
- Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44, 778-792.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- Palmeri, T. J., Goldinger, S. D. & Pisoni, D. B. (1993). Episodic encoding of speaker's voice and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 309-328.
- Patterson, D. & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, 58, 254-275.
- Peperkamp, S. & E. Dupoux (in press). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9*.
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Amsterdam: John Benjamins.
- Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101-139). Berlin: Walter de Gruyter.
- Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177-228). Cambridge, MA: MIT Press.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Doherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, segment, prosody* (pp. 90-

- 117). Cambridge: Cambridge University Press.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89(6), 2961-2977.
- Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30, 271-287.
- Sankoff, G. (2004). Adolescents, young adults and the critical period: Two case studies from *Seven Up*. In C. Fought (Ed.), *Sociolinguistic variation: Critical reflections* (pp. 121-139). New York: Oxford University Press.
- Strange, W. (1995). Cross-linguistic studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-linguistic research* (pp. 3-48). Timonium, MD: York Press.
- Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Folkes & G. Doherty (Eds.), *Urban voices: Accent studies in the British Isles* (pp. 203-222). London: Arnold.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41, 177-204.
- Whalen, D. H., Best, C. T., & Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25, 501-528.
- Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t,d/ in American English. *Journal of the Acoustical Society of America*, 66(4), 1039-1050.

## Figure Captions

Figure 1: Production (left) and perception (right) architecture for primary finding of Maye et al. (submitted) and Peperkamp and Doupoux (in press). Generalization occurs because of realignment at the level of phonemic encoding (represented by dotted arrow).

Figure 2: Mean percentage of [t<sup>h</sup>] outcomes for /t/ in strong and weak positions.

Figure 3: Mean percentage of [r] outcomes for /r/ in strong and weak positions.

Figure 4: Correlation between Generalization1 and mean Week2 performance, by subjects, for /r/ in strong position.

Figure 5: Correlation between Generalization1 and mean Week2 performance, by subjects, for /r/ in weak position.

Figure 6: Percentage of successful Glaswegian allophones, Week2, by conditions.

Figure 7: Percentage of successful Glaswegian allophones, Week2, by order.

Figure 8: Percentage of flaps in the /r/ items in strong positions, intervocalic (33) vs. non-intervocalic (15), plus percentage for /r/ in weak positions.

Figure 9: Mean percentage of flap recruitment and innovated (non-American) outcomes, /r/ in strong position.

Figure 10: Mean percentage of flap recruitment and innovated (non-American) outcomes, /r/ in weak position.

Figure 11: Waveforms for three subjects' productions of the phrase *new road*.

11a. Subject C5a, Training2 block: /r/ → [r]

11b. Subject B1a, Generalization1R block: /r/ → [z]

11c. Subject C6a, Training1 block: /r/ → [ɾ]

Figure 12: Perception (left) and production (right) architecture for proposed model in which realignment occurs at the level of subphonemic encoding. **(A)** Lexical learning occurs when specific lexical items are associated with new subphonemic variants. **(B)** Categorical reassignment occurs by generalizing over instances of lexical learning.

Table Caption

Table 1: Recording conditions by week

## Endnotes

---

<sup>1</sup> Similar interactions of phonological generalization with lexical items can also be captured in cascading connectionist models (Goldrick & Blumstein, in press; Baese & Goldrick, 2006).

<sup>2</sup> In this initial study, the presence of orthographic support from the written script maximized the chances that subjects would exhibit generalized learning. In a similar task without orthographic support, we would expect to see the effects of subjects failing to recognize words.

## Appendix 1: Target items

### Block A

1. The class does yoga on the matting.
2. Some day, he will find some courage.
3. The hill-dwelling monks can be seen building a temple.
4. The baby consumed a bowl of rice.
5. I suppose the illness caused his delirium.
6. He does the job, though he sounds funny when he talks.
7. We will need the long rope.
8. The family's chubbiness was mainly genetic.
9. Chuck always goes on the ferry.
10. Place the hassock inside the room.
11. The slimy animal by the pool is a toad.
12. The damp wind made him all sweaty.
13. The castle was held by a rebel.
14. My niece likes playing with Tonkas.
15. Leah was planning a vacation in Florence.
16. The yelling of the fans was muted.
17. The memo was funny because of a typo.
18. By the end of the movie, love was found by the heiress.
19. Deep in the woods was an old cottage.
20. He smoothed the edges with a rasp.

21. Chicago has a famous marathon.
22. This essay will be done on time.
23. The peace negotiation was plagued with racism.
24. The ball had seemed unhittable.
25. The candy bin is full of toffee.
26. The kids sang a silly rhyme.
27. How many books can you carry?
28. No one in the family believed Uncle Bob was batty.
29. An unlucky buck has a wide rack.
30. Jonathon likes milk in his porridge.
31. Good things come in twos.
32. He's thinking of the beans he's been eating.
33. The pancakes could be good with syrup.
34. The chalice was classified as a relic.
35. The swamp was the location of a big battle.
36. Though a snake, the python is tame.
37. Jack landed the salmon in a riffle.
38. The window glass was held in with putty.
39. In college, you buy books by the ton.
40. Bad shampoo can be made with oranges.
41. In the zoo, he saw a lonely rhino.
42. By one a.m., they deployed the shuttle.
43. The boy swallowed mud because he was curious.

44. Lyme disease is often blamed on ticks.
45. Ned's love of walking could be called fanatic.
46. Civil though she may be, his feelings could be ruffled.
47. The shah was in a fury.
48. The maid needs help with this task.

#### Block B

1. With heavy use, the cloth became a rag.
2. I believe the fish of the day is whiting.
3. The valley is unlivably arid.
4. I seldom see Melanie in town.
5. The log was the home of a raven.
6. A nice pie will be made with the berries.
7. The cook slowly made the beef patties.
8. He gave away his only token.
9. Then he skillfully sings an aria.
10. The village was enslaved by the Romans.
11. Selma's clothing was always fashionable and exotic.
12. Bush has an obsessive love of low taxes.
13. The couple enjoyed choosing a ring.
14. The passage of this bill is vital.
15. The new Nanolab will have unique tools.
16. In the old days, you soothed a baby with marrow.

17. The chess club held one final meeting.
18. Fax me a copy of his resume.
19. The boss came in a toga.
20. The message lacks an obvious moral.
21. The young couple should speak with a rabbi.
22. Someone should clean the tiles.
23. The flu is caused by a virus.
24. The gossip in the school was awfully petty.
25. The small flying thing is a wren.
26. In the evening, he munches cereal.
27. The navy loaned him a tank.
28. The news channel mentioned a UFO sighting.
29. Jacques lives cheaply in Paris.
30. This wood will be used in making a table.
31. He was deafened by the rifle.
32. The bed was below the folds of netting.
33. The essay should have a specific topic.
34. The sickly youth has no endurance.
35. Leo saw something askew in the rhombus.
36. Len's business office was inside the city.
37. He did fax them one query.
38. These bananas look ripe.
39. In the field I found a Mayan fetish.

40. You spoke slowly on the tape.
41. His whole life, Jack had been in a hurry.
42. Excess cleavage in an office is unsuitable.
43. I saw Andy in the hall with his twin.
44. The small child won the race.
45. The lamb seemed happy, though amazingly little.
46. The infection began in his tonsils.
47. Picasso designed this epic mural.
48. A Chicago dog always comes with relish.

#### Block C

1. On the weekends, old men walk along the Thames.
2. The sheep dog was lying by the rock.
3. Mrs. Jackson came up with a new theory.
4. The cheese by the olives is feta.
5. In the necklace was a humongous ruby.
6. The judge scolded the jury.
7. The ocean has both low and high tides.
8. The thief thinks she can escape all notice.
9. The policeman sounded his siren.
10. The ad was awfully racy.
11. Happiness is only fleeting.
12. The small dog was wagging his tail.

13. Climbing in the Himalayas involves many risks.
14. Insulation is made with batting.
15. Good fishing begins with good tackle.
16. She's in Mexico, climbing a Mayan pyramid.
17. A guinea pig is amazingly pettable.
18. The house had become an old ruin.
19. Galahad was anxious when he was in peril.
20. With his pencil, he keeps an ongoing tally.
21. Clownfish and sea anemones live on the reef.
22. Well, the man has had some experience.
23. Life was no fun among the Ottomans.
24. Five people play on the team.
25. The milk was sold in the dairy.
26. This salad needs six kinds of lettuce.
27. Along the lake, the couple cycled in tandem.
28. The slugs will avoid the roses.
29. The company sells useless insurance.
30. The flaw is on the tip.
31. The leak in the hull was sealed with a special resin.
32. The old donkey was given a heavy beating.
33. The diva was accompanied by a full chorus.
34. This season, the high-heeled shoe is all the rage.
35. He was killed by an unknown toxin.

36. Old wigs belong in an attic.
37. The king has no loyal men in the realm.
38. The coffee will keep Jan cozy inside the tollbooth.
39. This cloud looks like a cirrus.
40. In five days, the blooms will lose some petals.
41. The nebula is visible with a telescope.
42. The milkshake was done up with a cherry.
43. The consul said they've been invited.
44. How do you like the new rug?
45. In Vilnius, you can buy amazingly spicy curry.
46. A missing copy was shown by all the tags.
47. The dog could become rabid.
48. The canoes should be inflatable.

#### Block D

1. His language was so foul, only one line was quotable.
2. My dad has many worries.
3. He lived his life by the wisdom of the Talmud.
4. Seafood gives me a rash.
5. Ms. Jones gave examples of Eskimo-Viking borrowings.
6. When you sneeze, please use a tissue.
7. The clouds opened up and she saw a rainbow.
8. The usefulness of this device is debatable.

9. A sunny vacation leaves you looking tan.
10. Will they have a good marriage?
11. All the family's belongings lay beneath the rubble.
12. With a knife and some wood, you could whittle.
13. Do you think Sheila saw the heron?
14. The possum climbed up on the roof.
15. People in confined spaces can become catty.
16. As a hobby, Kim plays the timpani.
17. The police chief made the mob leave the area.
18. How many novels have you completed?
19. His cap held a long, golden tassel.
20. Sue could think of a good reason.
21. The special comes with pita.
22. The young amphibian became a tadpole.
23. Mike was amazed when he won the raffle.
24. Somehow she can deal with his snoring.
25. Sam found a casino and began betting.
26. His only companion is an unspeaking wrasse.
27. Even with my glasses, my vision is blurry.
28. Melvin hailed a taxi.
29. In the oven, she is baking some rolls.
30. How do hyenas find carrion?
31. The woman smiled with pity.

32. The café gave him a choice of teas.
33. I think Kim and Mike sound serious.
34. You always pick wrong.
35. The biology class was discussing a beetle.
36. Did one of the halfbacks pull a tendon?
37. The chess game was won with the rook.
38. The whale is a mammal and aquatic.
39. Business is done on the telephone.
40. This ice cube has a funny appearance.
41. We will film the movie in a jungle setting.
42. He sold me a ribbon.
43. Becky enjoys chewy candy like taffy.
44. The lilac bush is beside the sorrel.
45. We'll need his decision on the new road.
46. He yelled when he chewed his tongue.
47. The guinea pig sleeps in a nice burrow.
48. The ocean waves pounded the jetty.

## Appendix 2: Non-target items

### Block 1

1. A display of the dig can be seen in the lobby.
2. Dolphins swim and play alongside the ship.
3. May I buy some chicken feed?
4. In the piano division, the champion was Michael Hawley.
5. The clock face glows dimly in the evening.
6. The second copy of the code has many bugs.
7. If the ball bounces on this wall, then the game ends.
8. Simply place the apple on the napkin with a bow.
9. Food supply in the developing nations should be closely followed.
10. Some books will always be appealing.
11. Some music could soothe the savage babies.
12. Olga was hoping the food would be well-done.

### Block 2

1. Why should he push himself, when he has all the money he can use?
2. The only way one could fail his class is by sleeping when he gives the quizzes.
3. The sheep in the field fell down in the wind.
4. Seven men will be assigned these five offices.
5. Melanie's solution is a classic in the field.
6. Physics labs will be open by noon in the fall.
7. The Buck company will spend a million and fix the building.

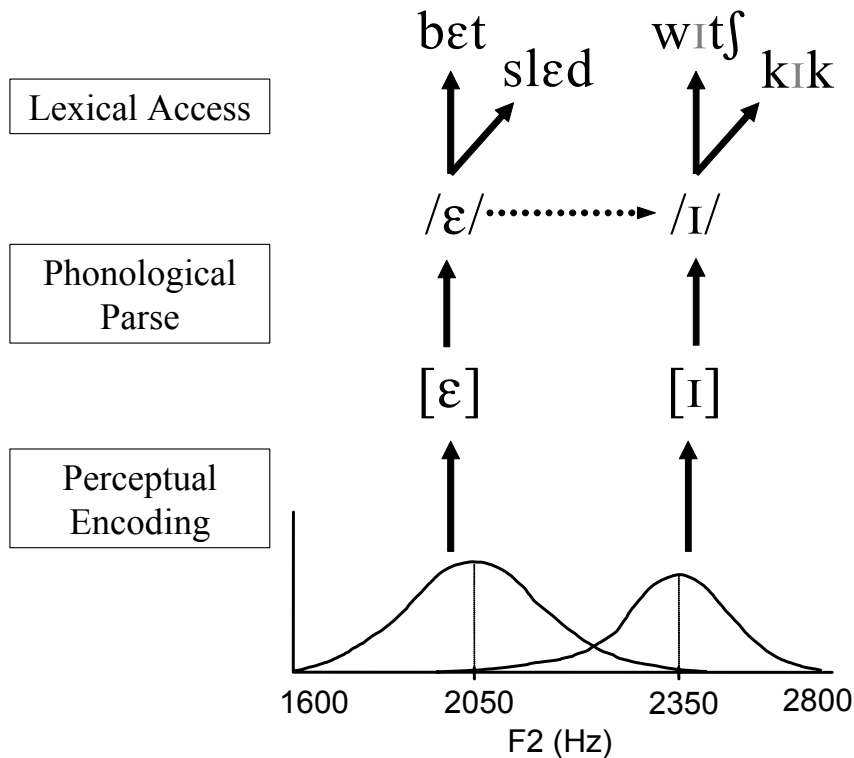
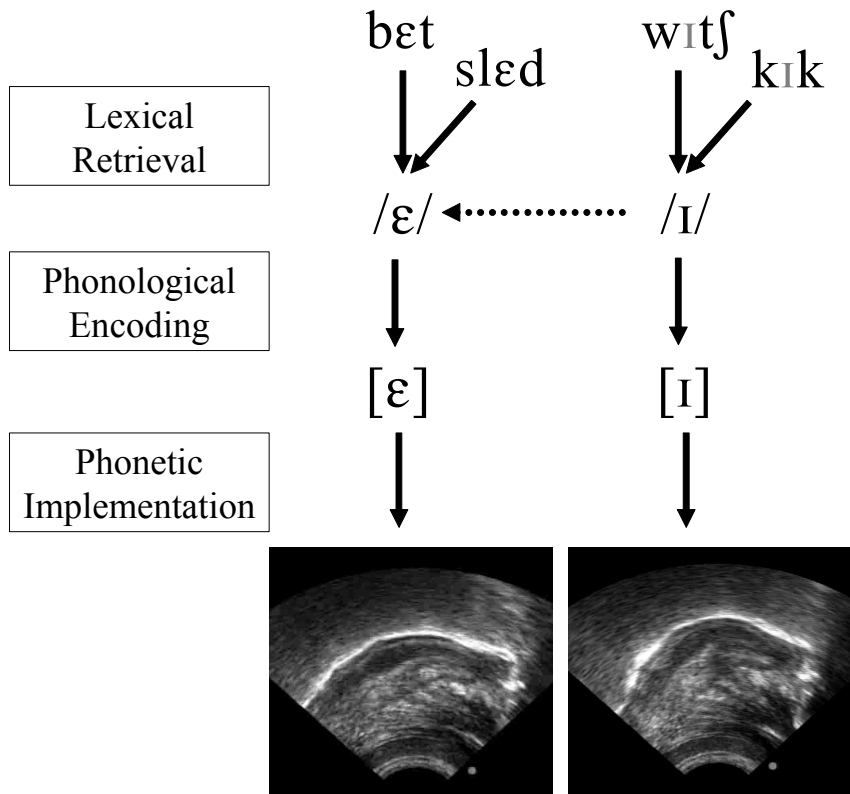
8. The koala seems so lovable and sleepy.
9. She finally gave up on being queen of all the lands.
10. Of this class, only one will be successful.
11. His mind shows signs of senile decay.
12. When five decades have gone by, you'll need a new guide.

### Block 3

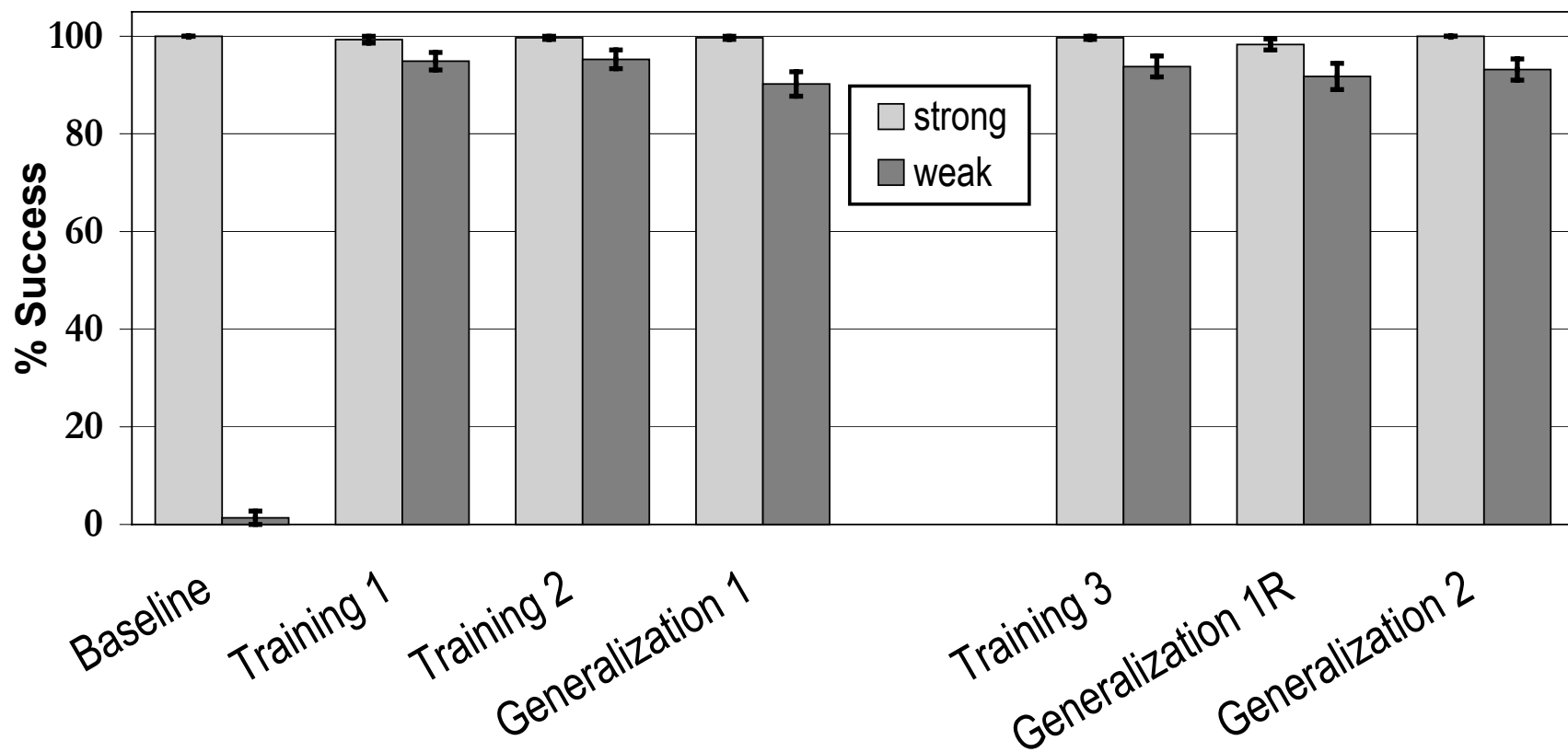
1. The glass was oozing a luminous fluid.
2. The gel was molding in the shape of a buffalo.
3. She found a bag of cash in the subway.
4. Globbs of muck fell along the sides.
5. Louis loves the sound of moaning voices.
6. Pam only knows osmosis, so she failed the biology exam.
7. The evil demon came in a puff of smoke.
8. The hail in Spain falls mainly on the mesas.
9. The invoice displayed the shipping and handling fees.
10. The film was shown in the evening.
11. The young musk oxen can be pleasingly affable.
12. Winning the game would be a bonus.

Table 1: Recording conditions by week

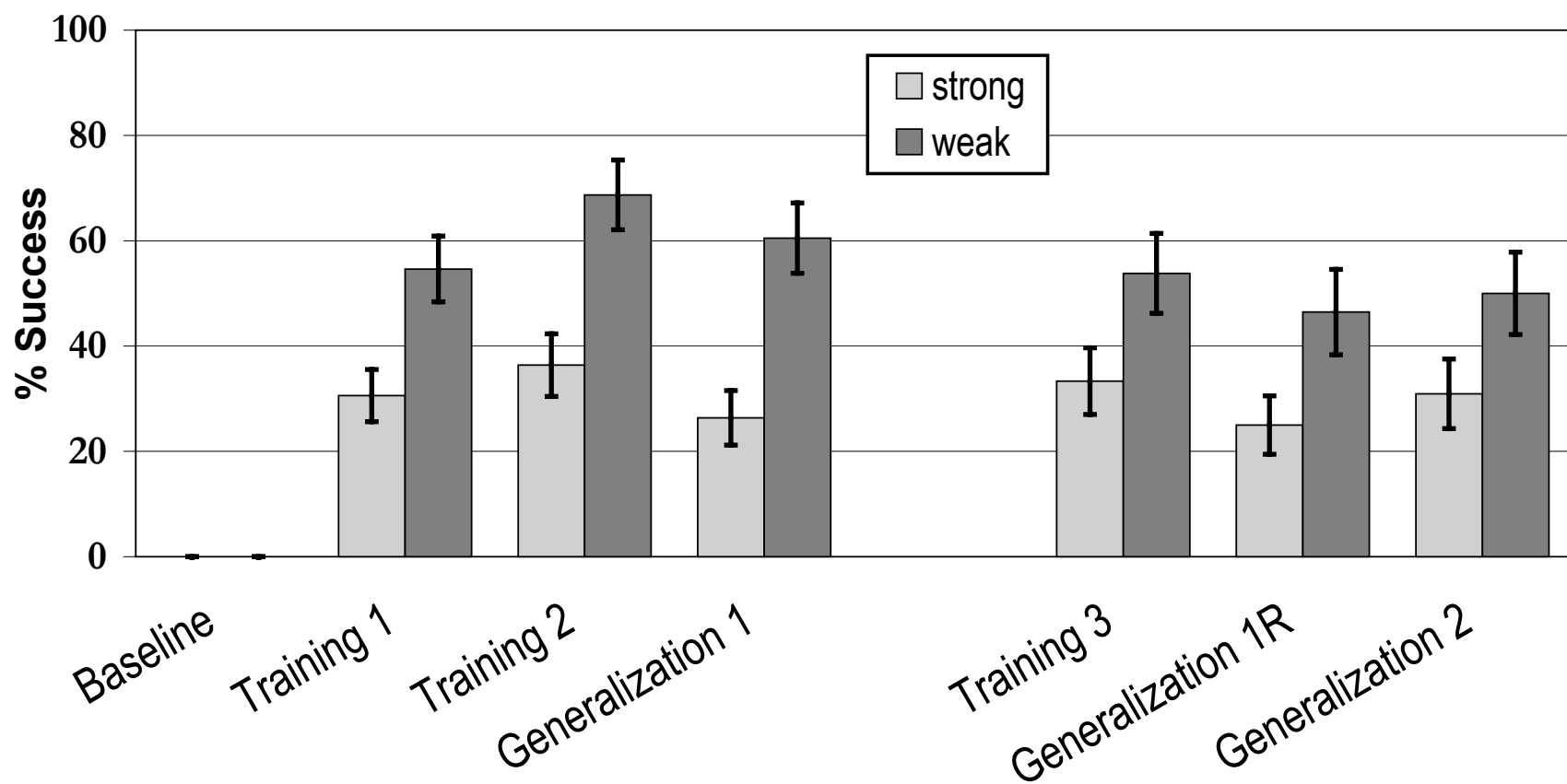
Week 1	Week 2
Baseline	
Training1, Training2 (with CD)	Training3 (without CD)
Generalization1	Generalization1R
	Generalization2



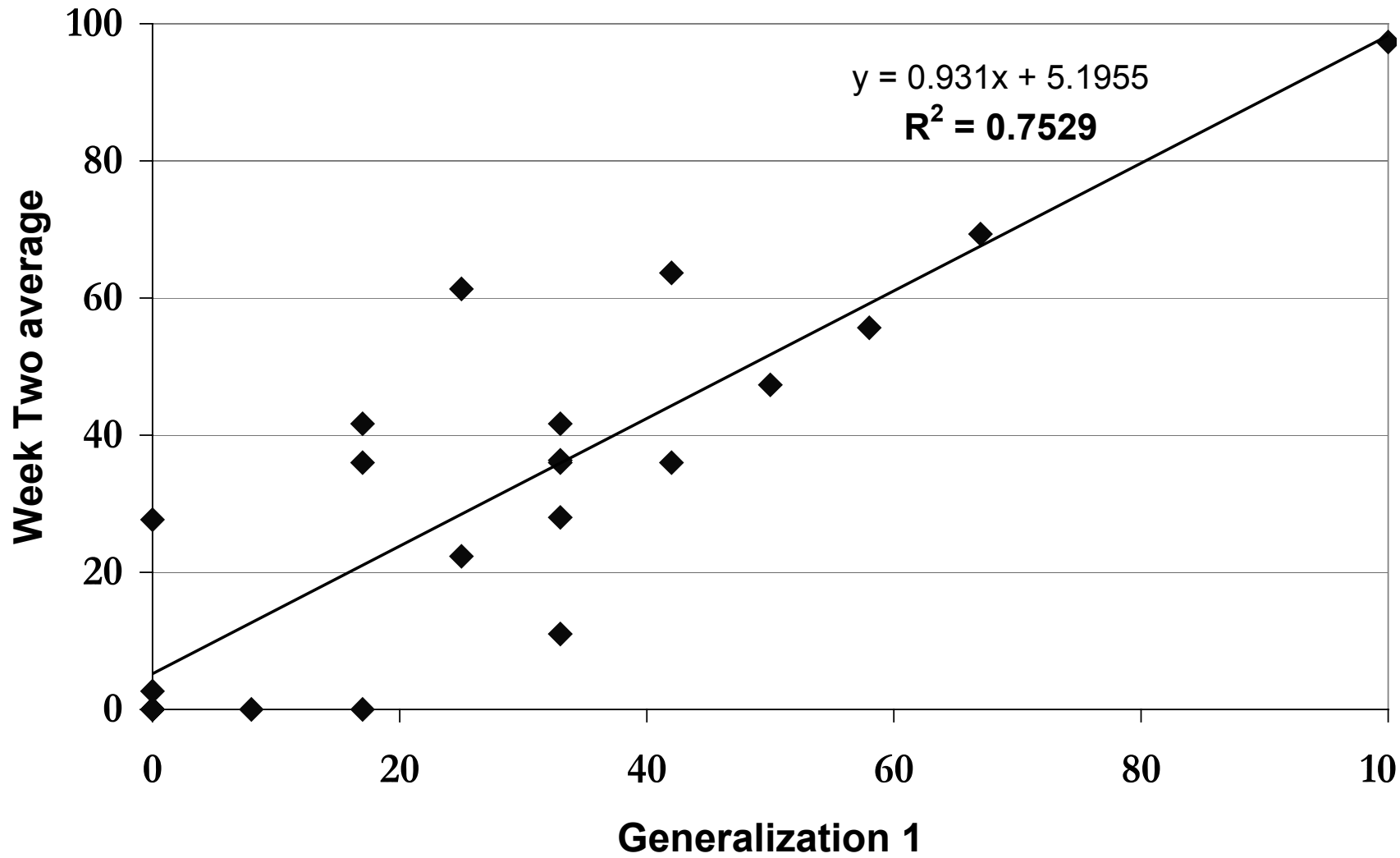
Figure



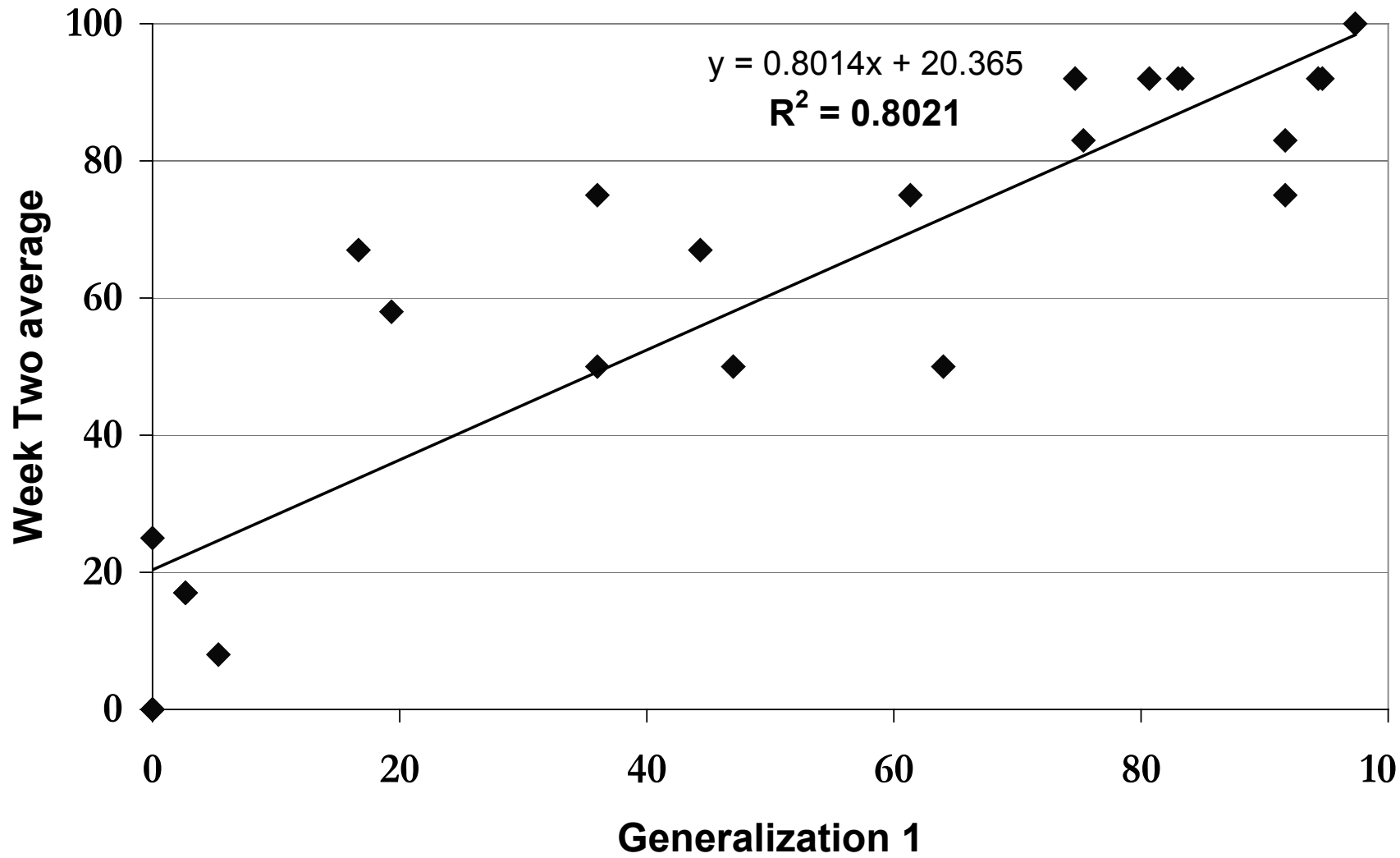
Figure



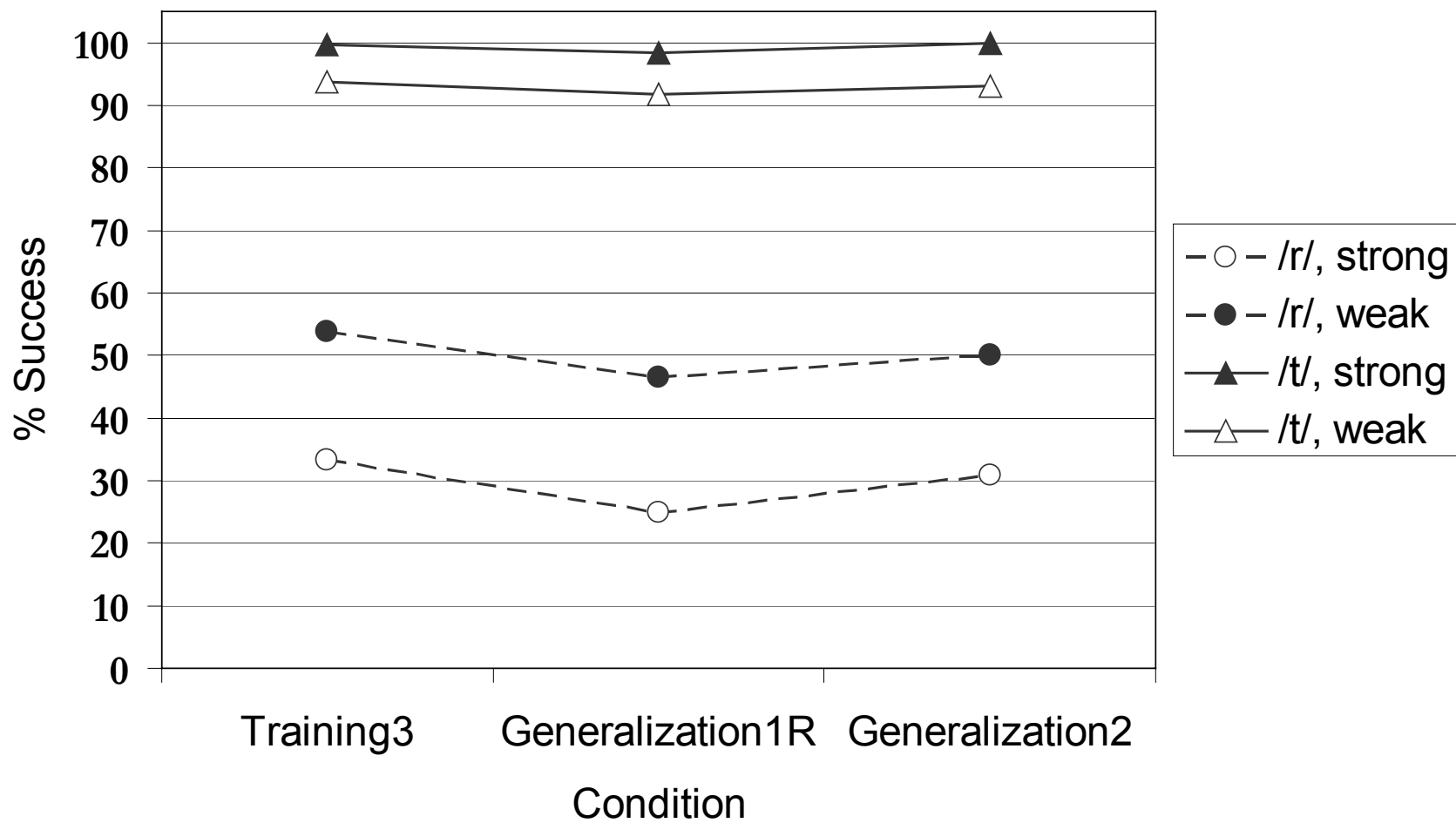
Figure



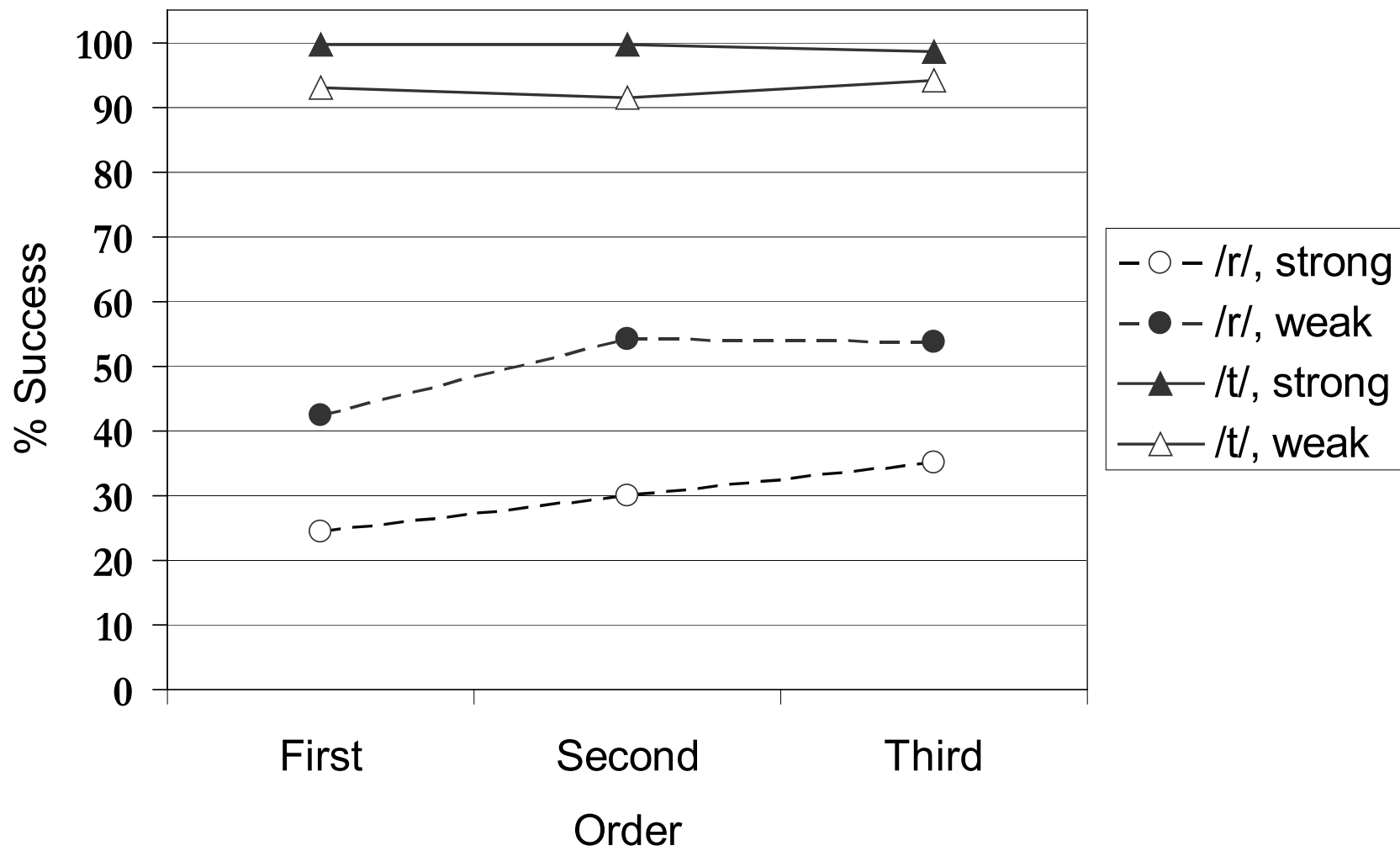
Figure



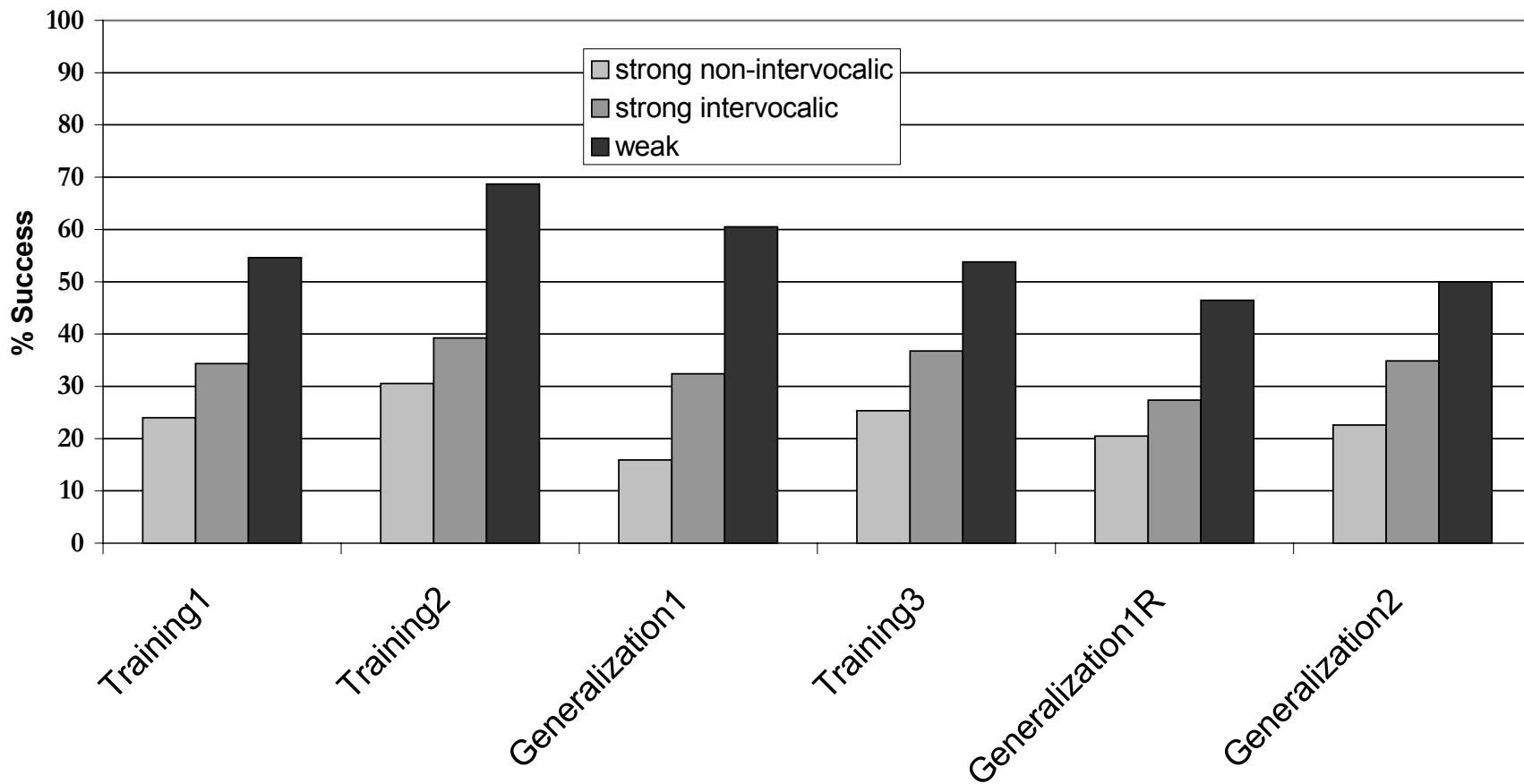
Figure



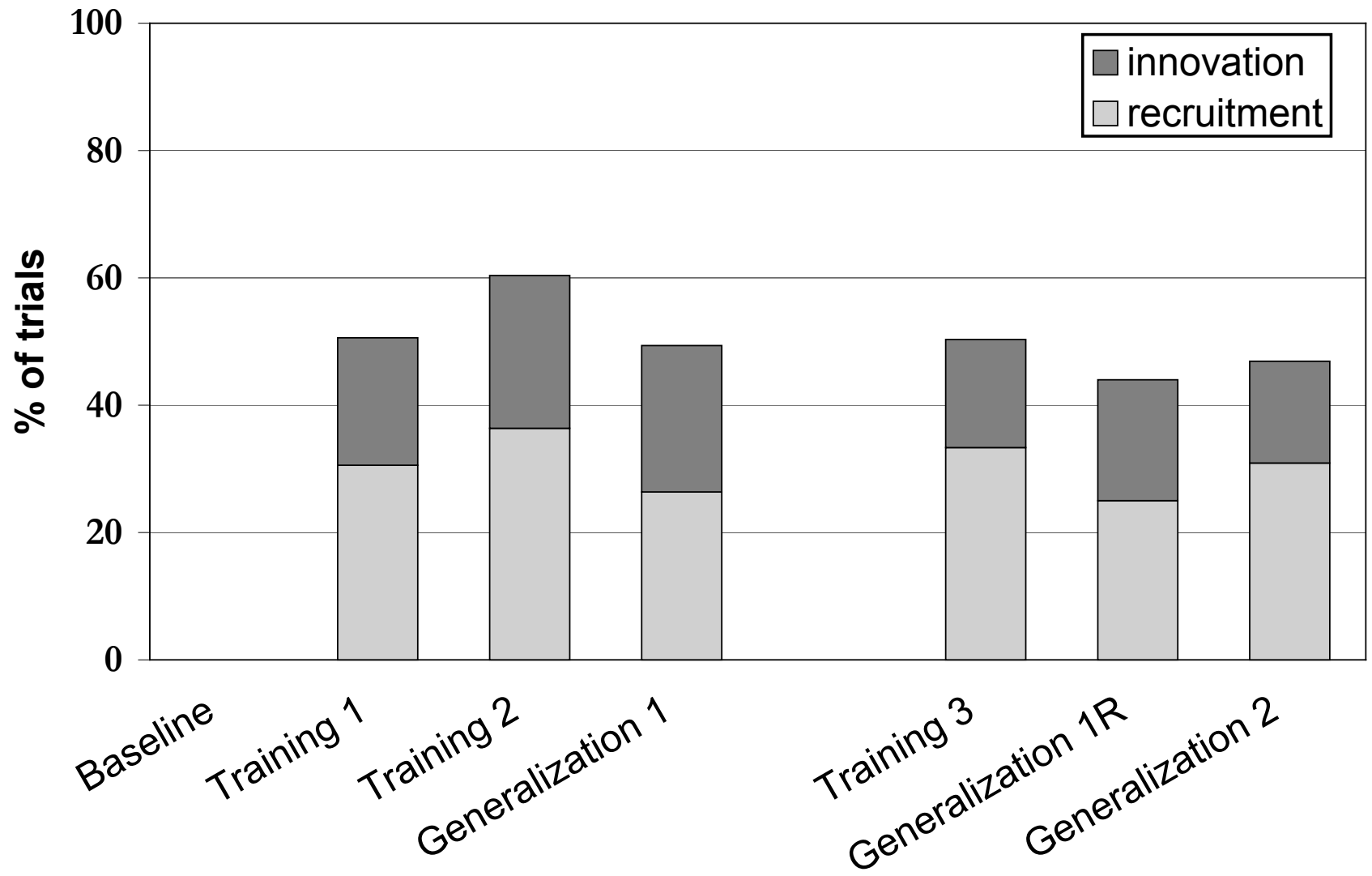
Figure



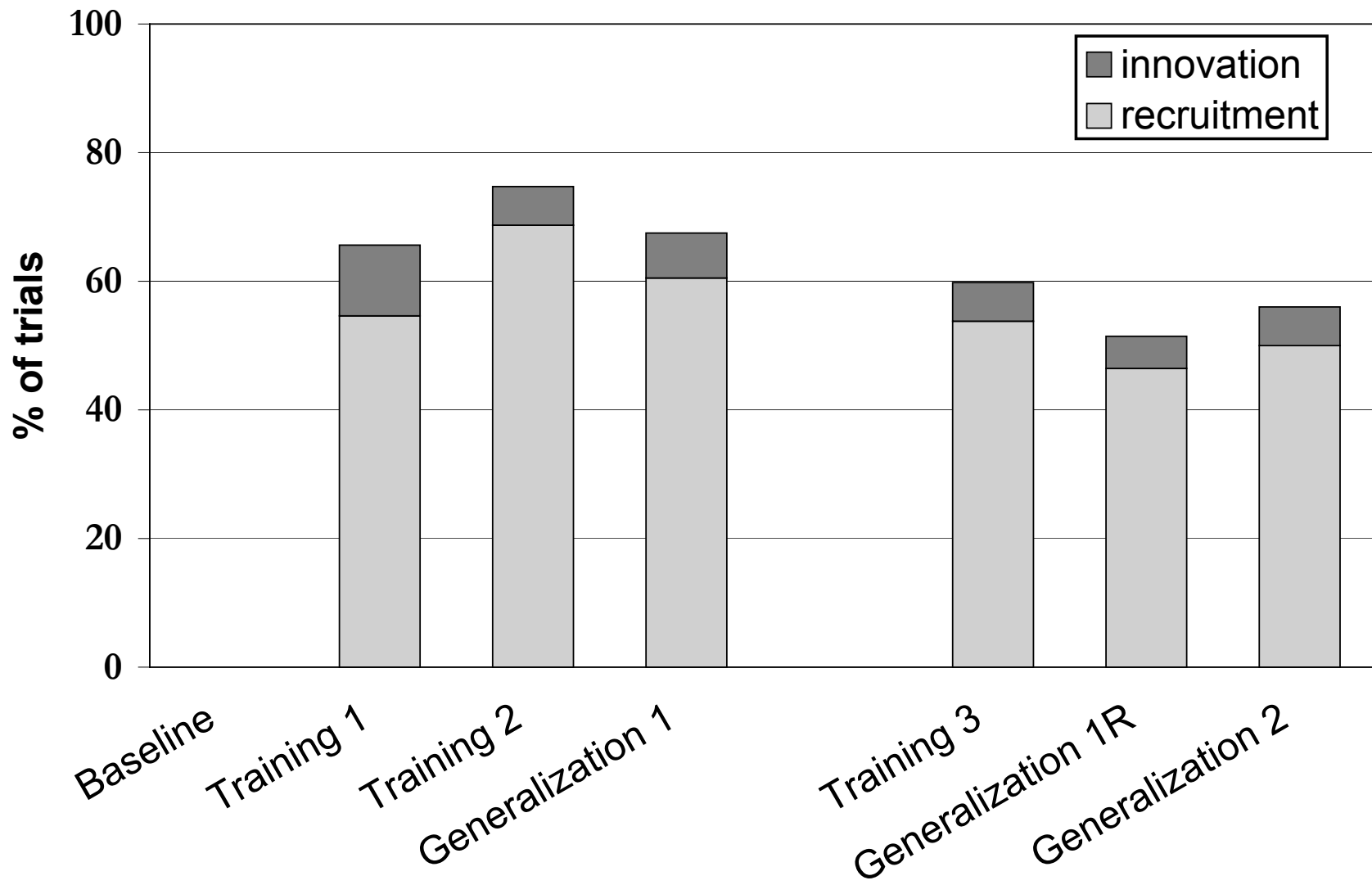
Figure



Figure

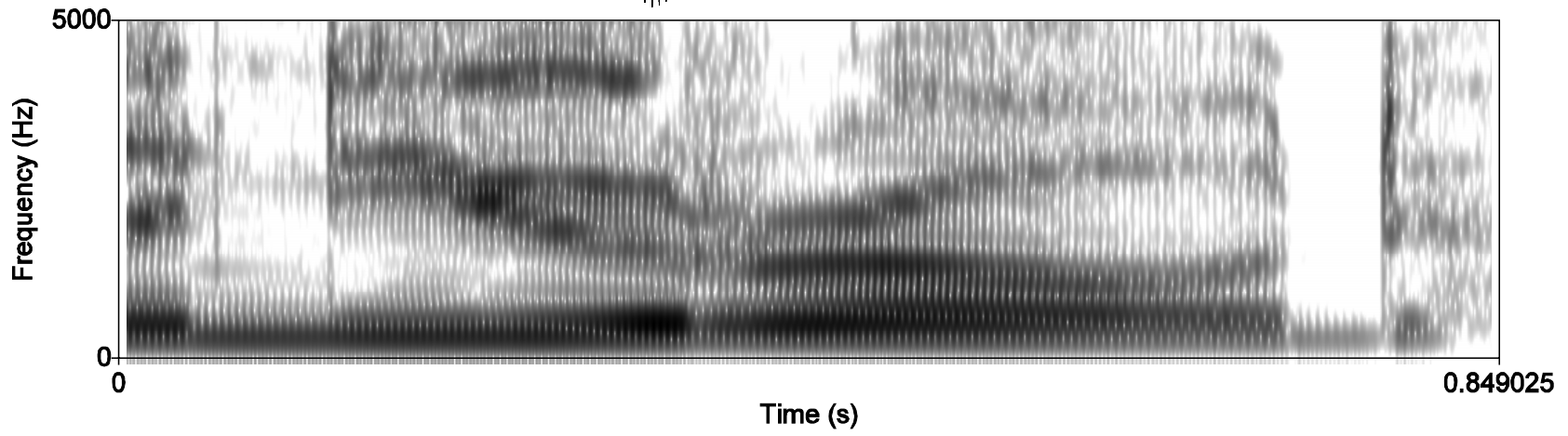
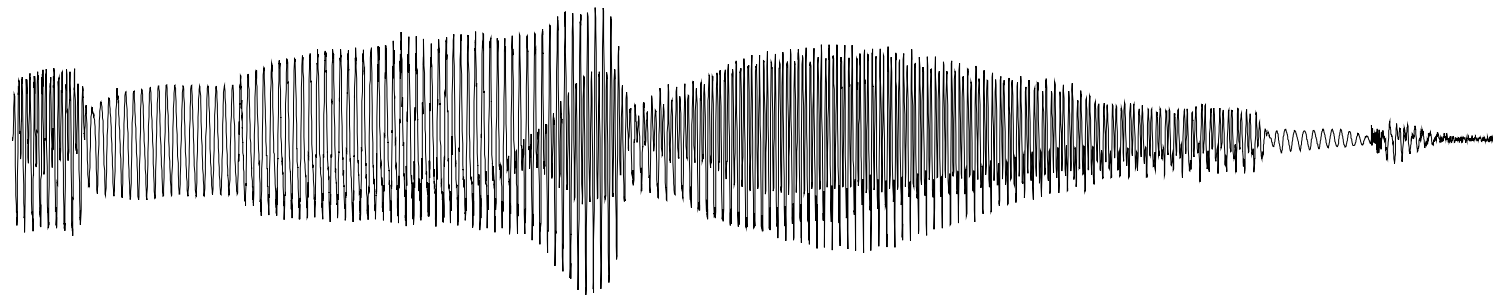


Figure

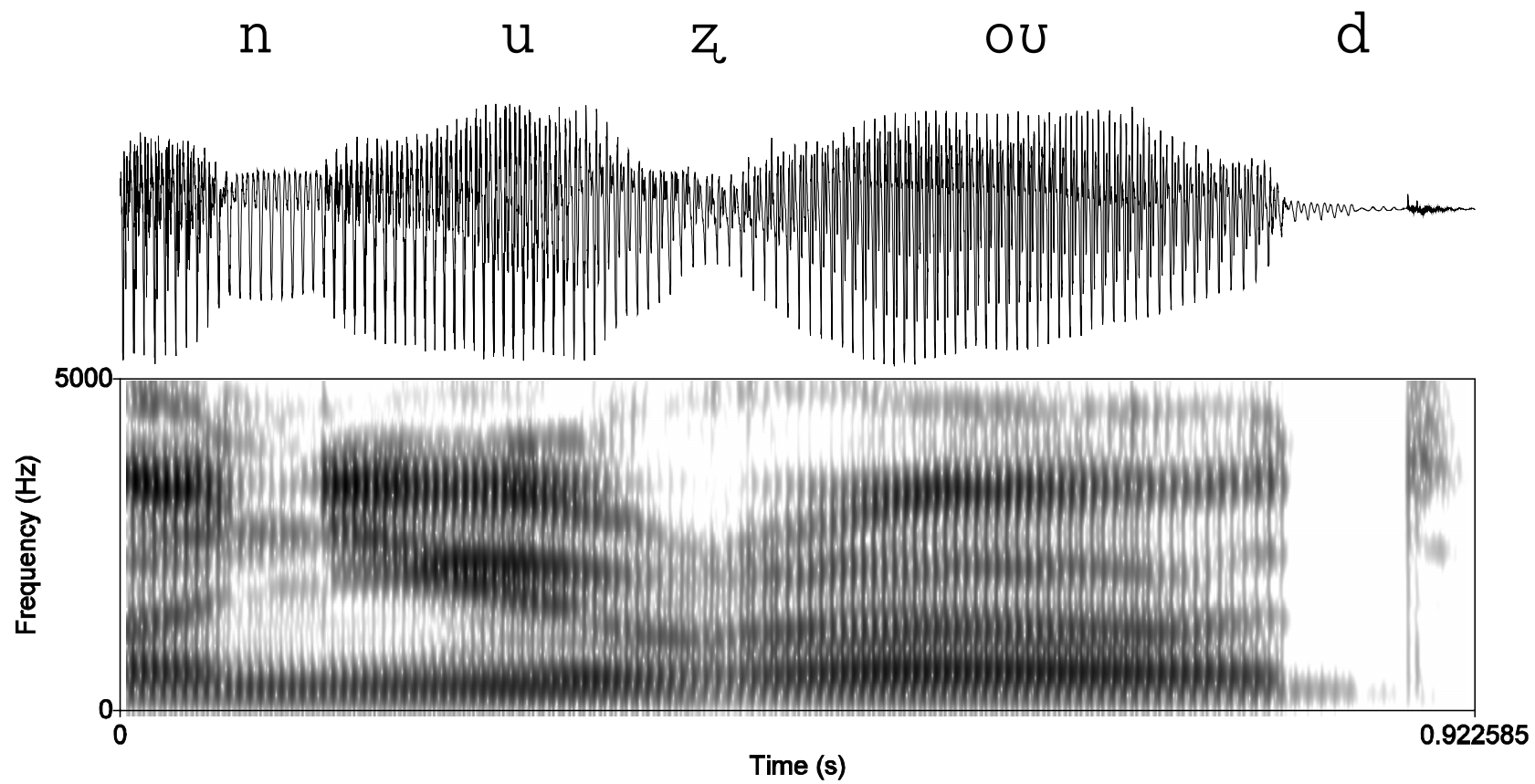


Figure

n u r ou d



Figure



Figure

