

Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions

DAVID EDDINGTON*

Abstract

The distribution of the flap allophone [ɾ] of American English, along with the other allophones of /t/, [t^h, t[̄], ɾ, t] has been accounted for in various formal frameworks by assuming a number of different abstract mechanisms and entities. The desirability or usefulness of these formalisms is not at issue in the present paper. Instead, a computationally explicit model of categorization is used (Skousen 1989, 1992) in order to account for the distribution of the allophones of /t/ without recourse to such formalisms. The simulations that were carried out suggest that they are not needed because analogy to surface apparent variables such as phones and word boundaries is sufficient to predict allophony.

In analogy, the particular allophone of /t/ (i.e. [ɾ, t^h, t[̄], ɾ, t]) that appears in a given context is determined on the basis of similarity to stored exemplars in the mental lexicon. From an acquisitional standpoint, categorization by analogy to stored exemplars dispenses with the need for rule induction although it does suggest that speakers group functionally related sounds into mental categories, a process that is influenced to a great deal by orthography.

Analogy also explains the stochastic nature of linguistic performance. In the present study, 3,719 tokens of the allophones of the phoneme /t/ were extracted from the TIMIT corpus and constitute the database from which analogs were chosen. The variables used included the three phones or boundaries on either side of /t/, and the stress of the syllables preceding and following /t/. The model proves quite successful in predicting the correct allophone, and the errors made are generally possible alternative pronunciations (e.g. moun[ɾ]ain, moun[t^h]ain). The success rate changes little when only small sub-samples of the database are incorporated. In addition, exemplar-modeling is found to be quite robust because even when a feature such as stress is eliminated, (a feature which is critical in most rule approaches), allophony is still highly predictable.

Keywords: analogy; flapping; tapping; American English; phoneme /t/; Analogical Modeling of Language; allophonic distribution

1. Introduction

The pronunciation of /d/ and /t/ as a flap is observed in many dialects of English to varying degrees but is particularly prevalent in the English spoken in North America. The flap is by far the most frequent variant in many lexical items (e.g., *butter*, *city*) and there is evidence to suggest that such words are stored in the mental lexicon with flaps rather than an underlying /t/ (Connine 2004). In fact, the association of the flap with the phoneme /t/ is most likely due to the effect of orthography since it comes about in to a great degree as children become literate (Treiman et al. 1994). However, flapping is not restricted to particular words but is a highly productive process that applies to neologisms and borrowings. Perhaps for this reason it has attracted the attention of so many scholars (Byrd 1994; Connine 2004; Davis 2003; de Jong 1998; Egido and Cooper 1980; Harris 1994; Kahn 1980; Kiparsky 1979; Laeufer 1989; Nespors and Vogel 1986; Parker and Walsh 1982; Patterson and Connine 2001; Picard 1984; Rhodes 1994; Riehl 2003; Selkirk 1982; Steriade 2000; Strassel 1998; Turk 1992; Zue and Laferriere 1979).

Exactly how the allophonic variants of /t/ developed in American English is not the focus of the present paper,¹ nor is how it is affected by sociolinguistic factors, although it has been shown to be affected by both sociolinguistic variables (Byrd 1994; Strassel 1998; Zue and Laferriere 1979) and linguistic factors (Egido and Cooper 1980; Gregory et al. 1999; Laeufer 1989; Nespors and Vogel 1986; Parker and Walsh 1982; Patterson and Connine 2001). The great bulk of the research on flaps and the other allophones of /t/ in English has focused on the phonetic environment in which flapping occurs and which of the aspirated, released, unreleased, and glottalized realizations of /t/ appears. In order to account for the allophones of /t/ these approaches make use of a number of formal mechanisms that are not surface-apparent (e.g., *rule ordering*: Jensen 1993; Kahn 1980; Kiparsky 1979; Nespors and Vogel 1986; *resyllabification*: Kahn 1980; Selkirk 1982; *prosodic feet*: Davis 2003; Giegerich 1992; Kiparsky 1979; Nespors and Vogel 1986; *phonetic licensing*: Harris 1994; *abstract features*: Kahn 1980; Kiparsky 1979; Nespors and Vogel 1986; Selkirk 1982). In this article, I assume that while such mechanisms may be useful in an analysis of linguistic competence, they cannot play a part in linguistic performance. For this reason, I present a psychologically plausible account of the allophones of /t/ which is based on surface-apparent properties. I demonstrate that the allophones of /t/ are predictable

without recourse to abstract entities by using a computationally explicit analogical algorithm. This framework proves to be highly robust because even when a supposedly crucial contextual cue such as stress is absent, allophonic distribution may still be accounted for.

2. Allophonic distribution as analogical categorization

Traditionally, those who assert that formal analyses mirror actual processing suggest that during acquisition children hear allophones such as [t^h], [r], and [ʔ] and come to relate them to the underlying phoneme /t/. Subconsciously they also intuit the context in which each of these occur, store that information, and use it in subsequent speech processing. A number of more recent formal approaches (Benua 1995; Burzio 1996; Kenstowicz 1996; McCarthy 1995; McCarthy and Prince 1994a, b; Steriade 1997, 1999, 2000) have begun to move toward the type of alternative approach I take in this paper; that is, they acknowledge the influence of fully specified surface forms on linguistic processing rather than on strict derivation from underlying forms. In like manner, many functional approaches recognize the important role that exemplars play in linguistic processing (Beckman and Pierrehumbert 2003; Bybee 2001; Hall 2005; Lakoff 1987; Langacker 1991; Silverman 2006; Solé 2003). I couch this article in Skousen's exemplar-based framework (1989, 1992) which I describe below.

In many formal models, a minimal lexicon is assumed that contains only unpredictable features. This means that a great deal of online processing must be performed on the underlying forms to produce the surface forms. Exemplar models shift this perspective to one in which the lexicon contains massive amounts of stored linguistic experience that includes even predictable, redundant, messy details. The information is stored in a network of highly interconnected entities. Connections are made based on semantic, morphological, phonological, orthographic, social, pragmatic, and other contextual similarities. Rather than assuming that people tacitly glean generalizations from the linguistic input and store them as separate entities of some sort, speakers refer to the database of stored experience in the course of linguistic processing to determine things such as allophonic distribution. This sort of storage is responsible for the probabilistic knowledge that speakers have about their language. People appear to learn probabilities associated with linguistic forms and put them to use in language processing (Labov 1994; Solé 2003).

Exemplar models are supported by psycholinguistic studies that demonstrate that not only are words stored as types, but individual tokens of the same word are stored in long-term memory (Goldinger 1997; Hawkins

2003; Palmeri et al. 1993). Words also appear to be stored with all of the phonetic detail present in the speech signal including information about frequency of usage, rather than in a form that has all redundant and supposedly irrelevant phonetic details abstracted away (Boomershine 2006; Brown and MacNeill 1966; Burton 1990; Bybee 1994, 2001; Coleman 2002; Fougeron and Steriade 1997; Kolers and Roediger 1984; Pisoni 1997).

The goal of the simulations described below is to demonstrate how the distribution of the allophones of /t/ in American English may be accounted for in an exemplar model. The definition of phoneme that I assume throughout this paper was probably first proposed by Baudouin de Courtenay (1895) and is ascribed to by more contemporary researchers such as Lakoff (1987), Nathan (1986), and Langacker (1987). According to this view, the phoneme is a conceptual category comprised of sounds that speakers feel are all instances of the same concept or unit. However, this does not mean that lexical items are stored as abstract phonemic units that are given allophonic realization in the course of production, only that certain sounds in a particular word or morpheme are conceptualized as being instances of the same category. For the purposes of the present article I discuss six allophones of /t/: [r], [ʔ], [ø], [t], [t⁼], and [t^h].

The aim of the present study is not to model how speakers come to categorize sounds into phonemes, but rather to show how categorized instances and analogy to them may underlie knowledge of allophonic distribution. Nevertheless, I assume that assigning sounds to phonemes is carried out based on factors such as phonetic similarity, morphology, semantics, and phonetic context (Bybee 1999; Mompeán-González 2004; Silverman 2006; Treiman et al. 1994). All of these converge on the word as a crucial unit. That is, people come to view all of the different pronunciations of /t/ in *sent*, for example, as instances of the same category because they all appear in the same word.

For literate English speakers another factor plays an important role in categorizing such phonetically disparate sounds as [r], [ʔ], [ø], and [t^h] together—the influence orthography (Skousen 1982; Treiman et al. 1994). Orthography is most likely responsible for the fact that English speakers tend to think of speech as containing a small set of sounds, while speakers that use a syllabary have a difficult time separating the components sounds in a syllable. The influence of the written word on the mental representation of phonemic categories probably differs between literate and illiterate speakers. Given the powerful effect of orthography on categorization, those who are uncomfortable with the notion of phoneme, (along with the fact that I have predetermined which allophones belong to the phonemic category /t/), may consider the simulations described below

as being designed to determine what stop or stop-like pronunciations orthographic *t* and *tt* receive in American English.

2.1. Modeling by analogy

In cognitive linguistics, knowledge is assumed to arise as a result of linguistic experience, with categorization of such experience playing a key role. Advances in computer science have resulted in a number of computer algorithms that may be used to test hypotheses about how categorized instances relate to cognitive processes (Aha et al. 1991; Daelemans et al. 2001; Medin and Schaffer 1978; Nosofsky 1988, 1990; Pierrehumbert 2001; Riesbeck and Schank 1989; Skousen 1989, 1992). There are numerous differences among these computational models, some of which have been discussed elsewhere (Chandler 2002; Daelemans et al. 1994; Shanks 1995). What is most important is what they have in common.

Consider the task at hand, which is to determine what allophone of /t/ should appear in a particular context. A database is needed that represents a sampling of a speaker's knowledge or experience with the allophones of /t/ which can be taken from natural language usage such as a corpus of utterances. For each entry in the database, a category variable specifies which of the allophones appears. Other variables could include information such as the phonetic, morphological, syntactic, pragmatic, and social context in which the allophone occurs. The algorithm's task is to take the variable vector as a test case and determine its similarity to the other vectors in the database. The determination of similarity is where the algorithms vary most radically from each other.

The particular model of analogy I use is Analogical Modeling of Language (abbreviated AM; Skousen 1989, 1992, 1995, 1998). AM makes its predictions on the basis of a given context, which is a vector of variables that represents linguistic information about the entity whose behavior is being predicted. The reader is referred to Skousen (1989, 1992) for a detailed treatment of the AM algorithm but a brief sketch of the model is in order. In the present study, the given context contains information about the context in which /t/ occurs. AM searches the database that represents the mental lexicon for database items that share variables with the given context. It then creates groups of database items with shared similarities called subcontexts. Variable vectors that have more in common with the given context will appear in more subcontexts. Subcontexts are further combined into more comprehensive groups called supracontexts. Upon inspection, some supracontexts will be homogenous, that is, the members agree or exhibit the same allophone of /t/ or they all share the same variable vector. Other supracontexts will have disagreements in that they

contain members with different allophones; these are said to be heterogeneous. By minimizing disagreements and eliminating members of heterogeneous supracontexts, database items belonging to the most clear-cut areas of contextual space (homogenous supracontexts) remain that are available to exert their influence on the choice of allophone for the given context. These make up the analogical set.

AM uses the members of the analogical set to calculate the probability that the given context will be assigned one of the allophones of /t/ found in the database. In general, what AM calculates is that the allophone in the database items that are most similar to the given context will predict the behavior of the given context, although the allophones of /t/ that appear in less similar database items have a small chance of applying as well provided that they appear in homogenous supracontexts. Allophony is determined in terms of a particular given context and no global characterization of the data is made. This implies that the variables which may be important in determining the allophone of /t/ in one given context may be not be important in determining the allophone in a different one (see Skousen 1995: 223–226 for an example).

The program gives the outcome in terms of the probability that each of the allophones will appear in a given context. For example, the probability may be five percent [t^h], 93 percent [r], zero percent [t⁼], one percent [ʔ], and one percent [t]. There are two ways in which this outcome may be interpreted (Skousen 1989: 82). The first, called selection by plurality, is used to determine the winner. Accordingly, the allophone with the highest predicted probability in the analogical set is applied to the given context, which in the above case is [r]. Random selection is the other method of interpreting the outcome. It uses the probabilities calculated by the algorithm that a particular allophone will appear in a given context. It essentially involves randomly selecting one of the members of the analogical set and applying the allophone of that member to the given context.² Members containing allophones that are more frequent in the analogical set have a higher probability of applying. Random selection reflects the sort of variability that occurs in actual language usage and that is hard to account for in formal approaches that predict one and only one outcome in a particular environment. In non-linguistic experiments children appear to utilize both random selection and selection by plurality when performing tasks involving probabilistic outcomes (Messick and Solley 1957).

The question naturally arises as to how closely this particular algorithm models the mental mechanisms speakers employ in the course of language production. A great deal of evidence exists that stored exemplars influence linguistic processing. Perhaps the most attractive part of an analogical

approach is its simplicity. It is based on the fairly uncontroversial idea that linguistic information is stored in the mind and retrieved as necessary. That groups of similar words can effect the behavior of other words with similar characteristics is well-attested in the psycholinguistic literature (e.g. Bybee and Slobin 1982; Stemberger and MacWhinney 1988). There is also ample evidence that behavior is based on stored exemplars (Chandler 1995; Eddington 2000; Hall 2005; Hintzman 1986, 1988; Hintzman and Ludlam 1980; Medin and Schaffer 1978; Murphy 2002; Nosofsky 1988; Schweitzer and Möbius 2004; Solé 2003). Analogical approaches are designed to model these effects. However, too little is known about the exact functioning of the brain to even begin to explain exactly how instances are stored, accessed, or categorized on the neural level. For this reason, it is impossible to conjecture about how faithfully AM or any other computer algorithm mirrors actual brain processes.

2.1.1. *Database for the simulations* The database used for the simulations was taken from the TIMIT corpus (Garofolo et al. 1993; Zue and Seneff 1996). TIMIT consists of a total of 6,300 utterances (2,342 different sentences) which were obtained by asking 630 speakers to read ten sentences each. Two of the sentences—called the dialect sentences—were identical for all of the speakers. Speakers were all native American English speakers.

From the time-aligned phonetic transcriptions of the training section of TIMIT, I searched the non-dialect sentences for instances of orthographic *t* that are traditionally thought to correspond to the phoneme /t/. Cases of *t* in words such as *catch*, *nation*, and *other* were, of course, not considered. Based on spectrographic analysis, the TIMIT transcription identifies four allophones of /t/: [r] as in *butter*, [ʔ] as in *gotten*, and deletion ([∅]) as in *percent of*. An unreleased [t⁼] as in utterance final *put* occurs when the spectrogram indicates stop closure but no release. The TIMIT transcription does not differentiate released and aspirated allophones, which is based on the gradient characteristic of voice onset time. Therefore, I determined a best guess estimate for a boundary between aspirated and unaspirated released allophones based on data by Davidsen-Nielsen (1974), Ladefoged (2005) and Lisker and Abramson (1964) in order to distinguish between these variants; accordingly, a VOT of 59ms or below was considered unaspirated, while VOTs of 60ms and above were marked as being aspirated.

Most of the 2,342 different sentences in the database containing /t/ were represented in the database, however, I avoided including the same sentence in the database twice, in particular the dialect sentences. The 3,719 resulting instances served as the database for the simulations. There

Table 1. Variables appearing in the database

Variables	
1	The phonetic realization of /t/.
2	The third phone or boundary to the left of /t/.
3	The second phone or boundary to the left of /t/.
4	The first phone or boundary to the left of /t/.
5	The first phone or boundary to the right of /t/.
6	The second phone or boundary to the right of /t/.
7	The third phone or boundary to the right of /t/.
8	The stress of the preceding syllable (primary, secondary, unstressed).
9	The stress of the following syllable (primary, secondary, unstressed).
10	The word /t/ appears in.

were 564 [ɾ], 234 [ʔ], 284 [ø], 629 [t], 860 [t^h], and 1,100 [t^h]. In addition, 48 instances of /t/ were voiced and much longer than a flap and were therefore transcribed as [d] in TIMIT (e.g., *carpenter*, *later*, and some words ending in *-ity*). In addition to the allophonic representations of /t/, the phonetic environment in which /t/ appeared was encoded (see Table 1). The three phones or boundaries to the right and left of the /t/ were identified and coded as variables. The boundary variables that could occupy one of the slots were either a phrase internal word boundary, a phrase internal pause, or a utterance initial or final pause/word boundary. When a sentence internal pause coincided with a word boundary a pause was coded. Three levels of stress were encoded from the syllable preceding and following /t/: no stress, primary stress, and secondary stress according to the TIMIT transcription. Lack of any phone or boundary in a given position was marked with a zero.

The syllable is such an integral part of most accounts of flapping, aspirating, and glottalizing that I was initially inclined to include it. However, determining boundaries is generally done on an a priori basis rather than on quantitative grounds. English speakers often differ in their intuitions about how to divide words into syllables (see Derwing 1992; Derwing and Neary 1991; Eddington et al., in print; Fallows 1981; Treiman et al. 2002; Treiman and Danis 1988; Treiman et al. 1992; Treiman et al. 1994; Treiman and Zukowski 1990; Zamuner and Ohala 1999). Because of the problems inherent in determining syllable boundaries, they were not included as variables. The encoding of the /t/ of *meet* as a flap in the sentence *I know I didn't meet her ... early enough* yields this variable vector:

- I. 1) ɾ, 2) word boundary, 3) m, 4) i, 5) word boundary, 6) ə, 7) pause 8) primary stress, 9) unstressed, 10) *meet*

Note that the tenth variable is the word in which /t/ appears. This is a way of including a rough degree of semantics and lexical identity into the representation.

This particular way of encoding the data was chosen to be as surface apparent as possible; it includes no overtly abstract features. Of course, it is not entirely devoid of abstractions because it incorporates categorical allophonic transcriptions rather than gradient formant frequencies, durations, information about articulatory gestures, etc. The use of stress is also an abstraction because the particular combination of duration, volume, and pitch that makes up stress in each instance is not directly represented. Nevertheless, these abstractions are not motivated by any theoretical preconceptions but are used because they were readily derivable from the TIMIT corpus in this form. The use of nominal linearly ordered variables is also a requirement in order for the data to be read by the computer algorithm. In sum, it is done for convenience and is not meant to support the notion that phonological processing involves symbolic manipulation of segments, although it does recognize that orthography may impose such a view to some degree. In cognitive terms, one exemplar could be thought to encode the semantic, acoustic, and sensory motor information of many instances that are similar enough to be considered essentially the same thing (Bybee 1994). Corpora that indicate detailed phonetic and gestural information will surely overcome this limitation on variables in future studies.

The TIMIT transcription was followed closely except in a few instances. [h] and [ɦ] were merged, as were fronted and non-fronted variants of [uw]. In TIMIT, some allophones were given different representations depending on whether they were stressed or not, therefore, I collapsed stressed and unstressed [ju], stressed and unstressed [ə], and stressed schwa, unstressed schwa, voiceless schwa and [ʌ] (which in American English differs little from schwa except for its stress). The selection of these variables to the exclusion of others does not mean that other factors do not influence allophonic distribution. Nor should the fact that some variables were excluded be taken to indicate that analogy would be incapable of handling them. The focus of the paper is not to evaluate sociolinguistic differences. Instead, the goal is to demonstrate that analogy can account for allophonic distribution by considering the phonetic and morphological context in which /t/ occurs.

2.1.2. *Simulations* For the purposes of the simulations, the data set described above is considered a subset of a speaker's linguistic experience with the phoneme /t/ and its allophones. In the simulations, each exemplar in the data set is removed one at a time and serves as the test case.

The allophone of /t/ for that exemplar is then determined based on analogy to the remaining exemplars according to AM's algorithm.

2.1.2.1. *Exact matches and perfect memory* The first simulations allowed access to all of the exemplars in the data set and also permitted exact matches in the database to influence the test item. In cognitive terms this represents lexical access to previously experienced exemplars. This simulation probably reflects what happens most often when producing a familiar word in a previously experienced context. If most previously experienced instances of *pretty* have a flap allophone then [pɹɪrɪ] will be produced. Under these circumstances, 99.7 percent of the test cases are correctly predicted. The success rate does not attain 100 percent accuracy due to the fact that some exemplars have /t/ in identical contexts, but with a different allophone. The influence of exact matches is evident when the possibility of their occurrence is diminished. Figure 1 summarizes the results of ten sets of simulations.

In the first, a random 10 percent sampling of the data set was chosen to analogize on. This was repeated ten times with a new 10 percent sample each time. The average success rate for the 10 runs is given for each of the allophones. The second set of simulations utilized random samplings of 20 percent of the data set, and so forth until the entire data set was available.

The ability of analogy to account for almost all allophony when exact matches in perfect memory are allowed reflects the idea that speakers know how a word is pronounced due to access to past experience, however, it is hardly a surprising finding. However, traditional phonological

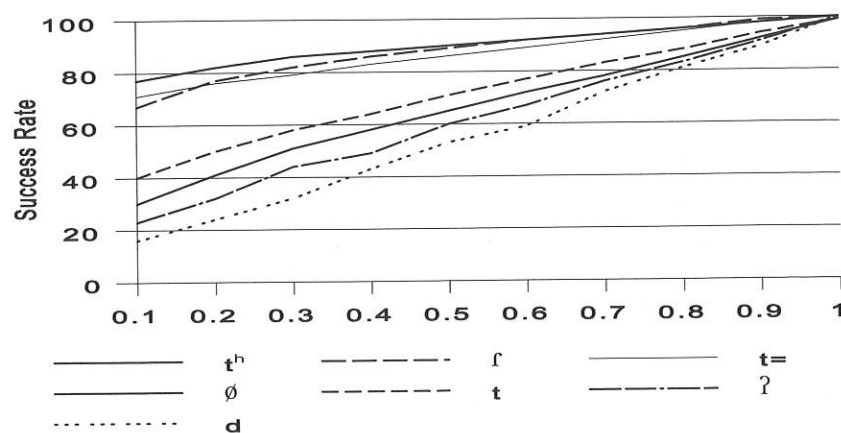


Figure 1. Success rates by allophone with varying percentages of the data set utilized

analysis takes one of two positions either tacitly or overtly. One stance is that the particular allophone of /t/ in a word or phrase is not stored in memory but is determined on the basis of a generalization of some sort (e.g., the /t/ in *seat* is glottalized because it follows a sonorant and appears utterance finally). The second holds that detailed pronunciation information may be stored in memory, but that a generalization about the allophonic distribution, separate from the stored instances themselves, is used to determine the pronunciation of a previously unexperienced item. Analogy, on the other hand, assumes that exemplars are stored with detailed phonetic information, but that no generalization about allophonic distribution is arrived at and stored as an entity separate from the exemplars themselves.³

2.1.2.2. *Benchmark simulation* Another way of testing analogy is to measure its success rate when the test items are treated as if they are previously unknown. This simulation was done to calculate a benchmark against which the remainder of the simulations could be compared. Each vector of variables representing an instance of /t/ was removed from the database and its allophone was predicted based on analogy to the remaining database items, however, exact matches were not allowed. Under these conditions, 65.3 percent of the items were predicted to have the same allophone of /t/ that appeared in the data set. Keep in mind that from a processing standpoint this simulation is less psychologically plausible because the underlying assumption is that the speaker has never heard or seen the word before and must determine its pronunciation strictly by analogy to other stored lexical items.

The success rate by allophone appears in the confusion matrix in Table 2 with the total number of instances in parentheses. The allophones [t^h], [r], and [t⁼] are the most successfully predicted, while [t], [ʔ] and [d] are the most difficult to predict. However, the general trends in the errors are quite interesting. For instance, many cases of deletion were predicted to appear with [t⁼] or [t^h] instead, yet both of the errors are quite plausible

Table 2. Confusion matrix of the predicted outcomes

Predicted outcome →	t ^h	r	t ⁼	ø	t	ʔ	d
t ^h (969)	83	3	4	1	8	0	0
r (564)	5	83	9	1	2	0	0
t ⁼ (860)	4	5	75	3	7	5	0
ø (284)	14	5	29	32	13	6	0
t (760)	24	8	20	4	43	1	0
ʔ (234)	4	14	56	6	2	18	0
d (48)	31	33	15	0	10	6	4

Table 3. Examples of predicted outcomes in terms of predicted probability

	t ^h	r	t ⁼	ø	t	ʔ	d
<i>site with</i>	2	0	46	0	<u>11</u>	41	0
<i>hiding out like</i>	0	7	20	10	10	<u>53</u>	0
<i>discount price</i>	4	0	63	<u>33</u>	0	0	0
<i>pronoun it carries</i>	12	2	<u>70</u>	9	5	2	0
<i>importance</i>	<u>2</u>	0	0	19	0	79	0
<i>bitter unresolving</i>	0	<u>98</u>	0	1	0	0	0
<i>dogmatically</i>	0	<u>76</u>	0	6	17	0	0

outcomes; *chest* can be [tʃes] as easily as [tʃes⁼]. The final /t/ in *comment on* can be deleted or realized as an aspirate or unreleased stop. In the same vein, it should not be surprising that [ʔ] is most often predicted to be [t⁼]; they are often interchangeable in the same context in words such as *Vietnam*, *nightmare*, and *light*. Additionally, [ʔ] and [t⁼] are acoustically quite similar in that they both result in only minor formant transitions of surrounding vowels (Silverman 2004: 170). Recall that in order for a phone to be coded as aspirated it needed to have a VOT of 60ms or greater. This somewhat arbitrary cutoff point may be partially responsible for many of the instances of [t] that were predicted to be [t^h] as well.

A number of specific outcomes are given in Table 3. The actual pronunciation is indicated with an underlined predicted probability. When this number is not the highest the instance is counted as an error. However, as is common with many such errors, the other highly predicted outcomes are often viable alternative pronunciations.

According to those who consider rules and constraints to be psychologically real, these entities are learned and manipulated subconsciously, which is why they cannot be overtly expressed. A linguistically naive speaker cannot describe the rule he or she uses to determine that the nonce word *chowty* would contain a flap but *catino* would not. It just sounds right. In the approach adopted here, this occurs because no tacit rule exists. *Chowty* sounds correct with a flap because there are many similar words in the mental lexicon with a flap.

2.1.2.3. *Interpretation of the success rate* The success rate of 65 percent may not initially appear very compelling, but a number of things need to be considered before passing judgment. First, the simulation was run to mimic what happens when a speaker cannot access previously stored instances of the word in a specific phonetic context from the mental lexicon which is probably not a common state-of-affairs. Second, there is a great

deal of variability in the allophone a speaker may select. Utterance final /t/ in *bat* may be realized as [t⁼], [ʔ], or [t], all of which are possible in fluent, native, American English speech. If [t⁼] is calculated to have the highest predicted probability and the success rate is determined by using selection by plurality (Section 2.1) the other outcomes would be deemed incorrect. Likewise, medial /t/ in *important* appears in the database with a flap, a glottal stop, and in some cases deleted in spite of the fact that it appears in the same phonetic context. The model predicts only one of these to be the most prevalent. Third, because I made no effort to sanitize the data by eliminating such overlapping cases, they are counted as errors.

Variability is also manifest in the database in other ways. For instance, *little* and *positive* were generally realized with a flap, yet one instance of deletion of /t/ appears in the database for both of these words. In general, aspiration of /t/ following a word initial /s/ is short enough that the /t/ is considered unaspirated (59 ms or lower). In theory, the aspirate should not occur in these contexts. In actual speech, however, there are outliers. In one instance of the word *still*, for example, the VOT after /t/ is 88 milliseconds. All such cases were coded as having an aspirate allophone. When the model predicted them to be unaspirated they were counted as errors. The data contain many instances of these kinds of variable pronunciations and the success rate of a formal model would encounter difficulties accounting for them as well.

It would seem reasonable to compare the resulting 65 percent success rate with that achieved by a rule approach applied to the same database. However, there are many things which make such a comparison difficult. Different rules have been devised, so which should be chosen? In order to be fair and impartial all accounts would need to be tested which would result in an extremely lengthy and tedious report. Unfortunately, (or fortunately) applying all or any of the formal approaches to the items in the database is impossible. First, some rules only deal with a few of the allophones of /t/ rather than the six predicted in the simulation. Second, traditional rules assume non-overlapping contexts which makes them incompatible with the considerable overlap evident in the database. Third, none of rules specify phonetic contexts that are purely surface apparent, but make use of abstract features, assumptions about syllable division, rule ordering, etc. Manipulation of abstract entities, (or at least controversial ones in the case of syllable breaks), allows one to account for anything; if two different allophones appear in exactly the same surface apparent context it could be claimed that it is due to a differing syllabifications, rule orderings, or values of an abstract feature, yet these entities are not observable and therefore not amenable to empirical analysis.

It could be possible to analyze the database itself in order to discover a general set of surface apparent contexts for each allophone. However, this would be problematic because the allophones are not in strict complementary distribution. For instance, the most general context for both [t⁼] and [ʔ] is that they appear preceded by a vowel or sonorant and followed by a consonant or word boundary. The second problem involves deciding how many contexts to describe for each allophone. Should only the single most general context be considered? If multiple contexts are described, how many instances should each context account for? 100? 50? 10? 2? At what point would a particular context no longer represent a valid generalization? Clearly, low success rates will occur when only one or two very general contextual statements are included. The greater the number of rule contexts allowed the higher the success rate will be.

When taken to the extreme a model such as Albright and Hayes (1999) would emerge in which all, often thousands of possible rules are generated from an inspection of the data. While such a model may achieve high success rates it lack in terms of learnability and psychological plausibility. Even rules based on surface properties allow a great deal of flexibility in deciding what is a true generalization and should be allowed. For these reasons, there is no way to objectively compare the success rates of rules based on observable contexts and the rates obtained from the simulations.

Nevertheless, it is possible to get a sense of the generalizations that exist in the data. To this end, I considered the boundary or phone before /t/ and the two that follow it. In order to make broad generalizations the variables were collapsed into vowel, sonorant, and obstruent (V, S, Ob) and the boundaries into phrase medial word boundary (#) versus phrase internal pauses and utterance final and initially boundaries (0). In addition, whether the syllable following the /t/ had either primary or secondary stress (1) versus no stress (0) was encoded. This yielded 148 attested combinations of these factors. The purpose of this was to find what general context each allophone appears in, therefore, I chose an admittedly arbitrary cut-off point of twenty instances in a cell as indicative of a generalization. All such cases appear in Table 4.

These generalizations account for 58.9 percent of the data. Of course, setting the cut-off point below 20 would result in a greater coverage of the data, but at the expense of more contexts per allophone which is already quite high, at least for a traditional phonological analysis. For example, there are ten contexts for [t^h] at the 20 instance cut-off. Combining all of these in a way that would be considered kosher in a rule-based framework would be a seemingly impossible task. One reason these

Table 4. General contexts for each allophone

	[ø]	[d]	[r]	[ʔ]	[t]	[t ^h]	[t ⁼]
#_S V 1						74	
#_V # 1					69	175	
#_V Ob 1						111	
#_V S 1						96	
0_V # 1						25	
Ob_# Ob 1	26						29
Ob_0 0 0						26	
Ob_S V 1					22	30	
Ob_V Ob 0					51		
Ob_V Ob 1					54		
Ob_V S 0					24		
Ob_V S 1					66		
S_# Ob 1							75
S_V Ob 0	23						
S_V S 1						48	
V_# Ob 0							56
V_# Ob 1				29	42		256
V_# S 1				70			61
V_# V 0			54				
V_# V 1			114				
V_0 0 0							59
V_Ob # 1					25		52
V_S # 0			36				
V_S Ob 0			21				
V_V # 0			40				
V_V Ob 0			117				
V_V Ob 1						46	
V_V S 0			51				
V_V S 1						39	
Total	49	0	433	99	353	670	588
% Predicted	17	0	77	42	56	61	68

contexts cannot be thought of as rules in the traditional sense is that many of their contexts overlap. For instance, the context V_# Ob 1 as in *at many* is shared by [ʔ], [t], and [t⁼].

As already mentioned, the ability to tweak the cut-off point in this analysis makes it possible to obtain a range of data coverage which is why it does not seem to offer an objective point of comparison for the overall success rate of the simulation. However, it does provide a rough measure of how many errors are actually possible alternative pronunciations. For example, in the context V_Ob # 1 (e.g., *chestnuts are*) there are 25 cases of /t/ and 52 of /t⁼/. Therefore, any instances of /t/ in this context that are predicted to be /t⁼/, (and vice-versa), could legitimately

be considered alternative pronunciations and not true errors. By applying this criterion to the overlapping generalizations in Table 4, 316 of the errors would be considered alternative pronunciations and the success rate of the benchmark simulation could be raised to 73.8 percent. Although it is admittedly a subjective measure, I examined the errors made in the benchmark simulation and divided them into those that according to my speech are plausible alternative pronunciations (e.g., *orien*[t^h]*jed* / *orien*[ø]*ed*, *firs*[t] *one* / *firs*[ø] *one*) and those that were odd pronunciations (e.g., *fros*[t^h]*bite*, *motiva*[r]*es*). According to this measure, the success rate is 91.3 percent.

There is another more objective way of calculating how significant the 65 percent success rate is. A purely random prediction of one of the six allophones would only achieve a 17 percent rate of success, while predicting the most frequent aspirated allophone in all cases would yield a 30 percent rate. A weighted prediction, based on the percentage of each allophone in the database would only correctly predict 19 percent of the cases. Therefore, the 65 percent success rate is significantly higher than the best chance rate of 30 percent ($X^2(1) = 929.14$, $p < 0.001$). Keep in mind that the 65 percent rate is attained by disallowing access to previously encountered instances. As seen in Section 2.1.2.1, a much higher rate of success is obtained when exact matches are permitted, which is a more plausible state of affairs in actual linguistic processing.

2.1.2.4. *Simulations with other variables* In the benchmark simulation, all 3,719 items were available as analogs. One question that arises is whether analogy is robust enough to make good predictions without consulting so many instances. An estimation of how many items need to be considered is easily obtained by limiting the number of database items used in the simulation. To this end, I ran ten sets of simulations in which exact matches were disallowed. In the first set, a random 10 percent of the database (about 372 items) was consulted. This was repeated ten times with a new sample drawn at random each time. The next set of simulations utilized 20 percent of the database, then 30 percent, and so on. The average success rate of each set of simulations appears in Figure 2.

The success rate obtained by consulting only about 372 items (55.4%) is only about 10 percent lower than the 65.3 percent rate that occurs when analogs are drawn from all 3,719 items. In other words, fairly good predictions are possible by using only a small subset of the storehouse of linguistic experience contained in the mental lexicon, although some improvement occurs as more items are consulted. In other words, the system of allophonic distribution of /t/ in English may be extracted from any random sampling of several hundred to a thousand instances even when

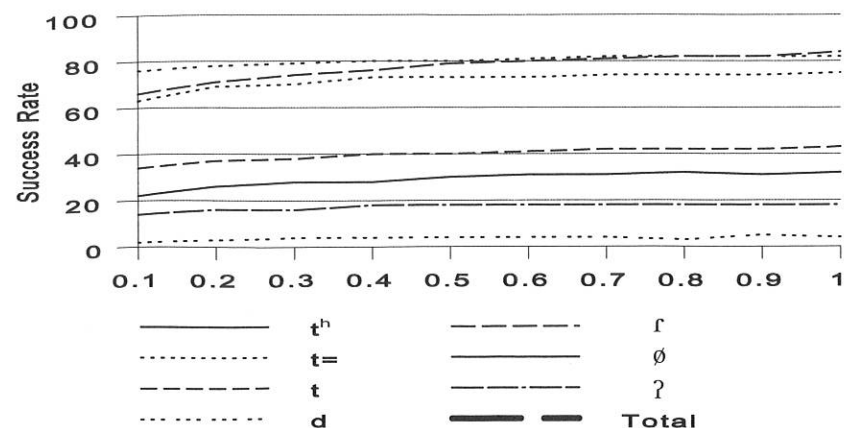


Figure 2. Success rates with differing data set sizes

the assumption is made that all instances of /t/ are new and previously unexperienced. As Figure 1 demonstrates, the same is true when access to previously experienced words is allowed resulting in much higher success rates.

In rule accounts of allophonic distribution emphasis is on finding the specific conditions that are necessary and sufficient to determine which allophone is called for. Besides abstract entities and mechanisms most rules for /t/ invoke stress as a key deciding factor. How does analogy fare without this crucial component? To determine this, another simulation was run without the stress variables and the success rate dropped from the benchmark of 65.3 percent to 62.7 percent which is significant ($X^2(1) = 5.49$, $p < 0.025$). However, elimination of this variable hardly deals a catastrophic blow to the ability of analogy to account for the data. In contrast, most rule approaches critically depend on stress and would simply be unable to make any sort of predictions without it.

Formal accounts of /t/ do not consider the effect that the phones or boundaries two and three slots to the left nor three slots to the right of /t/. They are generally not considered important, yet when these four variables are removed 63.9 percent of the items are still correctly predicted. That is, removing supposedly unimportant variables such as these results in a slight drop in success rate (1.4%) roughly similar to that obtained when eliminating the supposedly fundamental variables that encode stress (2.6% drop). The reason for eliminating variables in order to measure their effect is to demonstrate that analogy does not depend on a few decisive variables. Its predictions remain robust even when crucial variables are removed or a smaller data sets are drawn from.

3. Conclusions

The primary purpose of this paper is to present an account of how speakers may process allophonic distribution. When it is viewed as due to analogy to past linguistic experience the necessity of abstract entities and mechanisms is eliminated as is the need in language acquisition for subconsciously formulating generalizations. Unlike formal approaches that require necessary and sufficient conditions for each allophone, analogy proves to be extremely robust; it continues to make solid predictions even when only a fraction of the items in the simulated mental lexicon are consulted, and does not suffer catastrophic breakdown when variables are removed, even a variable such as stress which is generally thought to be an essential component of most accounts of the distribution of the allophones of /t/ in English. The types of errors predicted by the model are also of interest because they are generally plausible alternative pronunciations.

Received 19 December 2005

Brigham Young University

Revision received 15 May 2006

Notes

* I express my thanks to Kyle Woods and Stephen Mouritsen for their help in preparing the database and also to Royal Skousen, Dirk Elzinga, Andy Wedel, Steve Chandler, and the anonymous reviewers for their helpful comments on the manuscript. Contact address: Department of Linguistics and English Language, 4064 JFSB, Provo, Utah 84602, USA; Author's e-mail: <eddington@byu.edu>.

1. See Silverman (2004) and Turk (1992) for treatments of this issue.
2. Actually, one of the pointers in the analogical set is chosen, but the role of pointers in the algorithm has not been discussed in this summary description.
3. It is possible that the analogical set (Section 2.1) is stored. This is similar to assuming that a group of lexical items is related because they have many different associational connections between them.

References

- Aha, David W., Dennis Kibler, and Marc K. Albert
1991 Instance-based learning algorithms. *Machine Learning* 6, 37–66.
- Albright, Adam and Bruce Hayes
1999 An automated learner for phonology and morphology. Unpublished manuscript. www.linguistics.ucla.edu/people/hayes/learning/learner.pdf
- Baudouin de Courtenay, Jan
1895 *Versuch einer Theorie phonetischer Alternationen: Ein Kapitel aus der Psychophonetik*. Strassburg/Crakow: Karl Trübner.
- Beckman, Mary E. and Janet Pierrehumbert
2003 Interpreting phonetic interpretation of the lexicon. In Local, John, Richard Ogden, and Rosalind Temple (eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 13–37.
- Benua, Laura
1995 Identity effects in morphological truncation. *Papers in Optimality Theory. University of Massachusetts Occasional Papers in Linguistics* 18, 77–136.
- Boomershine, Amanda
2006 Perceiving and processing dialectal variation in Spanish: An exemplar theory approach. In Face, Timothy L. and Carol A. Klee, (eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, 58–72.
- Bromberg, Sylvain, and Morris Halle
2000 The ontology of phonology (revised). In Burton-Roberts, Noel, Philip Carr, and Gerard Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford: Oxford University Press, 19–37.
- Brown, R. and D. Mc Neill
1966 The ‘tip of the tongue’ phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5, 325–337.
- Burton, Peter G.
1990 A search for explanation of the brain and learning: Elements of the synchronic interface between psychology and neurophysiology. *Psychobiology* 18, 119–161.
- Burzio, Luigi
1996 Surface constraints versus underlying representations. In Durand, J. and B. Laks (eds.), *Current Trends in Phonology: Models and Methods*. Salford: European Studies Research Institute, University of Salford Publications, 97–112.
- Bybee, Joan L.
1994 A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics* 5, 285–305.
1999 Usage-based phonology. In Darnell, Michael, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics*. Amsterdam: John Benjamins, 211–242.
2001 *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, Joan L. and Dan I. Slobin
1982 Rules and schemas in the development and use of the English past tense. *Language* 58, 265–289.
- Byrd, Dani
1994 Relations of sex and dialect to reduction. *Speech Communication* 15, 39–54.
- Chandler, Steve
1995 Non-declarative linguistics: Some neuropsychological perspectives. *Rivista di Linguistica* 7, 233–247.
2002 Skousen’s analogical approach as an exemplar-based model of categorization. In Skousen, Royal, Deryle Lonsdale, and Dilworth B. Parkinson (eds.), *Analogical Modeling: An Exemplar-Based Approach to Language*. Amsterdam: John Benjamins, 51–105.
- Coleman, John
2002 Phonetic representations in the mental lexicon. In Durand, J. B. Laks (eds.), *Phonetics, Phonology, and Cognition*. Oxford University Press, 96–130.
- Connine, Cynthia M.
2004 It’s not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11, 1084–1089.

- Daelemans, Walter, Steven Gillis, and Gert Durieux
1994 Skousen's analogical modeling algorithm: A comparison with lazy learning. In Jones, D. (ed.), *Proceedings of the International Conference on New Methods in Language Processing*. Manchester: UMIST, 1–7.
- Daelemans, Walter, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch
2001 *TiMBL: Tilburg Memory-based Learner, Version 4.1 Reference Guide: Induction of Linguistic Knowledge Technical Report 01-04*. Tilburg: ILK Research Group, Tilburg University.
- Davidsen-Nielsen, Niels
1974 Syllabification in English words with medial *sp*, *st*, *sk*. *Journal of Phonetics* 2, 15–45.
- Davis, Stuart
2003 Capitalistic vs. militaristic: The paradigm uniformity effect reconsidered. In Downing, Laura J., T. A. Hall, and Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press, 107–121.
- De Jong, Kenneth
1998 Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics* 26, 283–310.
- Derwing, Bruce L.
1992 A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for five diverse languages. *Language and Speech* 35, 219–235.
- Derwing, Bruce L., and Terrance M. Neary
1991 The 'vowel-stickiness' phenomenon: Three experimental sources of evidence. *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 3. Aix-en-Provence: Publications de l'Université de Provence, 210–213.
- Eddington, David
2000 Spanish stress assignment within the analogical modeling of language. *Language* 76, 92–109.
- Eddington, David, Dirk Elzinga, Rebecca Treiman, and Mark Davies
In print The syllabification of American English. Manuscript, Brigham Young University.
- Egido, Carmen and William E. Cooper
1980 Blocking of alveolar flapping in speech production: The role of syntactic boundaries and deletion sites. *Journal of Phonetics* 8, 175–184.
- Fallow, Deborah
1981 Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17, 309–317.
- Fougeron, Cécile and Donca Steriade
1997 Does deletion of French schwa lead to neutralization of lexical distinctions? In Kokkinakis, G., N. Fakotakis, and E. Dermatas (eds.), *Proceedings of the 5th European Conference on Speech Communication and Technology*, vol. 7. Rhodes: University of Patras, 943–937.
- Garofolo, John S., Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, and Nancy L. Dahlgren
1993 *DARPA TIMIT: Acoustic-phonetic Continuous Speech Corpus*. Philadelphia: Linguistic Data Consortium.
- Giegerich, Heinz J.
1992 *English Phonology: An Introduction*. Cambridge: Cambridge University Press.

- Goldinger, Stephen D.
1997 Words and voices: Perception and production in an episodic lexicon. In Johnson, Keith and John W. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego, CA: Academic Press, 33–65.
- Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier, and Daniel Jurafsky
1999 The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society* 35, 151–166.
- Hall, Kathleen Currie
2005 Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. In Alderete, John, Chung-hye Han and Alexei Kochetov (eds.), *Proceedings of the Twenty Fourth West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadia Press, 191–199.
- Harris, John
1994 *English Sound Structure*. Oxford: Blackwell.
- Hawkins, Sarah
2003 Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31, 373–405.
- Hintzman, Douglas L.
1986 Schema abstraction in a multiple-trace memory model. *Psychological Review* 93, 411–428.
1988 Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95, 528–551.
- Hintzman, Douglas L., and Genevieve Ludlam
1980 Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory and Cognition* 8, 378–382.
- Jensen, John T.
1993 *English Phonology*. Amsterdam/Philadelphia: John Benjamins.
- Kahn, Daniel
1980 *Syllable-based Generalizations in English Phonology*. New York: Garland.
- Kenstowicz, M.
1996 Base-identity constraints and uniform exponence: Alternatives to cyclicity. In Durand, J. and B. Laks (eds.), *Current Trends in Phonology: Models and Methods*. Salford: European Studies Research Institute, University of Salford Publications, 363–393.
- Kiparsky, Paul
1979 Metrical structure assignment is cyclic. *Linguistic Inquiry* 10, 421–441.
- Kolers, Paul A. and Henry L. Roediger
1984 Procedures of mind. *Journal of Verbal Learning and Verbal Behavior* 23, 425–449.
- Labov, William
1994 *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.
- Ladefoged, Peter
2005 *Vowels and Consonants*, 2nd ed. Malden, MA: Blackwell.
- Laeuffer, Christiane
1989 French linking, English flapping, and the relation between syntax and phonology. In Kirschner, Carl and Janet Decesaris (eds.), *Studies in Romance Linguistics*. Amsterdam: John Benjamins, 225–247.

- Stemberger, Joseph Paul and Brian MacWhinney
 1988 Are inflected forms stored in the lexicon? In Hammond, Michael and Michael Noonan (eds.), *Theoretical Approaches to Morphology*. San Diego: Academic Press, 101–116.
- Steriade, Donca
 1997 *Lexical Conservatism and its Analysis*. Manuscript, University of California, Los Angeles.
 1999 Lexical conservatism. In Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, vol. 4. Hanshin: Linguistic Society of Korea, 157–180.
 2000 Paradigm uniformity and the phonetics-phonology boundary. In Broe, Michael and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology 5*. Cambridge: Cambridge University Press, 313–334.
- Strassel, Stephanie
 1998 Variation in American English flapping. In Paradis, Claude, Diane Vincent, Denise Deshaies, and Marty Laforest (eds.), *Papers in Sociolinguistics: NWAV-26 à l'Université Laval*. Quebec: Nota Bene, 125–135.
- Treiman, Rebecca, Judith A. Bowey, and Derrick Bourassa
 2002 Segmentation of spoken words into syllables by English-speaking children. *Journal of Experimental Child Psychology* 83, 213–238.
- Treiman, Rebecca and Catalina Danis
 1988 Syllabification of intervocalic consonants. *Journal of Memory and Language* 27, 87–104.
- Treiman, Rebecca, Jennifer Gross, and Annemarie Cwikiel-Glavin
 1992 The syllabification of /s/ clusters in English. *Journal of Phonetics* 20, 383–402.
- Treiman, Rebecca, Kathleen Staub, and Patrick Lavery
 1994 Syllabification of bisyllabic nonwords: Evidence from short-term memory. *Language and Speech* 37, 45–60.
- Treiman, Rebecca and Andrea Zukowski
 1990 Toward an understanding of English syllabification. *Journal of Memory and Language* 29, 66–85.
- Turk, Alice
 1992 The American English flapping rule and the effect of stress on stop consonant durations. *Working Papers of the Cornell Phonetics Laboratory* 7, 103–133.
- Zamuner, Tania S., and Diane K. Ohala
 1999 Pre-literate children's syllabification of intervocalic consonants. In Greenhill, Annabell, Heather Littlefield, and Cheryl Tano (eds.), *Proceedings of the 23rd Annual Boston Conference on Language Development*. Somerville, MA: Cascadilla Press, 753–763.
- Zue, Victor W. and Martha Laferriere
 1979 Acoustic study of medial /t, d/ in American English. *Journal of the Acoustic Society of America* 66, 1038–1050.
- Zue, Victor and Stephanie Seneff
 1996 Transcription and alignment of the TIMIT database. In Fujisaki, Hiroya (ed.), *Recent Research Toward Advanced Man-machine Interface Through Spoken Language*. Amsterdam: Elsevier, 464–447.