Ohala, J.J. (1981). Speech timing as a tool in phonology. Phonetica 38, pp. 204-212.

Phonetica *38:* 204–212 (1981)

Speech Timing as a Tool in Phonology

JOHN J. OHALA

Phonology Laboratory, Department of Linguistics, University of California, Berkeley, Calif., USA

Abstract. Phonological questions of the sort 'Is segment or feature [x] a surface-phonetic event or a property of the underlying mental representation?' may be answered in some cases by considering the durations of the speech sounds in the word in which [x] appears. Here the status of the epenthetic stop [p] in words such as *teamster* [-tʰimpstɚ] was clarified by measurement of the duration of the preceding vowel + nasal (VN) sequence: a shortened VN implies the [p] is underlying; a long VN implies that it is a surface-phonetic event.

Introduction

It is quite an old observation that the source of widespread, so-called 'natural' sound changes is to be found within all human speakers, past and present. Thus it is possible to examine the speech of a present-day speaker of English and find there the same 'seeds' that led to a major tonal split in Middle Chinese several centuries ago: the Middle Chinese tones developed higher and lower variants depending on whether the syllable-initial consonant was voiceless or voiced, respectively; similarly, English speakers also show higher and lower pitch on vowels following voiceless and voiced obstruents [HOMBERT et al., 1979]. Of course, although the pitch pattern evident in English (and every other language investigated on this point) parallels the Chinese sound change in many respects, e.g., the segment types involved, the direction of the variation, etc.; they are not *entirely* the same. In Chinese (in most dialects) the voicing contrast was lost in the process of the sound change that led to the tonal split – the tonal variation thereafter was not tied to the voicing of the syllable-initial consonant; in English, however, the pitch differences on the vowels are mechanically linked to the voicing of the preceding obstruent. In Chinese the speaker actively programs the pitch difference via changes in laryngeal tension; there is no evidence that the English speaker has to actively vary the tension of the vocal cords to achieve the observed pitch differences, rather, they are apparently unintentional by-products of the voicing difference. The crucial difference, then, is that the immediate source of the sound pattern, viewed synchronically, is the underlying mental lexicon of the speaker in the case of Chinese (henceforth, the 'underlying level'), whereas it still stems from a purely peripheral vocal tract constraint in the English speakers (henceforth, the 'surface level' or 'physical phonetic level')[1].

In cases such as the consonantally induced pitch differences, it is rather easy to differentiate the sound change from the purely phonetic process that exists only at the surface level since, in the case of the former, the conditioning environment was lost. But if the conditioning environment is not lost, it may not be very easy to tell the two types of sound patterns apart.

Epenthetic Stops: Underlying or Surface?

One such ambiguous case is that which involves the so-called epenthetic stops which appear between nasals and following obstruents [Passy, 1890; Grandgent, 1896; Millardet, 1911; Ohala, 1972, 1974, 1975]. To consider examples from English, we find on the one hand cases in which these stops are clearly fossils, i.e., independent, lexically specified aspects of the pronunciation, since even the standard spelling of the words includes them, e.g., *Thompson (<Thom + son)*, *dempster (<deem + ster)*, *Alhambra* (*<al hamra* 'the red [one]', Arabic), *thunder (<thunor)*. But, on the other hand, there are numerous cases of epenthetic stops in words where these stops are not represented in the spelling, e.g., *something* ['sʌmpθɪŋ], *teamster* ['tʰimpstɚ], *Chomsky* ['tʃɒmp-ski], *dance* [dænts], *length* [lɛŋkθ], *youngster* ['jʌŋkstɚ]. In some of these words the stop seems to be quite an essential part of the word and, indeed, Kenyon and Knott [1944] in their pronouncing dictionary

[1] My use of the terms 'underlying' and 'surface' deviates slightly from standard generative usage. My 'surface' level is that stage of the speech production or perception mechanisms which is human-universal, i.e., all the anatomical and neuroanatomical 'hardware' serving speech. My 'underlying' level is everything further 'upstream', in particular, those cognitive structures that can be modified by learning and experience.

record variants with the stop included. In other cases, the stop seems intuitively to be unintentional, a low-level transitional element. (As a purely phonetic entitity the stop arises due to the partial denasalization of the nasal consonant as the configuration of the velum anticipates that required for the following obstruent [see OHALA, 1972, 1974, 1975]). How can we tell whether such stops are present at the underlying or the surface level? [Cf. also the discussion of this problem by HARMS, 1973].

Word games could provide the answer [CHAO, 1934; SHERZER, 1970; HOMBERT, 1973, 1976]. A word game which would rearrange the parts of the word in such a way as to destroy the environment producing the epenthetic stops could reveal whether the stop was purely a surface phenomenon triggered by the phonetic environment or whether it was an independent underlying segment. For example, SHERZER [1970] used a Cuna word game (first syllable shifted to the end of the word) to clarify the status of the medial vowel in a word transcribed phonetically as [bíriga] 'year'. The word was anomalous because it deviated from the otherwise consistent pattern of penultimate stress placement. When restructured by the word game, however, it became [gabir] not * [rigabi]. The medial vowel was apparently not present at the underlying level, but rather was only an epenthetic surface vowel.

If such a word game existed in English (or if it could be introduced), it could be used in a similar way to reveal the status of the epenthetic stops in *teamster* and *youngster*, e.g., would *youngster* become [stɚjʌŋk] or [stɚjʌŋ]? If the former, then we might conclude that the [k] was in some sense an underlying /k/; if the latter, then it could be judged as just a physical phonetic transitional element. Although such evidence would be revealing, it could be subject to many extraneous distortions, e.g., subjects' knowledge of spelling, their knowledge of the constituent morphemes, etc. Of course, it might be possible to utilize word games or other tests using sufficient controls to eliminate confounding influences [see OHALA, in press]. I propose, however, that in this case (and some others) measurements of segment durations may provide us with a more direct answer to our question [see, in addition, LEHISTE, 1970, 1975; HERBERT, 1975].

There have been many phonetic studies showing that the duration of certain speech segments are influenced by the structure of the syllable, the word, and the phrase in which they appear, as well as by their inherent phonetic chatacter [HOUSE, 1961; LEHISTE, 1970; LIND-

BLOM and RAPP, 1972]. If our phonological question can be stated as 'does this syllable (word, etc.) have some specified underlying structure?' and if we can assume that timing of the parts of this structure are determined at the underlying level, then the durational measures of these parts will be helpful to us. The question regarding the character of the epenthetic stops meets these criteria. In words such as *teamster* and *youngster* we can rephrase our question to 'Is the first syllable closed with a voiceless stop ([tʰimp-], [jʌŋk-]) or a nasal ([tʰim-],[jʌŋ-])?[2]. If the former, then the epenthetic stops are underlying; if the latter, then they are just surface elements.

There is considerable evidence that in American English a stressed vowel and any following nonobstruent continuants are markedly shortened when they appear before a tautosyllabic voiceless obstruent [MALÉCOT, 1960; HOUSE, 1961; LEHISTE, 1970; BARNWELL, 1971; LOVINS, 1978]. Therefore, to determine if the stop is present at the underlying level (where we assume timing relations are determined), the *vowel* + *nasal* (VN) sequence should be measured to see if it is very short. To test the viability of this reasoning, the following 'calibration' of the technique was done.

Study 1

Methods

24 native speakers of American English, all of them students at the University of California, Berkeley (but not linguistics or language majors), were recorded during individual interviews whose purpose, they were told, was to elicit their reaction to some new English words [cf. OHALA, 1974]. They were told that one way new words were formed was by adding suffixes to existing words, e.g., "'hood' added to 'cousin' gives the new word 'cousinhood'". Subjects (Ss) were orally presented with several pairs of suffixes and stems and were asked to pronounce the new word that would result from the addition of the suffix to the stem. They were also asked to define the new word and to report whether they thought it would be used. In fact, only the pronounciation was of interest to the examiner (E), but the other questions lent credibility to the task. The E always pronounced the suffix before the stem in order not to inadvertently suggest to Ss some particular pronunciation of the combination.

At the beginning of the session the new word *clam* + *ster* was elicited. The last new word elicited was *clamp* + *ster*. (The intervening words were of no interest to this study and can be regarded as 'filler' words.) It was hoped that the word *clamster* would, in at least some cases, show an epenthetic [p] which it could be assumed would be a surface and not an underlying /p/ (since the word was completely new and any [p] that did show up

[2] It seems safe to rule out the possibility that the stop would be part of the second syllable, i.e., [-pstɚ], [-kstɚ].

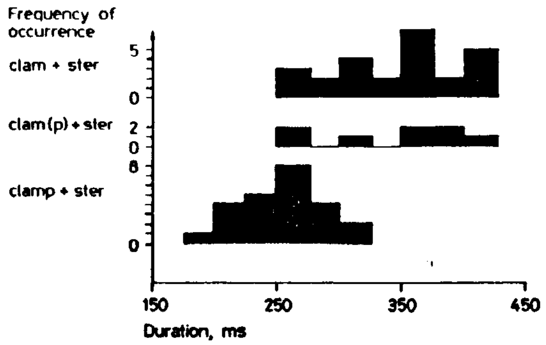

Fig. 1. Histograms showing the distribution of durations of the *vowel* + *nasal* sequence in 25 tokens of *clamster* (top), 8 of those tokens of *clamster* showing a clear epenthetic stop (middle), and 24 tokens of *clampster* (bottom).

would not have had time to be fossilized yet.) The word *clampster*, on the other hand, would contain what was assumed to be an underlying /p/, i.e., the final /p/ in the constituent morpheme *clamp*.

Oscillographic representations were made of Ss' recorded pronounciations of these words. Relevant segment durations were measured by hand. Most Ss pronounced these words more than once; nevertheless, many tokens were unmeasurable. In total 25 tokens of *clamster* were usable and 24 of *clampster*.

Results and Discussion

The distribution of the durations of the VN sequences in the two words are shown as histograms in figure 1. Two histograms are given for the word *clamster*, one for all the tokens of the word and a second one for the subset (8) which showed an epenthetic stop. (It is possible that some tokens with epenthetic stops were missed since the acoustic signal was not always clear.) Obviously the distributions of the two words are quite different ($p < 0.001$, one-tailed t test), even though there was no attempt to normalize these durations by taking into consideration possible between-speaker variations in the rate of speaking.

Thus, as predicted, it is possible to find durational differences in words which may be segmentally identical in their phonetic transcription, i.e., ['kʰlæmpstɚ], but which presumably differ in their underlying forms. Given this result, an attempt was made to use this technique to determine the status of epenthetic stops in existing words. Only preliminary results are available from this second study.

Study 2

Methods

Only a few Ss have been processed for this study. The Ss were (and will be) similar to those in the first study. Ss were told that the purpose of the study was to investigate possible differences in the stressing of new words as opposed to existing words. These instructions presumably justified the study to Ss, but deflected their attention from the precise point of interest. In addition to recording Ss' voice with a microphone, their oral air pressure was sampled via a pressure transducer connected to a short stiff tube held (as a pipe would be) between their lips at the corner of the mouth. The open end of the tube projected less than 1 cm beyond the lips and thus picked up pressure built up behind labial constrictions, but not behind constrictions made further back in the mouth. Instructions regarding the formation of new words were as in the first study. E orally cued each S about the word to utter and S then pronounced it three times in the frame sentence 'I'll say ____ once'. Only the pronounciation was elicited. In addition to the target words *teamster* and *Samson* (words likely to have epenthetic stops), there were also 'calibration' words, e.g., *team, Sam, sum, sumster, sump, sumpster, cam, camster, camp, campster*. (As before, the E cued the S to produce the derived words and the target words by putting the word ending first, e.g., 'Add *-ster* to *team*'.)

The duration of the VN sequences were measured by hand as before. Two quantitative criteria were used to determine the status of the epenthetic stops in the target words. These were applied to the data of individual Ss, not data pooled from many Ss. First, the duration of the VN sequence in a target word, e.g., *teamster* (assuming the pressure trace showed that it did have an epenthetic stop), was compared with the duration of the VN sequence in an appropriate 'calibration pair', e.g., *sumster* and *sumpster*[8]. If the VN of the target word was durationally similar to the calibration word not having the underlying /p/, then the target word was judged also to lack an underlying /p/. The second criterion was to examine the ratio of, e.g., the duration of VN in *team* to that in *teamster* and to compare it with ratios similarly formed from *sum* with respect to *sumster* and *sum* with respect to *sumpster*. It was assumed that an underlying /p/ was indicated if the ratio of *team* to *teamster* was similar to that of *sum* to *sumpster*, otherwise a surface [p] was indicated.

Results and Discussion

Figure 2 shows sample traces of the microphone and oral air pressure signals for one S. The duration of the sequence [im] in *teamster* (which by the oral pressure trace is revealed to have an epenthetic stop) is similar to that in *sumster*, not *sumpster*. Thus, by the first criterion, this token of *teamster* ['tʰimpstɚ], is judged to have only a surface [p]. In general, approximately 70% of those tokens of *teamster* and *Samson* which showed an epenthetic stop could be unambiguously

[8] It is desirable to make the comparison between vowels having the same intrinsic duration, ideally, the same vowel. The English lexicon does not always provide us with enough minimal pairs of the sort *sum* and *sump*, however, which (a) can plausibly be derived with the suffix *-ster*, (b) have an easily measurable VN sequence, and (c) contain the same stressed vowel as the existing target word. The comparison between *teamster* and *sumster* or *sumpster* is probably legitimate since [i] and [ʌ] have similar intrinsic durations [Peterson and Lehiste, 1960; House, 1961].

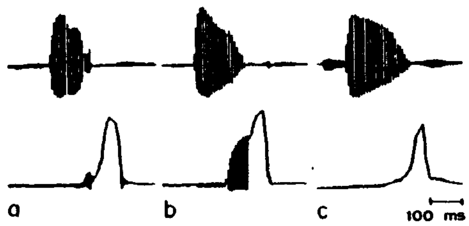

Fig. 2. Tracings of the oscillograms obtained from 1 speaker of the microphone signal (top) and oral pressure sampled just behind the lips (bottom) for the initial stressed syllables of the words *sumpster* (a), *sumster* (b) and *teamster* (c).

analyzed according to the criteria specified above. (In most cases, these criteria showed the epenthetic stops to be surface events.) The 30% which could not be analyzed either had values intermediate between those provided by the two types of calibration words or the two criteria gave conflicting results. It is possible that a more refined temporal analysis would work on these items, e.g., by taking into consideration individual Ss' idiosyncratic intrinsic durations of vowels, and by normalizing the tokens for within-speaker rate variations.

General Discussion

There may be overriding phonotactic reasons why a 'true' underlying /p/ would not be found in a word such as *teamster*: English apparently has a fairly pervasive constraint against the tautosyllabic sequence *tense vowel* + *nasal* + *voiceless stop*. The few exceptions, e.g., *paint, joint, pint, wont, mount*, generally involve alveolar nasal + consonant clusters. This constraint would not apply to *Samson* or similar words with lax vowels before the nasal + consonant cluster.

The following possible objections can be raised against the analysis presented here:

(1) A sound change may have taken place in English creating two underlying series of stops, i.e., a normal /p/ and an epenthetic /p*/, each of which creates different durational effects in the words in which they appear. Thus, contrary to my claims, both types of stop would be present at the underlying level. This point cannot be dismissed casually, since similar diachronic developments have created long

versus short segments in many other languages (e.g., Sanskrit *sarpa* 'snake' > Prakrit *sappa*), although the conditions are usually quite different from those involved in the formation of epenthetic stops. In any case, it is not possible to deal with this objection until a serious, fully documented case is made for it.

(2) Contrary to my assumption, the durational characteristics of the segments measured in this study may be determined at the surface, not the underlying level. This is the expressed view of, among others, HALLE and STEVENS [1967] and CHEN [1970]. Nevertheless, I am convinced by the arguments of LISKER [1974] that there is no phonetic basis for such effects and that they are instead part of the sound pattern of English, i.e., properties of the underlying level. (On the other hand, I accept JAVKIN's [1979] arguments that many of these effects originated diachronically due to phonetic factors.) This case underscores the fact that before any use may be made of speech timing (or any other phonetic parameter) for phonological questions, a great deal of phonetic research, such as that reviewed by LISKER [1974], is necessary.

Acknowledgments

I am grateful to HARUKO KAWASAKI and MICHELLE CAISSE for their assistance in these studies. This research was supported by a grant from the Committee on Research, University of California, Berkeley.

References

BARNWELL, T.P., III: An algorithm for segment durations in a reading machine context. MIT Res. Lab. Electronics, Techn. Rep., No. *479* (1971).

CHAO, Y.-R.: The non-uniqueness of phonemic solutions of phonetic systems. Bull. Inst. History and Philology, Academia Sinica *4:* 363–397 (1934).

CHEN, M.: Vowel length variation as a function of the consonantal environment. Phonetica *22:* 129–159 (1970).

GRANDGENT, C.H.: Warmpth. Publ. Mod. Lang. Ass. *11:* NS *4:* 63–75 (1896).

HALLE, M.; STEVENS, K.: On the mechanism of glottal vibration of vowels and consonants. MIT Q. Prog. Rep., Res. Lab. Electronics *85:* 267–271 (1967).

HARMS, R.T.: Some non-rules of English (Indiana Linguistics Club, Bloomington 1973).

HERBERT, R.K.: Reanalyzing prenasalized consonants. Stud. Afr. Linguistics *6:* 105–123 (1975).

HOMBERT, J.-M.: Speaking backwards in Bakwiri. Stud. Afr. Linguistics *4:* 227–236 (1973).

HOMBERT, J.-M.: Word games: some implications for analysis of tone and other phonological processes. Working Papers in Phonetics, vol. *33*, pp. 67–80 (University of California, Los Angeles 1976).

HOMBERT, J.-M.; OHALA, J.J.; EWAN, W.G.: Phonetic explanations for the development of tones. Language *55:* 37–58 (1979).

HOUSE, A.S.: On vowel duration in English. J. Acoust. Soc. Am. *33:* 1174–1178 (1961).

JAVKIN, H.R.: Phonetic universals and phonological change. Rep. Phonology Lab., Berkeley, No. 4 (1979).

KENYON, J.S.; KNOTT, T.A.: A pronouncing dictionary of American English (Merriam, Springfield 1944).

LEHISTE, I.: Suprasegmentals (MIT Press, Cambridge 1970).

LEHISTE, I.: The role of temporal factors in the establishment of linguistic units and boundaries; in DRESSLER, MAREŠ Phonologica 1972, pp. 115–122 (Fink, München 1975).

LINDBLOM, B.; RAPP, K.: Reexamining the compensatory adjustment of vowel duration in Swedish words. Occasional Papers, Language Centre, Essex, vol. *13*, pp. 204–224 (1972).

LISKER, L : On 'explaining' vowel duration variation. Glossa *8:* 223–246 (1974).

LOVINS, J.B.: Nasal reduction in English syllable codas. Papers from the Regional Meeting, Chicago ling. Soc., vol. *14*, pp. 241–253 (1978).

MALÉCOT, A.: Vowel nasality as a distinctive feature in American English. Language *26:* 222–229 (1960).

MILLARDET, G.: Insertions de consonnes en suédois moderne. Revue Phonétique *1:* 309–346 (1911).

OHALA, J.J.: How to represent natural sound patterns. Project on Linguistic Analysis Reports, Berkeley, vol. 16, pp. 40–57 (1972).

OHALA, J.J.: Experimental historical phonology; in ANDERSON, JONES Historical linguistics. II: Theory and description in phonology, pp. 353–389 (North-Holland, Amsterdam 1974).

OHALA, J.J.: Phonetic explanations for nasal sound patterns; in FERGUSON, HYMAN, OHALA Nasálfest: papers from a symposium on nasals and nasalization; pp. 289–316 (Language Universals Project, Stanford 1975).

OHALA, J.J.: Articulatory constraints on the cognitive representation of speech; in MYERS, LAVER, ANDERSON The cognitive representation of speech (North-Holland, Amsterdam, in press).

PASSY, P.: Etudes sur les changements phonétiques (Librairie Firmin-Didot, Paris 1890).

PETERSON, G.E.: LEHISTE, I.: Duration of syllable nuclei in English. J. acoust. Soc. Am. *32:* 693–703 (1960).

SHERZER, J.: Talking backwards in Cuna: the sociological reality of phonological descriptions. Southwestern J. Anthropol. *26:* 343–353 (1970).

Received: August 19, 1980; accepted: August 24, 1980

JOHN J. OHALA, Phonology Laboratory, Department of Linguistics,
University of California, *Berkeley, CA 94720* (USA)