PERCEPTION OF SYNCOPE IN NATIVE AND NON-NATIVE AMERICAN ENGLISH

Joann Fokes and Z.S. Bond

Ohio University, Athens, Ohio, U.S.A.

ABSTRACT

Native and non-native English speaking subjects made forced choice identifications of word triads embedded in phrases as spoken by three different English speakers. The triads consisted of 1) words with initial unstressed [sə] syllables, 2) words created by vowel syncope resulting in s-clusters, and 3) words containing s-clusters. A three way analysis of variance revealed a significant interaction between the two subject groups, word triads, and the speakers. Native subjects were better able than the non-natives in identifying tokens even though there were no differential patterns in production. There was some bias in terms of speaker and particular word stimuli.

1. INTRODUCTION

Both native and non-native speakers alter the pronunciation of English in casual speech, but perhaps in different ways. For example, native Americans frequently employ syncope or vowel loss in the pronunciation of unstressed syllables. This phenomenon is well documented [3] in the case of internal unstressed syllables and appears to be correlated with word stress patterns. Such reductions seem to be more common in English than other languages because of its polysyllabic rhythm. Typically, syllables containing strong beats fall at irregular intervals and are surrounded or flanked by syllables with weak beats. Reductions also occur in initial unstressed syllables as in the casual pronunciation of s'pose for suppose. In fact, vowel syncope may spill over into more formal styles as in the network news commentary reporting recent "S'preme Court decisions".

In the preceding example, vowel syncope results in a word with two juxtapositioned consonants resembling a dictionary word which does indeed contain a cluster. For example, vowel syncope in support results in the production of s'port which them becomes a possible homonym with sport. Just how listeners identify words containing vowel loss which become homonyms with real words is the question of interest in this investigation. It can be hypothesized that correct word identification is based on the semantic content of the message. On the other hand, there could be confusions in the perception of the target word unless the phonetic characteristics of the utterance provide for cues in its correct perception. Thus, if the content is ambiguous, there could be phonetic information to aid in the perception of the intended word.

Before the perception of words containing vowel syncope can be adequately studied, the actual production of such items require description. The phonetic detail of clusters resulting from vowel syncope was previously investigated by Fokes and Bond [4,5]. They tape recorded ten American English speaking subjects and four non-native English speakers who read a series of six phrases or sentence sets. Each set contained a triad of test words embedded in the same phrase: 1) a word beginning with an unstressed syllable in the form of [sə] followed by [p] or [k], 2) a real word containing an initial cluster consisting of [sp] or [sk], and 3) a word containing an artificially created [sp] or [sk] cluster resulting from vowel syncope. The subjects reported no more difficulty in pronouncing such items as s'port than the other members of the triad, sport and support. Five tokens of each phrase for all subjects were analyzed spectrographicly. No group patterns were found for either American or non-native English speakers in their ability to differentiate real from artificial clusters in their speech. The stops in artificial clusters were not always aspirated. In addition, these data did not show the systematic reduction in length of /s/ in clusters as opposed to singletons reported by Klatt [5] and by Crystal and House [1,2]. Instead, individual subject patterns in the duration of the initial fricative, voice timing, or stop closure plus vowel were noted. Such individual patterns were not found among the non-natives. Rather, they lacked consistency within their own individual productions as if attempting alternate productions in a trial and error approach. As expected, they also inserted vowels within the real clusters which the Americans never did.

Since there were no consistent group patterns in the productions of subjects in differentiating words with unstressed syllables, real clusters or artificial clusters, one might predict that listeners would be unable to distinguish between the real and artificial clusters when embedded in the same phrase. Alternatively, if listeners are able to perceive artificial clusters as their target words with an unstressed initial syllable, there is likely information in the speech stream that was undetected in the studies by Fokes and Bond [4,5]. Of interest also was whether differentiation between real and artificial clusters is an ability restricted to American listeners or whether non-native listeners also are capable of making distinctions resulting from vowel syncope.

2. METHOD

2.1. Materials

The stimuli for the present study were the productions from the previous investigation and consisted of tape recorded readings of short phrase or sentence triads containing test words 1) with an initial unstressed syllable beginning with [s], 2) a real [sp] or [sk] cluster, and 3) an artificial [sp] or [sk] clus-

ter. Each member of a triad was inserted into the following phrase sets:

On (succumbing, scumming, s'cumbing) at parties.
He (secured, skewered, s'cured) the meat.
The (supplies, splice, s'plies) of tape.
My (support, sport, s'port) of baseball.

Four tokens of each item spoken by three native Americans and one proficient non-native speaker who had been speaking English since childhood were recorded in random order to make a listening tape of 192 items. The speakers were selected on the basis of clarity of the tape and the absence of any trace of an unstressed vowel in words containing either the artificial or real clusters. The reduced vowel was present in the test words with the unstressed syllables.

2.2. Subjects

The two groups of subjects were college students: 15 native American English listeners and 10 non-native listeners. The non-native groups' experience with English was limited to academic training in English in their homeland and from two to three years English contact at Ohio University.

2.3. Procedure

The subjects made forced-choice identifications (ex: splice/supplies) of each of the tape recorded tokens. Subjects listened via headphones in a quiet listening laboratory.

3. RESULTS

The percent identifications of the triads by both groups of listeners are given in Table 1. The American listeners identified real clusters and two syllable words nearly 100% of the time. They heard the artificial clusters as two-syllable words at variable rates ranging from 56.6% for one of the native American productions to only 7% for the non-native proficient speaker.

Non-native listener identifications of real clusters ranged from 79% to 90% and from 86% to 96% for two-syllable words. They identified artificial clusters as two-syllable words from 15% for the non-native speaker to 47% for one of the native speakers. Interestingly, the non-native subjects perceived the proficient non-native speaker's artificial clusters as the target word more often than the native subjects.

Identifications were also lexically dependent; s'cumb was rarely heard as succumb (8%), while s'port and s'cured were identified as two-syllable words 64% of the time. In fact, with the word scum removed from the analysis, identification of the artificial cluster rose to 59% for Speaker Four's productions and to 69% for Speaker 2. Identification also rose to a level of 38% for the non-native speaker productions as well.

Identification scores were submitted to a 2 by 3 by 4 repeated measures analysis of variance consisting of one between factor (two listener groups), and two within factors (4 English speakers and word triads). The Greenhouse-Geisser adjusted degrees of freedom were used to test the interaction and main effects. There were the following significant interactions: speaker by listener group ($F = 4.74$; $df = 2.27, 52.11$; $p<.01$); speaker by word triad ($F = 36.11$; $df = 3.26, 75.02$, $p<.0001$); and speaker by listener group by word triad ($F = 5.81$; $df = 3.26, 75.02$; $p<.0009$). There was no listener group by word triad interaction. In determining the source of the interactions, Speaker One was clearly different in that her artificial clusters could not be identified as intended by native subjects but were identified at somewhat higher rates by non-native subjects. Also significant were the main effects of listener group ($F = 23.35$; $df = 1, 23$; $p <.0001$); speaker ($F = 45.97$; $df = 2.27, 52.11$: $p <.0001$); and word triads ($F = 464.22$; $df = 1.44, 33.11$, $p <.0001$).

4. CONCLUSIONS

The American native subjects were better able to identify artificial clusters as the target word containing the unstressed initial syllable than the non-natives. This ability cannot be credited to semantic cues only since the test words were embedded in the same phrase. Subjects, however, were highly influenced by specific words and the linguistic background of the speaker.

Because there was no single invariant acoustic pattern separating real from artificial clusters, we speculate that both groups of listeners were using multiple cues as a basis for perceptual judgments. That is, any one speaker may have used a set of cues which, in turn, may have signaled the intended target word.

In addition, listeners may have the facility of adapting to the peculiarities of individual speakers and their intentions. Apparently listeners are able to perform in this manner even when given a minimal amount of speech data.

5. REFERENCES

[1] Crystal, T. H. and House, A. S., (1988a), Segmental Durations in connected- speech signals: current results, Journal of the Acoustical Society of America, 83, 1553-1573.

[2] Crystal, T. H. and House, A. S., (1988b), Segmental durations in connected- speech signals: syllabic stress, Journal of the Acoustical Society of America, 83, 1574-1585.

[3] Delattre, P. (1966). A comparison of syllable length conditioning among languages, International Review of Applied Linguistics, 4, 184-196.

[4] Fokes, Joann and Bond, Z.S., (1989). S'pose a vowel is lost, Spring Meeting of the Acoustical Society of America, Syracuse, New York.

[5] Fokes, Joann and Bond, Z.S., (1989). Vowel syncope in native and non-native English, American Speech-Language Hearing Association Convention, St. Louis Missouri.

[6] Klatt, D.H. 1975, Voice onset time, frication, and aspiration in word-initial consonant clusters, Journal of Speech and Hearing Research, 18, 686-705.

Table 1. Means and 95% confidence intervals for native and non-native English subjects in identifying the stimulus triads.

	Native		Non-native	
TRUE CLUSTERS	Mean	95% C.I.	Mean	95% C.I.
S1	99.2	97.9-100	90.4	82.6-98.3
S2	98.9	97.1-100	83.8	74.0-93.0
S3	97.8	96.3- 99.1	79.2	69.0-89.3
S4	99.2	97.9-100	86.7	78.4-95.0
TWO SYLLABLE WORDS				
S1	99.2	98.2-100	86.7	76.9-96.5
S2	99.7	99.1-100	96.7	92.1-100
S3	99.4	98.6-100	95.0	92.3-97.7
S4	100	100 -100	95.4	89.6-100
ARTIFICIAL CLUSTERS				
S1	7.2	3.2-11.3	15.8	7.5-24.1
S2	55.6	48.8-62.3	35.7	29.8-43.5
S3	48.1	39.0-57.1	42.5	28.3-56.8
S4	45.3	37.1-53.5	46.7	32.8-60.6