

# THE PERCEPTION OF RATE-INDUCED RESYLLABIFICATION IN ENGLISH: NATIVE ENGLISH LISTENERS\*

Byung-jin Lim, Ken de Jong, and Kyoko Nagao  
Indiana University

Stetson (1951) noted that repeating singleton coda consonants (VC) causes them to be perceived as onset consonants (CV) as the rate of repetition increases. The current study examines whether naïve listeners perceive such resyllabification, whether such perceptions are affected by details of the splicing technique, and whether they are affected by the voicing of the stop consonants. Stimuli were extracted from productions in which talkers repeatedly produced singleton voiced and voiceless stops in CV or VC position. Speakers tracked a metronome which increased tempo from 450 to 200ms/syllable. Open-set identification revealed that 1) while slow VCs are identified as such, fast VCs are identified as CVs a majority of the time, 2) CVs are rarely identified as VCs; however, 3) both CVs and VCs at fast rates are often identified as CVCs, especially when the consonant is voiceless. A forced-choice identification task indicates that fast VCs and CVs are both identified as CVs 80% of the time at fast rates. These perceptions are not appreciably affected by either the splicing details or the voicing of the consonant.

## 1 Introduction

This study is part of a larger project pursuing Stetson's (1951) observation that coda stops become resyllabified as onsets when they are produced at fast rates. Recent acoustic production studies (de Jong, 2001a and 2001b), show that repeated codas become similar to onsets at fast rates, and repeated onsets also become similar to codas at fast rates. Furthermore, rate scaling affects the temporal structure of onsets proportionally, while the temporal structure of codas is resistant to changes. In the current study, we examine naïve listeners' identifications of the speech from these earlier works. We 1) examine naïve listeners' responses to repeated onsets and codas at a range of rates replicating previous studies, 2) examine responses across stimulus editing techniques, and 3) study the effect of stop voicing on syllabification.

## 2 Experiment 1: Open-set Labeling Test

In order to see what sorts of labels naïve listeners would use to describe rate-varied repeated syllables, an open set identification test was carried out.

### 2.1 Subjects

Subjects were 24 undergraduate students at Indiana University. Among them, two were non-native speakers of English: one was Spanish and the other was Korean. Since it was our main concern to see the perceptual patterns of native speakers of American English, here we report the results of native speakers of English only. No listeners reported any history of speech or hearing disorders.

---

\* The current work is supported by the NIDCD (grant #R03 DC04095-01A2) and by the NSF (grant # BCS-9910701). The earlier version of this paper was presented at the 141st Meeting of the Acoustical Society of America at Chicago, Illinois. Abstract: *Journal of the Acoustical Society of America*, 109: 2311, 1pSC4.

## 2.2 Stimuli

Stimuli for the current study were extracted from a production experiment (de Jong 2001a, 2001b) in which four English talkers repetitively produced syllables with voiced and voiceless labial-stop onsets and codas as shown in Table 1. Repetition rates were controlled with a metronome, which started slow (450ms/syllable) and increased in rate throughout each trial to a rate of 200ms/syllable. Thus, production rates range across the stable production range from 2Hz to 5Hz as reported in Nelson *et al.* (1984).

	CODAS	ONSETS
Voiced	'eeb'	'bee'
Voiceless	'eep'	'pea'

Table 1. Corpus.

Three-syllable segments were extracted from rate-varying trials. In the first experiment, syllables were taken from the rate extremes - slow rate stimuli including syllables 7 ~ 10, and fast rate stimuli including the last three syllables for each utterance. Hence they had periods of approximately 450ms for the slow rates and 200ms for the fast rates. These segments were digitally excised at voicing offset from the vowel previous to the stimulus syllables, and at voicing onset from the vowel following the stimulus syllables, as shown in Figure 1. The three excised syllables were presented to the listeners, who were asked to identify the repeated word. Subjects were presented with 32 trials (4 speakers \* 2 speech rates \* 4 types). Each trial included three repetitions of one of the above four syllables.

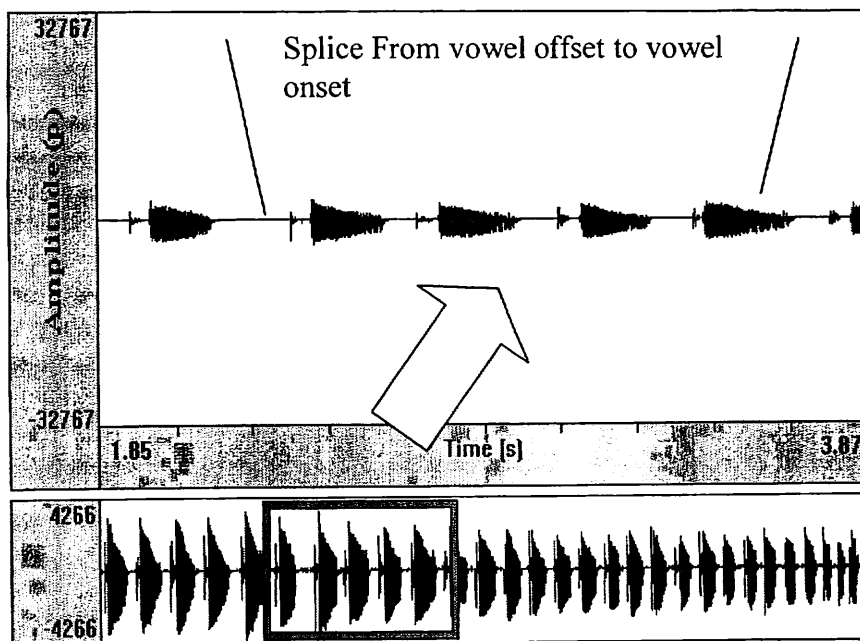


Figure 1. A three-syllable stimulus extracted from a complete utterance.

## 2.3 Procedure

Stimuli were played from a Sony mini-disk player through powered speakers in a classroom. Subjects were instructed to write down what they heard. They were informed that each recording had

been cut out of a longer utterance in which the speaker repeated a syllable many times. They were also warned that what the speaker was repeating might not be a real word. After listening to each recording, they were asked to write down what syllable they thought the speaker was repeating with regular English spelling. If they wanted to listen to the recording again, they raised their hands, and a stimulus was repeated. Before the experiment, there was a practice session to familiarize the subjects with the task. The session lasted about 30 minutes.

## 2.4 Results

The stimuli were labeled with the types of labels as summarized in Table 2. Pooled across rates, 72% of the responses were one of the intended syllable types; that is, they were reported as onsets or codas but not both. Productions at the slow rate were labeled as one of the intended syllables 85% of the time, while those at the fast speech rate were labeled as such 60% of the time. Interestingly, 23% of the responses split the consonant into two consonants (e.g., a repeated stimulus 'eep' was identified as 'beep') and a small number of responses were VCV such as 'ebee'.

		Response Type			
		CV(onset)	VC(coda)	CVC(both)	Other
Slow	CV stimuli	76%	2%	20%	2%
	VC stimuli	3%	88%	5%	4%
Fast	CV stimuli	59%	3%	34%	4%
	VC stimuli	52%	6%	35%	7%

Table 2. Results of open-set labeling test.

The results indicate that listeners usually identified the stimuli as one of the intended structures, CV or VC, though their responses did not necessarily match the intention of the speaker. At fast rates, there was an increase in onset responses especially for VC stimuli (52% favoring onset vs. 6% for codas) indicating subjects were much more likely to perceive stimuli as onsets than they were at slow rates.

## 3 Experiment 2: Closed-set Perception Test

The open-set identification test showed that listeners perceived resyllabification of repeated syllables at fast rates, and that their responses were generally one of the intended syllables; that is, either onset or coda. In Experiment 2, we presented a much larger set of stimuli taken from the entire range of speech rates.

### 3.1 Subjects

Twenty-three native speakers of American English were recruited from the Indiana University student population and they were all paid for their participation. Of 23 listeners, 20 were female and most were undergraduates, with an average age of 21 years. Most were born and had grown up in Midwestern areas (13 out of 18). All were monolingual speakers of English with no hearing problems. About a third (39%) of the participants were students in Speech and Hearing Science and most of the participants (83%) had taken some linguistics courses.

### 3.2 Stimuli

The stimuli consisted of three-syllable segments taken from all portions of the original utterances. These segments were digitally excised as in Experiment 1 as shown in Figure 1. Each stimulus included three repetitions of one of each of the above four syllables listed in Table 1.

### 3.3 Procedures

After listening to the stimuli through headphones, listeners were asked to identify what they heard by clicking one of the four labeled icons on the computer monitors: 'pea', 'bee', 'eep, and 'eeb'. In addition, there was a scroll bar ranging from 0 to 100 for subjects to indicate the confidence level of each choice. Confidence ratings are not analyzed in the current study.

Prior to the experiment, a brief practice run of 8 trials was conducted to familiarize the listeners with the task. None of the listeners showed any difficulty in performing the task. During the experiment, a five-minute break was given after every 100 trials.

### 3.4 Results

Figure 2 shows identification as a function of rate for the voiced consonant tokens. The vertical scale indicates the proportion of onset responses. The horizontal scale indicates the location of the three-syllable segments in the original utterance, ranging from slow repetitions at the left and fast on the right. Circles mark the original CV stimuli while squares mark VCs.

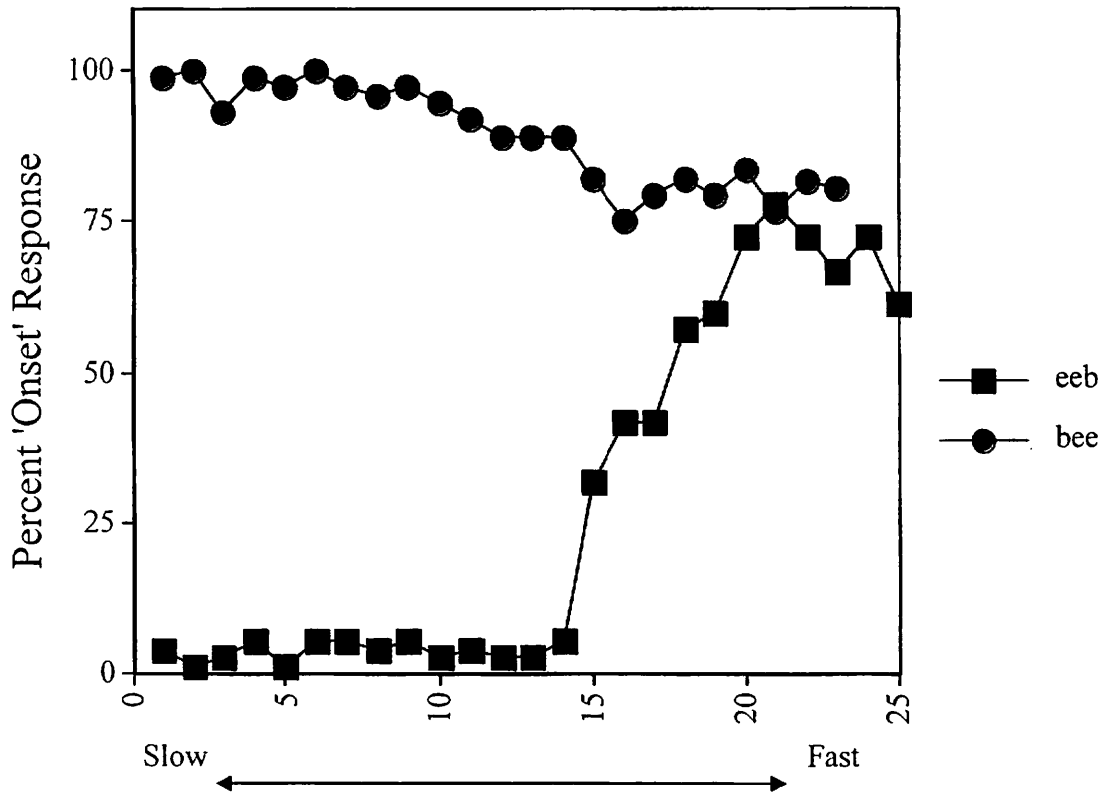


Figure 2. Percent onset response for voiced consonants.

The labels of intended codas (squares) change from nearly 100% coda (low onset responses to the left) to mostly onset responses suddenly around stimulus number 15. However, CV tokens at slow rates similarly are labeled nearly 100% at such, and as rates increase tend to be heard more often (about 20% of the time) as codas. Thus, at fast rates both intended onsets and intended codas elicit roughly the same proportion of CV and VC responses at fast rates.

Figure 3 displays identification functions for voiceless consonants. The pattern is very similar to the voiced consonants in Figure 2. This suggests that voicing does not substantially affect the identification shifts. However, a closer examination shows subtle effects of the voicing contrast on such perceptions. For example, even at fast rates, almost 90% of 'pea' productions are labeled as onsets. Furthermore, intended 'eep' seems to be labeled more often as an onset, evidenced by shifts beginning around stimulus number 11. Also, the identification function shows a shallower change in perceived syllable type than for the 'eeb' productions in Figure 2.

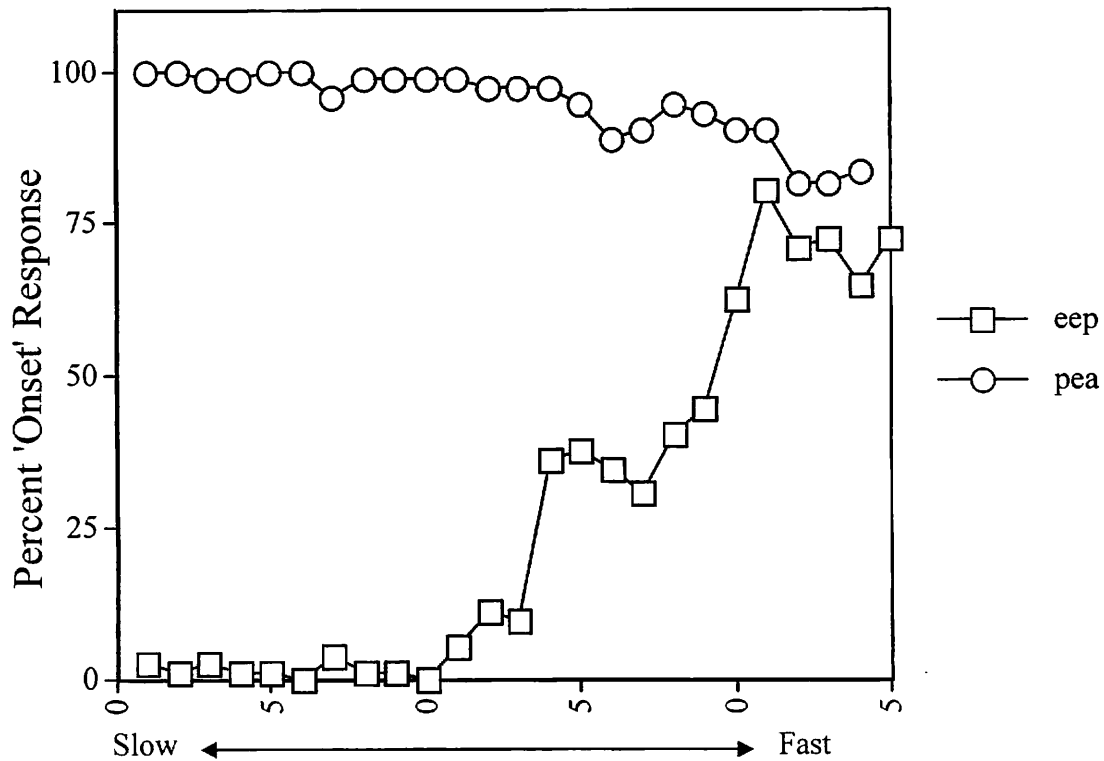


Figure 3. Percent onset response for voiceless consonants.

These results indicate that there is an asymmetry between intended onsets and intended codas in terms of identification functions. The tendency to label fast rate onset types as codas is much smaller than the tendency to label fast rate coda types. In addition, the voicing value of stop consonants seems to have very little effect on perceived resyllabification.

#### 4 Experiment 3: Replication with Different Splicing Technique

One aspect of the results of Experiment 2 was not expected, namely the incomplete resyllabification of the coda types into onsets. Experiment 3 was designed to test the possibility that the inclusion of a stop release at the final edge of each stimulus might account for this incomplete resyllabification.

##### 4.1 Procedure

In contrast to the method used to splice stimuli in Experiments 1 and 2, new stimuli were spliced at voicing onset and voicing offset as shown in Figure 4. If the existence of a stop release at the final edge influences listener identification, the new stimuli would cause different perceptual patterns since they have no stop release at the final edge of the coda. In addition, to make the change symmetrical, the stop releases at the beginning of the stimuli were also omitted.

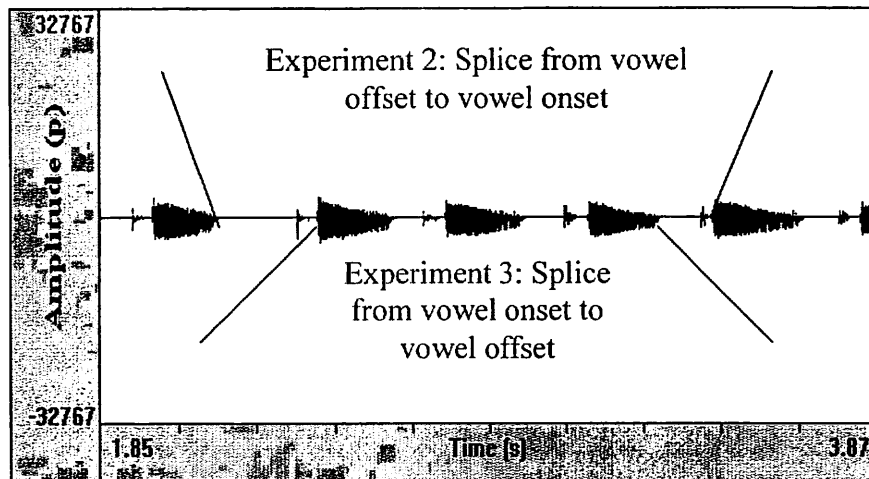


Figure 4. The alternative splicing technique used in Experiment 3, from vowel onset to vowel offset.

The new stimuli were presented to 23 subjects, six of whom had participated in Experiment 2. Except for stimuli and subjects, other procedures were the same as described in Experiment 2.

#### 4.2 Results

Figure 5 shows identification functions for Experiment 3 superimposed on the identification functions given in Figure 2. Results show no appreciable difference between the two stimulus sets, indicating no effect of the presence of a stop release at the final edge on identifications. Subjects were consistent in focusing on the middle portions of the repeated syllables; removal of consonant transients (especially stop burst releases) from the edges of the stimuli did not affect the identification functions. The proportion of onset responses for each stimulus in the first set (with stop transients) was regressed against the matched stimulus in the second set (without stop transients), yielding a very strong linear correlation ( $R^2=0.953$ ,  $m=0.93$ ).

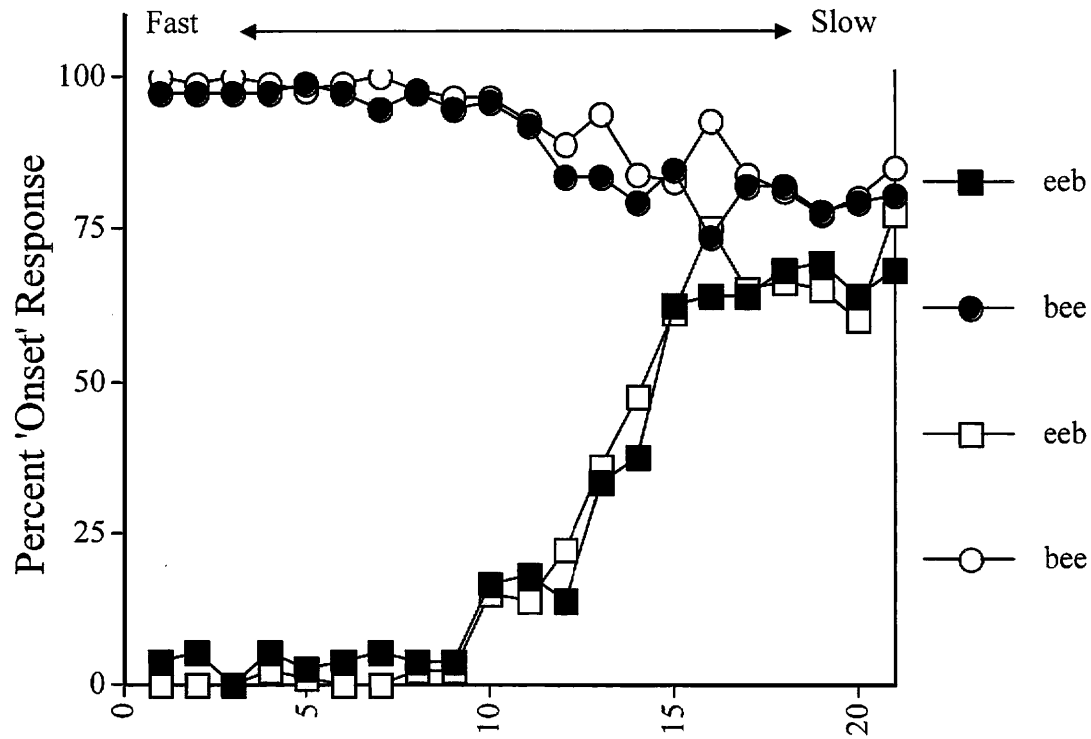


Figure 5. Identification functions for the two stimulus sets.

## 5 Summary

This study shows that naïve listeners perceive resyllabification at fast rates. Furthermore, perceptions are very robust to splicing techniques and also generalize across stops with different voicing specifications. Concerning the lack of a splicing technique effect, these results indicate that native listeners in general are quite consistent at assigning syllable affiliation, and are able to ignore the effects of the experimental context. This result is quite remarkable, when one notices that even the details of the identification functions are replicated in the two experiments. These details include proportional inconsistencies in the identification of fast rate stimuli. Thus, the listeners are able to ignore the effects of splicing and arrive at the same amount of disagreement as to the identity of the repeated syllables. This strongly suggests that the fast rate syllables are truly ambiguous, and different stimuli are more or less like canonical CVs. Concerning the essential lack of a voicing effect on syllabification, consonant voicing does not affect syllabification perception. Thus, the resyllabification effect is robust to different glottal configurations, suggesting that the effect is indeed one of syllable affiliation, and hence this phenomenon can tell us something about the nature of the syllable as a general entity, rather than just about the local organization of various glottal and oral gestures. Future work will address further questions of the generality of the phenomenon, as well as what acoustic and articulatory attributes of the speech give rise to the shift in apparent affiliation, as well as the inconsistency in the perception of resyllabification.

## References

- Boucher, V.J. (1988). A parameter of syllabification for VstopV and relative timing invariance. *Journal of Phonetics*, 16: 299 – 326.
- Davidson-Nielsen, N. (1974). Syllabification in English words with medial *sp*, *st*, *sk*. *Journal of Phonetics*, 2: 15 – 45.

- de Jong, K.J. (2001a). Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech production task. *Journal of Speech, Language, and Hearing Research*, 44 : 826 - 840
- de Jong, K.J. (2001b). Rate-induced resyllabification revisited. *Language and Speech*, 44: 197 - 216.
- Stetson, R. H. (1951). *Motor Phonetics*. Amsterdam: North-Holland.
- Tuller, B., and Kelso, J. A. S. (1991). The production and perception of syllable structure. *Journal of Speech and Hearing Research*, 34, 501-508.