The Interaction of Syllabification and Voicing Perception in American English

Kenneth de Jong, Kyoko Nagao, and Byung-jin Lim
Department of Linguistics, Indiana University

1   Introduction

1.1   Two ways of explaining phonology by phonetic facts

Being scholars who have been experimentally investigating language phenomena within the discipline of linguistics, it is heartening to see phonological theory over the past 50 years incorporating more and more physical attributes of the speech communication process into the explanation of phonological phenomena. As, we hope, interest continues in seeing what aspects of phonetic behavior can account for what aspects of phonological systems, it will soon become apparent that there are a wide variety of ways this can be done. In particular, there are two very different types of approaches which are often not clearly distinguished in discussions of phonetic explanation.

The first type explains parts of previous phonological analyses as being due to some specified phonetic strategy. Often, attributes previously accounted for in a phonological grammar may be accounted for by a specified set of production strategies that speakers use to accomplish a more sparsely specified phonological structure. A good example of this is in Pierrehumbert & Beckman's (1988, also Beckman & Pierrehumbert, 1986) model of Japanese tonal structure. Most phonological descriptions of Japanese tone structure previous to their work posited tonal specifications on each mora, along with context sensitive rules to account for, among other things, the apparent lack of an initial low tone in words with an initial accent. In Pierrehumbert & Beckman's model, the phonological specification of tone is more sparse and regular, the apparent loss or adjustment of tones in various contexts is due to the phonetic spell out of these tones. A conventionalized phonetic execution of a slimmed down phonological structure accounts for much of what one observes. This sort of approach does not require an emphasis on production. Kingston & Diehl (1994) employ a similar general strategy, though in their model the production strategies are heavily driven by the expected needs of the perceiver. What these models have in common with one another is 1) the on-line phonetic skills of the speaker explain phonological phenomena, and 2) the phonetics is either smart (as in Pierrehumbert & Beckman), or very smart (as in Kingston & Diehl).

The second type of approach, while perhaps not employed as commonly in current research, also has a very long track record. In this approach, a phonological phenomenon is due to very low-amplitude phonetic pressures constantly acting on the communication process over an extended period of time. Here, phonetic facts take a long time to create a phonological phenomenon. What makes this approach different from that outlined above is that the phonetic skills in individuals may be essentially irrelevant. It is historical dynamics which does the work of putting phonetic facts into the phonological grammar. Two excellent examples of this sort of model are Quantal Theory (Stevens, 1989) and 'Adaptive Dispersion Theory' (Liljencrants & Lindblom, 1971; Lindblom, 1986, Lindblom & Maddieson, 1988;

Lindblom & Diehl, 2001). Both of these models attempt to explain typological phonological facts in terms of phonetic facts, non-linearity in the 'links of the speech chain', and the shape of the articulatory space modulated by the auditory system, respectively. Note that, while attempts have been made to determine how such phonetic facts impact the phonetic skills of individuals (such as Perkell & Nelson, 1985, Perkell & Cohen, 1989; and Beckman et al., 1995, for Quantal Theory), such models do not need the individual phonetic skills to be demonstrably constrained by the phonetic factor which is doing the work. Also note that, while both models are formulated to explain typological facts, they are both amenable to explaining particular phonological facts of particular languages, as is demonstrated by de Jong & Obeng (2000).

Distinguishing between these two types of models is particularly important, then, since the predictions made by any particular model with respect to how any given individual will behave in an experimental setting depends on which view of the phonetic incorporation it embodies. Also, we cannot properly say we have understood a phonological phenomenon unless we can actually link the part of the phonology we are attempting to explain to the phonetic factor. Hence, with the second type of model, it is also necessary to understand how the phonetic factors actually give rise to the historical dynamics which produce the phonological phenomenon.

1.2   Syllable level voicing allophony

The current paper explores these two sorts of phonetic explanations of the relationship between syllabic position and the voicing contrast in American English. It has long been observed that the contrast between, for example, /p/ and /b/ is expressed differently, depending on the position of the stop with respect to the vowel. Preceding a vowel within a syllable, the contrast is largely one of aspiration. /p/ is aspirated, while /b/ is voiceless, or in some dialects voiced or even an implosive. Following a vowel within a syllable, both /p/ and /b/ both tend to lack voicing in the closure and the contrast is expressed largely by dynamic differences in the transition between the previous vowel and the stop. Here, vowel and closure duration are negatively correlated such that the /p/ has a shorter vowel and longer closure duration. This difference is often enhanced by the addition of glottalization to /p/. In addition to these differences, there are additional differences connected to higher-level organization involving stress and feet edges. To make the current discussion more tractable, we will restrict ourselves to the two conditions (CV and VC) laid out above.

The most straightforward traditional phonological approach to these facts would be simply to specify a context sensitive rule which either changes the identity of the two stops to something more closely approximating what appears phonetically, or to have language specific realization rules affecting rendering of some underlying feature, say [voice].

However, there are two facts which point out the shortcomings of this approach. 1) the type of allophonic variation we get in American English is also attested in a number of different languages, as pointed out by Keating, Linker & Huffman (1983). The pattern seems to be more general than would be predicted by an array of language-specific rules. 2) the allophonic difference between stops of the same type in different location seems to reflect a more general property of consonants in those syllabic positions. As is shown in, e.g., Turk (1993); de Jong (1998, in press), and others, consonants in prevocalic position tend to be more extreme in articulation and more temporally compact than consonants in post-vocalic position. Hence, expressing the voicing contrast in VC position in terms of vowel dynamics rather than in terms of contrasts linked to the consonant closure seems to 'make sense'. A

general property of consonants in that location is encoded in the allophonic variants of the stops.

Jakobson, Fant, & Halle (1952) have a brief discussion of the relationship between the voicing allophony and syllabic position which captures this non-independence. In their discussion of Danish, they state that /t/ is 'strong' and /d/ is 'weak'. Here, other features, such as aspiration and voicing are redundantly specified to enhance the main contrast. Syllable positions also are 'strong' or 'weak'. "Two positions are discernible in the Danish word - strong and weak. In monosyllabic words the strong position for a consonant is at the beginning of the syllable and the weak position, at its end." (p. 5) Hence onsets and /t/ have an affinity, while codas and /d/ also have an affinity. Allophonic interaction between the two then is simply expressed as an additive relationship between the two factors, as illustrated in Figure 1.
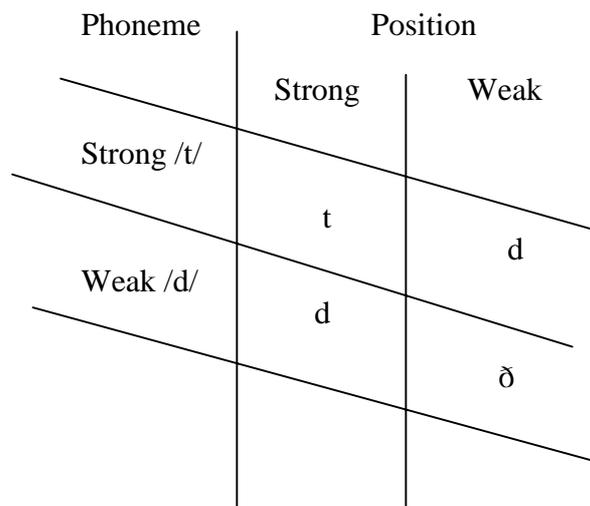

Figure 1. A schematic of the interaction between syllable position and the stop voicing contrast. (After Jakobson et al., 1952, Figure 1.)

Critical for this explanation is the existence of syllabic locations as having physical properties, here 'strong' and 'weak'. Thus, this sort of explanation runs counter to a trend in phonological theory toward explaining syllable-level effects as being due to some aspect of segmental sequencing, and not due to syllable level prosody at all. One such model is that of Kawasaki (1982) and Ohala (1990), who propose that syllable-level sequencing constraints are the result a long-term pressure toward acoustic modulation between successive segments. Similarly Mattingly (1984) notes that collation of segments at the level of the syllable tends to maximize the use of the frequency range used for speech transmission. Hence, any pressure toward long-term optimization of the transmission will tend to produce syllable level sequencing constraints. More recently, others, such as Wright (1996), Silverman (1996), and Steriade (1999), have articulated models in which specific syllable-level effects are due to the demands preserving specific segmental contrasts. In each of these models, the syllable is essentially an epiphenomena, and the syllable level phenomena are to be explained in terms of segmental contrast and sequencing.

Along these lines, Silverman (1998) presents a brief analysis of the voicing allophony in American English. The nature of this analysis as indicating a purely historical analysis in which syllable-level organization plays no part is spelled out in a more recent development of this work (Silverman, ms). In this analysis, he adopts Jakobson et al.'s (1952) description of the contrast between /b/ and /p/ as one of strength; /b/ is lenis, and /p/ is fortis. The syllable-conditioned allophony is generated by mechanisms specific to pre-vocalic and post-vocalic

sequencing. In the CV situation, there is a historic tendency to shifting the lenis consonants toward voicelessness, and a subsequent subtle perceptually-driven bias applied to the distribution of fortis consonants which tends to separate them from the now voiceless lenis stops. The VC situation is somewhat different in that the response to a historic tendency to push both the fortis and lenis consonants toward voicelessness is a switch to a perceptually more salient dimension involving the dynamics of the stop closing movement. Fortis consonants involve relatively fast closing movements and lenis consonants involve relatively slow closing movements. This contrast is subsequently enhanced by the addition of glottal stoppage in the fortis case, further shortening the acoustic transition.

In this analysis, Silverman repeatedly makes the point that allophonic variation is not generated by speakers optimizing their production system, but rather is due to low-level, perhaps undetectable biases in the transmission of speech patterns from one generation to another. Here, speakers simply copy the distributions of phones from previous generations. Syllable structuring is irrelevant for the allophony. Perceptual selectivity tends to produce certain types of effects which are sensitive to the temporal sequencing of sounds, and so can *post-hoc* be abstracted into syllabic units. However, one does not need syllable-level parsing to explain syllable-level allophony.

2 The current study

2.1 Rationale for the experimental approach

The current paper examines the degree to which synchronic aspects of the speech communication system can be implicated in the appearance of syllable-level voicing allophony in American English. Simply put, if the phonological facts of syllable-level allophony are to be explained by phonetic facts, are such facts detectable in the production and perception system of individual speakers and hearers? Or do phonetic facts create allophony only through long-term historical dynamics which are not robustly detectable in experiments on individual speakers?

The current project probes the relationship between voicing[1] and syllable position by systematically controlling rate in a repetitive speech production task. This variable and this task were chosen due to an observation by Stetson (1951, and documented by Tuller and Kelso, 1991) that syllabic parsing is sensitive to rate of repetition. Stetson noted that fast rates induce perceptual resyllabification of codas (such as the /b/ in 'eeb') into onsets (as in 'bee'). Hence, this task allows us to investigate how consonants get parsed into syllables. By including consonant voicing differences, we can investigate the relationship between syllabic parsing and the voicing category of the consonant being parsed.

Thus, our inquiry into the occurrence of syllable-level allophony is indirect in two ways, in terms of technique, and in terms of specific questions posed of the data. First, in terms of technique, the project includes both a production component and a perception component. Speakers produce various forms with variation controlled explicitly in a production experiment. Their productions, then, are evaluated not in terms of their physical

---

[1] Throughout the description of the current data set, we will use the term 'voicing' to indicate the contrast between /p/ and /b/. This is purely a convenience and is not meant to indicate that the contrast is actually expressed or perceived on the basis of the occurrence of voicing in the stop closures. We will use the terms 'fortis' and 'lenis' when discussing interpretations of the data which implicate a model such as Jakobson, et al.'s or Silverman's.

make-up, but rather in terms of the category structure of perceivers. By evaluating productions in terms of perceptions, this study allows us to examine the degree to which the production and the perception system match one another, and also allows us to examine how the control variables in the production experiment affect the relationship between production and perception.

Second, in terms of questions posed of the data, we will pursue the question of how allophonic differences in stops and syllabic parsing affect one another. One type of hypothesis would be that they do not affect one another at all, speaker productions simply match perceptions. This would be the case if there is essentially no on-line relationship between allophony and syllabic location; speakers just produce items which correspond to the categories listeners have. However, another possibility is that there is a systematic misalignment between the productions and perceptions due to some production factor associated either with syllabic position or voicing which tends to create mismatch between production and perception. These mismatches would tend to push the phonological categories in particular directions over the long run.

2.2   Methods

Stimuli for the perception experiment were obtained by means of a production experiment involving four native speakers of some variety of mid-western American English. There were two female and two male speakers, all in their 30's, one of which is the primary author. Speakers were given one of four visual stimuli, 'eep', 'pea', 'eeb', or 'bee', and then asked to produce the form one time for each click of a metronome. A metronome was used to control the rate of repetition in order to insure a wide range of rates which are comparable across the four speakers. The stimuli used in the perception experiment were extracted from a condition in which the metronome started at a slow repetition rate of 450 ms/syllable, and which increased in rate by 12.5 ms increments until reaching a fast rate of 200 ms/syllable. This task did not prove difficult for the speakers, who performed it with little practice and with only a very small number of dysfluencies. In addition, previous analyses of the data show that speakers were quite good at tracking the rate of the metronome with their speech. Detailed analyses of the acoustics of these productions related to syllabification and rate changes are reported in de Jong (2001a) and de Jong (2001b), respectively.

The current paper, however, examines perceptual responses to these productions. In order to get judgements of voicing and syllabification at times throughout the utterances, three-syllable pieces were spliced from the original utterances, so that each syllable in the utterance was the middle syllable of one of the three-syllable stimuli. Two different splicing techniques were used, one which included consonantal transients and one in which vocalic portions formed the edge of the stimuli. Responses for the two techniques were virtually identical, indicating that listeners were very good at abstracting away from the spliced edges of the stimuli. (These results are reported in more detail in de Jong et al., 2001). Results presented below are for tokens with consonant transients included. Only the last 23 syllables were included, thereby eliminating the first syllables which tended to vary while the subjects entrained to the metronome. Thus, there were 21 stimuli for each utterance, one utterance for each speaker by word form were included for a total of 4 speakers X 4 word forms X 21 stimuli = 336 stimuli.

These stimuli were then presented to 23 native listeners of American English by means of a *Matlab* protocol on PC workstations. Listeners were presented with a stimulus, told that the sounds were spliced out of a longer utterance in which a speaker was repeating a syllable, and asked to identify the repeated syllable. They were given the same four choices

as the speakers were presented with originally, 'eep', 'pea', 'eeb', and 'bee'.[2] The task was self-paced, and the listeners could repeat each stimulus as many times as they wished before making a judgement.

2.3   Predictions concerning perceptual identification

With respect to the syllabification of the stimuli, we expected, based on informal evaluation, that listeners would perceive fast rate VC's as being CV's, just as Tuller & Kelso had found. However, the point of the current analysis is if and to what degree the voicing of the consonant affects this pattern of syllabic judgement. Analyses such as Jakobson et al.'s seem to predict that this perception of VC's as CV's should be encouraged by the consonant being a fortis /p/. Hence, 'eep' should be more likely to be identified as a CV than 'eeb'. If this effect is strong enough, it might also even induce a misperception in the opposite direction, such that a lenis /b/ in onset position might tend to be misidentified as a VC form. In contrast to this, an explanation of allophony such as Silverman's would predict no necessary effect of voicing on syllabification, since allophony is only indirectly related to syllabic parsing.

With respect to voicing judgements, the same sorts of predictions can be made concerning the syllabic location of the consonants. Again, analyses such as Jakobson et al.'s seem to predict that CV's would tend to be identified as fortis /p/, while VC's would tend to be identified as lenis /b/. A purely historical account, especially one in which listener's perceptions are based on an internal model of the distribution of productions would seem to predict no relationship between syllabification and voicing. However, a complication here is that historical accounts such as Silverman's rely on various phonetic effects of markedness to drive the changes which lead to allophonic variation. Specifically, lenis /b/ occupies the unmarked voiceless position, while fortis /p/ was driven away from the unmarked position by means of perceptual selectivity. This asymmetry would suggest that fortis /p/'s would tend to appear in the unmarked position, especially when the effects of the production factor (here, speech rate) are strong. Hence, one might find CV's would tend to be identified as lenis /b/, the category which occupies the unmarked position. It is unclear whether such an argument can be made for VC's since the markedness of the dynamics of the voicing effect in the vowel is unknown.

3   Results

3.1   Voicing effects on syllabic parsing

Rate effects on the perception of syllabic affiliation are presented in Figure 2. Here, the horizontal axis plots the position of the stimulus in the original utterances, and hence is an index of speech rate. Stimuli with low numbers appeared closer to the beginning of the utterance, and hence had slower rates than did ones later in the utterance appearing to the left in the figure. The vertical axis plots the proportion of CV identifications by all listeners combined.

---

[2] The listeners were also asked to give an indication as to how confident they were in their assessment by clicking on a scroll-bar, but these results are not considered in the current paper. More on this is presented in Nagao, et al. (2001).
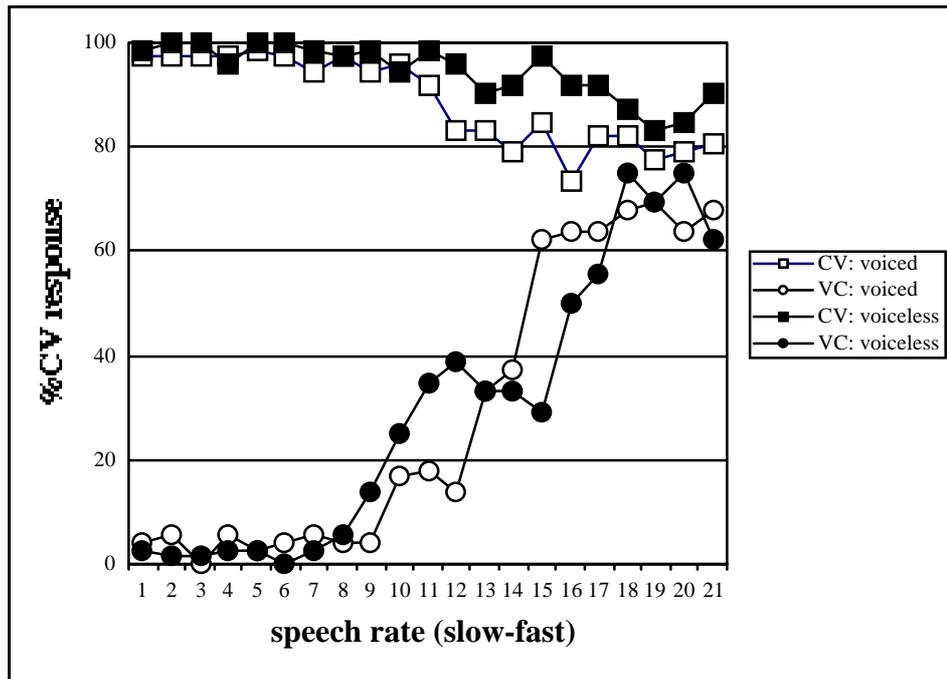
Figure 2. Proportion of CV responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for the four different produced syllables.

The general pattern of results for VC productions (circles in Figure 2) is what is expected from Stetson's and Tuller & Kelso's observations. VC productions are consistently identified as VC forms when repeated at slow rates, as is evident in the functions lying very close to zero. As rate increases, however, one sees a sharp increase in CV identification, trailing off at a level of approximately 75% at the fastest rates. There is an additional aspect of the current results which is not mentioned in previous works, which is a complementary shift which happens with CV productions (squares in Figure 2). At slow rates, CV productions are identified as such, the functions lying very close to 100%. As rate increases, there is a smaller downward shift which occurs at roughly the same location in the utterance, again to a level of approximately 75%. Analyses of this effect are currently under way; preliminary results suggesting that part of this effect is due to a generally greater degree of difficulty in identifying fast rate tokens. However other aspects of the analyses suggest that this effect is due to genuine ambiguity of the tokens with respect to whether the consonants belong with the previous or following vowel.

What is particularly of interest in the current discussion, however, is the difference between the voiced and voiceless forms (filled and hollow symbols, respectively). There is essentially no difference between the two sets of functions. Both sets show the same shifting pattern at the same locations in the utterances and to the same degree. There are subtle differences in the two functions, especially in the middle portions of the utterances; however, these differences do not seem to reflect a consistent bias in one direction or another. Some voiceless forms show more CV identifications, while others show less. In addition, an increment in proportion of CV responses for VC productions tends to get counteracted by a greater decrement for CV productions. The overall result of all this is that there is no discernable pattern of voicing effect on perceived syllabification.

To quantify these observations, we conducted a series of regression analyses, linking production types and rate variation to acoustic attributes and then to identification responses. Table 1 summarizes the relationship between the production control variables, voicing, syllabification, and rate, and perceived identification. As expected, most of the perceived

syllabification in the corpus corresponds to the stimulus given to the speaker, while a smaller but still substantial proportion of perceived syllabification is predicted by rate increases. The proportion accounted for by consonant voicing is remarkably small, less than one tenth of a percent of the variability in perceived syllabification.

Table 1. Summary of regression of proportion CV responses against production control variables.

	R^2	*p*
Syllabification (Intended)	.633	<.001
Rate	.061	<.0001
Voicing	.001	.503

As predicted by a model such as Silverman's, syllabic parsing is not affected by the voicing of the consonant being parsed. Thus, while the location of the consonant with respect to a vowel may be historically responsible for a shift in the phonetic nature of the consonant, these differences in the phonetic nature of the consonant are not used to parse the location of the consonant with respect to the vowel.

3.2 Syllabification effects on voicing identification

Figure 3 is similar to Figure 2, except that the vertical dimension here is proportion of voiceless responses rather than proportion of CV responses. Also, only CV productions are plotted. The general pattern with respect to rate is quite similar to that presented in Figure 2; voicing identification neatly matches the stimulus presented to the speakers at slow rates, but stops matching at faster rates. Specifically, there is a fairly large effect of produced voiced syllables ('bee' productions) being identified as voiceless ('pea'). Hence, there is a clear effect of rate inducing a shift in the productions which crosses the category boundary between /p/ and /b/.
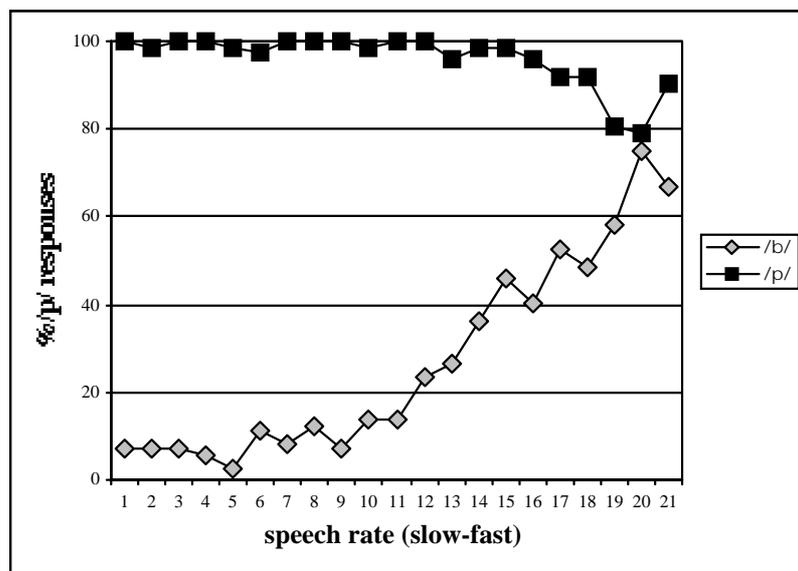

Figure 3. Proportion of voiceless (/p/) responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for onsets only.

Figure 4 is the same as Figure 3, except that it plots VC productions rather than CV productions. Here we also see a mismatch between production and perception, however, in the opposite direction. Here, voiced productions are consistently identified as voiced, while voiceless productions tend to be misperceived as voiced. A rate effect is evident here as well, however it is more subtle, and tends to be continuous across all of the rates.
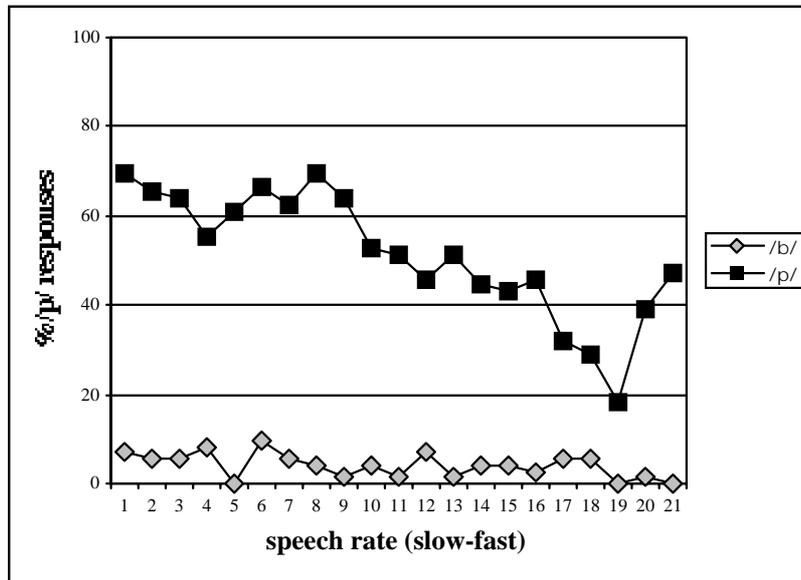

Figure 4. Proportion of voiceless (/p/) responses as a function of position in utterance which acts as an index of rate. Functions given here indicate responses for codas only.

Thus, it appears that some of the expectations based on Jakobson et al.'s analysis are born out in the current data. Onsets tend to be identified as voiceless, while codas tend to be identified as voiced. To illustrate this point, Figure 5 compares the schematic from Figure 1 to average data from the current experiment. Here, averages are calculated for the four stimuli presented to the speakers. Averages are in terms of proportion of times in which an item is identified as voiceless. As can be seen, the relationship apparent in the data to the right in Figure 5 is essentially that schematized to the left in Figure 5. What is interesting is that the vertical axis of the schema is labeled as 'strength', while the vertical axis in the data figure is proportion of /p/ responses. Thus, here, 'strength' is indexed in the likelihood that a perceiver will label a consonant as being a /p/. Hence, /p/'s in onsets are better /p/'s than ones in codas, and /b/'s in codas are better /b/'s than ones in onsets. Or, to restate Jakobson et al.'s analysis, being in an onset and being produced as a /p/ correlate, and being in a coda and being produced as a /b/ also correlate in the opposite direction.
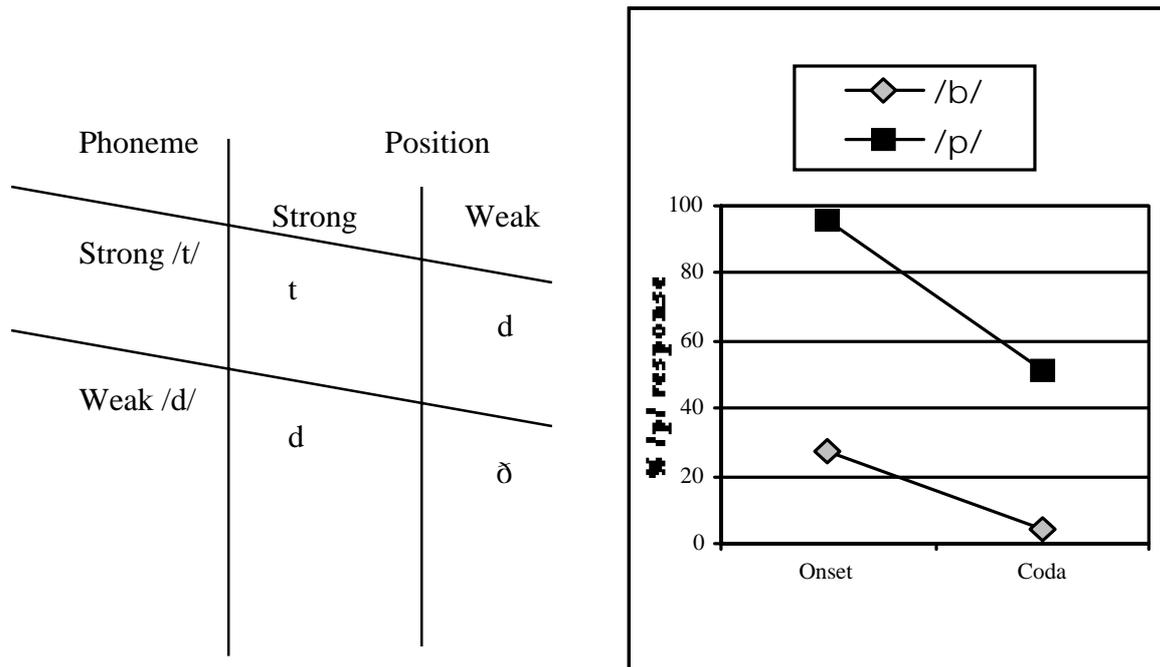
Figure 5. Average proportion of identifications as voiceless for each stimulus presented to the speakers. The schema from Figure 1 is repeated here for comparison.

4   Discussion and conclusions

At first glance, it appears that the current study has produced inconsistent results with respect to the general models being tested here. Stating these results with respect to our extension of Jakobson et al.'s analysis of Danish makes the point quite well. We predicted on the basis of this analysis that /p/'s make better onsets and /b/'s make better codas, and hence should tend to bias the identification of syllabic structure accordingly. This proved to be wrong; no bias is evident in the current data. We also predicted that onsets should make for better /p/'s and codas should make for better /b/'s. This prediction was born out. Hence, there is an interaction between location of the consonant and its identification, while there is no interaction between the identity of the consonant and its perceived location.

What these results suggest is a fundamental difference between the two identification tasks with respect to the production task. The syllabification task is a parsing judgement concerning the grouping of the material in the speech signal. It is heavily affected by rate scaling, a task which, according to detailed analyses of the current corpus in de Jong (2001a, and 2001b) has a disproportionate effect on the portion of the signal at the edge of each syllabic constituent. The identity of the consonant does not heavily affect the acoustics of these edges; both voiced and voiceless consonants have roughly the same global temporal structure. The voicing identification task, however, is a paradigmatic categorization task which relies on a wide range of much more subtle information. The difference in the robustness of the information on which these judgements rely has been implicated in a study which is parallel to the current one involving non-native listeners (Lim, de Jong, & Nagao, in

press; Nagao, de Jong, & Lim, 2001). Non-native listeners were speakers of Korean and Japanese, two languages which are notably different from English in both the voicing categories and in syllabic inventory. While the non-native listeners show a number of differences from native listeners in voicing categorization, they are remarkably similar to natives in their performance on syllabification judgement. What this suggests is that the parsing task is easier than the categorization task.

Hence, to reevaluate the conclusion reached above that the lack of an effect of voicing on syllabification supports a purely historical model of allophony, it is apparently the case that syllable level allophony is not there to help with syllabic parsing. This is consistent with a historical model of the allophony. However, it is also not the case that syllable level allophony is not synchronically connected to syllabic organization. Being produced in a certain syllabic position does, apparently, affect the production of voicing cues in the signal. Being produced as an onset tends to make a stop fortis, while being produced as a coda tends to make a stop 'lenis'. The direction of these effects is what would be expected of the pressure which would have driven the allophonic variation observed in the current system. That is, it is in the direction suggested by Jakobson et al's analysis. In addition, increasing rates tends to make this effect larger, suggesting that there is some sort of production factor underlying the effect.

In conclusion, diachronic modelling of the segmental phonetic pressures has some merit for explaining syllable-level allophony. However, the current study suggests that the pressures are not just apparent in a subtle generation-to-generation shift, detectable only through sampling entire production distributions. They are actually detectable in an experimental study which examines the relationship between production and perception under systematically varied conditions. In the current study, the production task employed here induces the appropriate changes, when evaluated with respect to the listeners' perceptual systems. That is to say, examining the relationship between production and perception in a communicative system, under conditions of controlled variation, gives a an indication of the sort of disequalibria which would over the long term tend to push phonological systems in particular directions. What this points out is that even long-term historical models are subject to experimental investigation, and that, within such models, the phonetics of individuals contain the structure which are found in phonological conventions.

References

Beckman, M.E., & J.B.Pierrehumbert (1986). Intonational structure in English and Japanese. *Phonolotgy*,3: 255 - 309.

Beckman, M.E., T.-P. Jung, S.-H. Lee, K. de Jong, A.K. Krishnamurthy, S.C. Ahalt, K.B. Cohen, & M.J. Collins (1995). Variability in the Production of Quantal Vowels Revisited. *Journal of the Acoustical Society of America,* **97**: 471 - 490.

de Jong, K.J. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics,* **26**: 283 - 310.

de Jong, K.J. (in press). Temporal constraints and characterizing syllable structuring. In. J. Local, R. Ogden, & R. Temple (eds.), *Papers in Laboratory Phonology VI.* Cambridge: Cambridge University Press.

de Jong, K.J., B.J. Lim, & K. Nagao (2001). The Perception of Rate Induced Resyllabification in English. (Abstract) *Journal of the Acoustical Society of America*, 109: 1pSC4.

de Jong, K.J., & S.G. Obeng (2000). Labio-Palatalization in Twi: Prosodic, contrastive, and quantal effects on linguistic micro-evolution. *Language*, 76: 174 - 195.

Jakobson, Fant & Halle (1952). *Preliminaries to Speech Analysis.* Cambridge, MA: MIT Press.

Kawasaki, H. (1982). *An Acoustical Basis for the Universal Constraints on Sound Sequences.* Ph.D. dissertation, University of California, Berkeley.

Keating, P.A., W. Linker, & M. Huffman (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277 - 290.

Kingston, J., & R.L. Diehl (1994). Phonetic knowledge. *Language*, 70: 419 - 454.

Liljencrants, J., & B. Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language*, 48: 839 - 862.

Lim, B.J., K.J. de Jong, and K. Nagao (in press). Cross-language perception of syllable affiliation: effects of voicing and language background. *Berkeley Linguistic Society,* 26.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J.J. Ohala and J.J. Jaeger (eds.), *Experimental Phonology.* Dordrecht: Foris.

Lindblom B., & I. Maddieson (1988). Phonetic universals in consonant systems. In L.M. Hyman, and C.N. Li (eds.), *Language, Speech and Mind.* London: Routledge.

Linblom, B., & R.L. Diehl (2001). Reconciling static and dynamic aspects of the speech process. (Abstract) *Journal of the Acoustical Society of America*, 198: 2380, 2pSC1.

Mattingly, I. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (eds.), *The Cognitive Representations of Speech,* pp. 415 - 420. Amsterdam: North-Holland.

Nagao, K., B.J. Lim, and K.J. de Jong (2001). Cross-language Perception of Rate Induced Resyllabification. (Abstract) *Journal of the Acoustical Society of America*, 109: 4pSC12.

Ohala, J.J. (1990). Alternatives to the sonority heirarchy. In M. Ziolkowski, M. Noske, & K. Deaton (eds.), Papers from the 26th Regional Meeting of the Chicago Linguistics Society: Vol. 2: The Parasession on the Syllable in Phonetics and Phonology. Chicago: Chicago Linguistics Society.

Perkell, J.S., & M.H. Cohen (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics*, 17: 123 - 133.

Perkell, J.S., & W.L. Nelson (1985). Variability in the production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America*, **77**, 1889 - 1895.

Pierrehumbert, J.B., & M.E. Beckman (1988). *Japanese Tone Structure.* Cambridge, MA: MIT Press.

Silverman, D. (1996). Phonology at the interface of phonetics and morphology: Root-final laryngeals in Chong, Korean, and Sanskrit. *Journal of East Asian Linguistics*, 6: 1 - 22.

Silverman, D. (ms). English alveolar stops, and the nature of allophony, University of Illinois at Urbana-Champaign.

Steriade, D. (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In O.Fujimura B.Joseph and B.Palek (eds.) *Proceedings of the 1998 Linguistics and Phonetics Conference,* pp. 205-242. The Karolinum Press.

Stetson, R. H. (1951). *Motor Phonetics.* Amsterdam: North-Holland

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17: 3 - 45.

Tuller, B., & Kelso, J. A. S. 1991. The production and perception of syllable structure. *Journal of Speech and Hearing Search*, 34: 501-508.

Turk, A (1994). Effects of Position-in-Syllable and Stress on Consonant Articulation, Ph.D. dissertation, Cornell University.

Wright, R.A. (1996). Consonant Clusters and Cue Preservation in Tsou. Ph.D. dissertation, UCLA.