Abstract. Increases in speaking rate and decreases in syllable stress have been thought to be accomplished with a common production strategy, and correspondingly, to produce comparable variations in vowel acoustics. If this is the case, it would be expected that a listener makes similar use of rate and stress information in vowel perception. In this study, the perceptual influences of rate and stress information were assessed for the perception of vowels in /pVp/ syllables. For each of nine vowels (i, I, e, er, a, o, u) a male talker recorded the carrier sentence "I think it's the yellow /pVp/" at slow and fast rates, while placing primary stress on the target /pVp/ or upon the preceding word "yellow." Synthetic vocalic regions of /pVp/ syllables were then generated on an OVE-III synthesizer, pairing the formant frequency values of natural slow-stressed syllables with the pitch, amplitude, and duration of the corresponding syllables produced in slow-stressed, fast-stressed, and slow-destressed carriers. These synthetic syllables were then edited into the natural carriers and presented to listeners in a vowel identification task. The pattern of identification errors differed for the three types of sentence carriers: short-long vowel pairs were confused infrequently and with about equal probability for the original slow-stressed carrier sentences, while the fast-stressed carriers were characterized by a predominance of long vowel-short vowel confusions and the slow-destressed carriers by short vowel-long vowel confusions. This difference is taken as evidence that rate and stress transformations are not perceptually equivalent, perhaps reflecting differences in how these factors are regulated in production.

Increases in speaking rate and decreases in syllable stress have often been characterized as having comparable consequences for vowel acoustics. Lindblom (1963) observed, for example, that the vowels in fast speech and destressed speech are shorter and reduced in vowel color relative to the vowels in slow-stressed speech. He therefore proposed that a single production strategy is employed to effect rate and stress changes. Invariant motor
commands are said to specify the articulatory targets that correspond to individual vowels, with rate and stress contrasts being defined by the time intervals between successive commands. In the case of rapid or destressed speech, these intervals would be too brief to allow the articulators to reach their targets, and this would result in shortened segment duration and reduced vowel color.

Acoustic measurements have not always shown an invariant pairing of shortened duration and spectral vowel reduction, however. We (Verbrugge & Shankweiler, 1977) found that the vowels in destressed /pVp/ syllables, spoken in the carrier sentence "I think it's the yellow /pVp/ 's call," were not appreciably shorter than their stressed counterparts, despite the fact that the two sets of vowels displayed marked spectral differences. Other investigators (Harris, 1978; Gay, 1978) have found complementarily that an increase in speaking rate may result in reduced duration with little change in center formant frequency values. These data suggest that rather than being regulated with a common production strategy, rate and stress are independently regulated, and they may differentially affect vowel acoustics; correspondingly, they may differentially affect vowel perception.

In a previous study (Verbrugge & Shankweiler, 1977), we conducted a perceptual test of listeners' use of the information for rate and stress in their identification of vowels. An adult male talker produced nine vowels in /pVp/ syllables in four sentence contexts: the target syllables were either stressed or destressed in sentences that were spoken at fast or slow rates. Listeners were then asked to identify the vowels when syllables were presented in their original sentence context or in a different context.

Identification errors increased significantly when fast syllables were presented in slow sentences. Somewhat surprisingly, a complementary influence of fast sentence context on slow syllables was not observed; nor was an effect of stress context observed for stress interchanged syllables, despite marked spectral differences in the syllables produced at different levels of stress. These negative findings may reflect a limitation of the experimental design, however. The target syllables themselves contained prosodic information that reflected their original sentence context. We suspected that this information was sufficient to specify their original stress level and, in the case of the slow syllables, to specify their original rate of articulation, even when these syllables appeared in an inappropriate sentence context. To overcome this limitation, the vocalic regions of /pVp/ syllables were produced synthetically in the current study, so that spectral information and prosodic information could be manipulated independently. In this way, target syllables could be produced with a prosody that matched the carrier sentence, but with a formant structure appropriate to a different context.

Three sets of stimuli were produced. In each set, the formant values for synthesis were those of natural stressed syllables spoken in a carrier sentence at a slow rate, with the formant values of the vowels /i, 1, e, a, o, a, ı, v, u/ being measured at the pitch pulse of maximum amplitude. Measurements for the first and second formants are displayed in Figure 1 (the points connected by the solid line), and it can be seen that these formants are representative for an adult male talker (cf. Peterson & Barney, 1952).
One stimulus set synthetically paired the slow-stressed formants with the prosody of their original slow-stressed carrier sentences, while a second set paired the same formants with the prosody of fast-stressed carriers. The resulting synthetic syllables were then edited into their corresponding carrier sentences, and the perceptual consequences of increased speaking rate were assessed by comparing responses to these slow-stressed and fast-stressed sentences. In the third stimulus set, the slow-stressed formants were paired with the prosody of slow-destressed carriers. A comparison of responses to the slow-destressed and slow-stressed sentences indexed the perceptual effect of destressing.

Subjects were asked to identify the vowels in the target /pVp/ syllables in one of two listening tests: they heard either a randomized list of slow-stressed and fast-stressed sentences, or a randomized list of slow-stressed and slow-destressed sentences. Responses were indicated by circling one of nine "p—p" words written in English orthography.

The parameters used in syllable synthesis are illustrated in Figure 3. As mentioned, the formant values were those of naturally produced slow-stressed syllables, with formants one, two, and three being iterated for the duration of synthesis. The duration was specified by the duration of the vocalic region of the natural syllable produced in the original carrier. Fundamental frequency was measured at the points of onset and offset of voicing and linearly interpolated between these points. Amplitude was also measured at onset and offset of voicing, and at the pitch pulse of maximum amplitude. The amplitude contour was interpolated from onset to the maximum and from the maximum to offset. Synthetic segments were generated on an OVE-III synthesizer and edited into carriers immediately after the natural /p/ release and before final /p/ closure. This pairing of natural and synthetic speech was effective; no subject detected a change of source and all subjects reported that the test sentences sounded natural.

To summarize, the target syllables in the three stimulus sets contained identical spectral information, but the pitch, amplitude, and duration of the syllables varied to reflect changes in rate and stress. This design permitted a direct comparison of the influences of rate and stress information on a perceiver's sensitivity to the spectral information for vowels. If, for example, fast speech and destressed speech have comparable spectral consequences in production, then one would expect to find similar perceptual responses to the fast-stressed and slow-destressed sentence carriers.

Responses were analyzed both in terms of the absolute accuracy of vowel identification and in terms of the pattern of confusion errors. With regard to the overall accuracy of identification, responses to the fast-stressed and slow-destressed carriers were similar, in that both groups showed an overall increase in identification errors relative to the original slow-stressed carriers. (Identification errors averaged 13% for the fast-stressed and 9% for the slow-destressed carriers, compared to 6% for the original slow-stressed carriers.)

Responses to the fast-stressed and slow-destressed carriers differed markedly in terms of the overall pattern of confusion errors, however. Confusion errors for the vowels /ɛ/ and /ə/ are displayed in Figure 4. The
open bars represent confusions of /ɛ/ with /ə/ and the hatched bars represent confusions of /ʊ/ with /ɛ/. Both types of errors are shown for the original slow-stressed carrier sentences, for the fast-stressed carriers, and for the slow-destressed carriers. It can be seen that when the slow-stressed formants appeared in their original slow-stressed carriers, /ɛ/ and /ʊ/ were never confused. However, when the same formants appeared in fast-stressed carriers, there were confusions of both types, but particularly confusions of /ʊ/ with /ɛ/. Finally, when the slow-stressed formants appeared in slow-destressed carriers, there were a number of /ɛ - ʊ/ confusions but no /ʊ - ɛ/ confusions. The pattern of confusion errors is clearly different for the fast-stressed and slow-destressed carriers.

It should be noted that the predominance of /ʊ - ɛ/ confusions in the fast-stressed carriers cannot be accounted for simply in terms of vowel duration. While it is the case that the fast-stressed syllables were shorter than the slow-stressed syllables, it is also the case that the fast syllables were presented in the context of rapidly articulated carriers. In this context, /ʊ/ was still relatively longer than /ɛ/. Presumably, then, the fast carriers provided sufficient information for accurately discriminating /ɛ/ and /ʊ/ on the basis of relative duration. Errors in the fast-stressed carriers are more likely due to alterations in perceptual sensitivity to spectral variables that distinguish /ɛ/ from /ʊ/.

Confusion errors for the vowels /ʌ/ and /ɑ/ are shown in Figure 5, with the open bars representing /ʌ - ɑ/ confusions and the hatched bars representing /ɑ - ʌ/ confusions. The pattern of errors is similar to that for the vowels /ɛ/ and /ʊ/. In the original slow-stressed carriers, /ʌ/ and /ɑ/ were confused infrequently and with equal probability. In the fast-stressed carriers, the frequency of confusions increased, and particularly, the frequency of /ɑ - ʌ/ confusions increased. Finally, in the slow-destressed carriers, while confusion errors dropped off overall, there were more /ʌ - ɑ/ than /ɑ - ʌ/ confusions.

It is evident that responses to the fast-stressed and slow-destressed carriers differ in terms of the direction of confusion errors, the fast-stressed carriers being characterized by a predominance of long vowel-short vowel confusions and the slow-destressed carriers by a predominance of short vowel-long vowel confusions. When responses are pooled over the short vowels /ɛ/ and /ʌ/, and the long vowels /ʊ/ and /ɑ/, an analysis of variance shows a significant interaction of carrier sentence type and the direction of confusion errors, \( F(1,22)=7.72, p<.05 \).

These data indicate that rate variation and stress variation both influence a listener's sensitivity to the spectral information for vowels. It appears, however, that their influences are not commensurate; an increase in rate and a decrease in stress invite qualitatively different vowel identification patterns. This conclusion is compatible with a growing body of acoustic data (Verbrugge & Shankweiler, 1977; Harris, 1978; Gay, 1978), and with production data. Harris (1978) has found, for example, that electromyographic activity associated with tongue body movement varies principally in terms of duration for rate changes and in terms of amplitude for stress changes.
The rate effects observed in this study deserve special mention. We have argued elsewhere (Verbrugge, Strange, Shankweiler, & Edman, 1976; Verbrugge & Shankweiler, 1977) that rate information serves principally to aid a listener in adjusting to variations in vowel duration. This observation has been based on the fact that when syllables were excised from sentence carriers and presented in citation form, or were swapped into sentence carriers spoken at different rates, errors tended to increase for short-long vowel pairs. However, in the current study, the interaction of rate and vowel duration was controlled through the synthesis of syllables with vowel durations that matched those of the natural syllables produced in each carrier. Nonetheless, rate effects were observed, suggesting that rate aids a listener in adjusting not only to variations in vowel duration, but to spectral variations as well.

Finally, these data can be at least partially addressed in terms of center formant frequency measures made for the natural syllables produced in the carriers. Formants in the natural slow-destressed syllables were reduced in vowel color relative to the slow-stressed syllables, particularly formant one (see Figure 1). Given this circumstance, it is not surprising that when the stressed formants were synthetically paired with destressed prosody, an increase of /E - æ/ and /ʌ - a/ confusions resulted.

A look at the formant space for the fast-stressed syllables is somewhat less helpful. The formants of fast-stressed syllables (Figure 2) do differ significantly from those of slow-destressed syllables (Figure 1), as would be expected given the results of this study. However, fast-stressed formants do not differ appreciably from slow-stressed formants (Figure 2), making an account of the rate effect observed here unclear. Some data reported by Gay (1974) may speak to the rate effects, however. Averaging over a number of tokens of the point vowels /i, a, u/ spoken by several talkers, Gay found that the vowel space for fast speech differed from that for slow speech in that first formant frequencies were higher for all vowels. Given this circumstance, one would predict a predominance of long vowel-short vowel confusions when slow-syllable formants are paired with fast prosody, and this is what we observed.

In conclusion, we have shown that the spectral information for vowels is perceived differently as a function of rate variation and stress variation. These distinct alterations in perceptual sensitivity correspond, at least in part, to the differing effects that these two transformations have on the spectral properties of vowels. Thus, the perceptual results provide indirect support for the hypothesis that rate and stress are independently regulated in production.

REFERENCES


Figure 1. Center frequencies of first and second formants of nine vowels spoken in /pVp/ syllables: slow-stressed and slow-destressed conditions.
Figure 2. Center frequencies of first and second formants of nine vowels spoken in /pVp/ syllables: slow-stressed and fast-stressed conditions.
Figure 3. Comparison of a natural /pVp/ syllable with a corresponding syllable in which the vocalic region has been synthesized.
Figure 4. Listener confusions of the vowels /ɛ/ and /æ/ for /pVp/ syllables with synthetic vocalic regions in which the formants of slow-stressed syllables were paired with the prosody of either slow-stressed, fast-stressed, or slow-destressed carriers.
Figure 5. Listener confusions of the vowels /ʌ/ and /ə/ for /pVp/ syllables with synthetic vocalic regions in which the formants of slow-stressed syllables were paired with the prosody of either slow-stressed, fast-stressed, or slow-destressed carriers.