PERCEPTUAL AND ACOUSTICAL ANALYSES OF VELAR STOP CONSONANTS

SARAH HAWKINS

Department of Linguistics
University of Cambridge
Cambridge CB3 9DA UK

KENNETH N. STEVENS

Research Laboratory of Electronics and
Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge MA 02139 USA

Abstract

An acoustic property that distinguishes velar consonants from labial, alveolar, and dental consonants is a prominent midfrequency "compact" spectral peak, usually in the frequency range 800-4000 Hz. In a series of perceptual experiments, synthetic syllables with initial voiced and voiceless stop consonants were generated, and the spectral characteristics of the consonant burst were systematically manipulated to yield various degrees of prominence of a midfrequency spectral peak. From listener responses to these stimulii, we have determined that the property of compactness depends in part on the amplitude of the prominent spectral peak in relation to a peak at about the same frequency in the following vowel. Spectral analyses of a number of naturally spoken stop consonants in English have shown that the amplitude characteristics of the midfrequency spectral prominence of the burst are consistent with the perceptual data. However, the degree of prominence often shows fluctuations throughout the region encompassed by the burst and voicing onset in the following vowel.

1. Introduction

The most distinctive acoustic characteristic of velar stops is usually said to be a compact spectral prominence, in the midfrequency range of 800-4000 Hz. In Fig. 1 we see smoothed spectra of the burst and vowel onset of a naturally-spoken /ga/, together with the waveform. The burst spectrum has the classical compact midfrequency prominence. Another attribute of the pattern in Fig. 1 is that the amplitude of the spectral peak in the burst is comparable (within about 5 dB) to the amplitude of the corresponding spectral peak in the vowel. This characteristic of the burst in relation to the vowel is consistent with data reported by several investigators [1, 2]. Velar stops also have a number of secondary characteristics, such as bursts that are longer, and first-formant transitions that tend to be slower than those for bilabials and alveolars. Nevertheless, spectra of velar stops vary a great deal, and the concept of "compactness" is poorly understood.

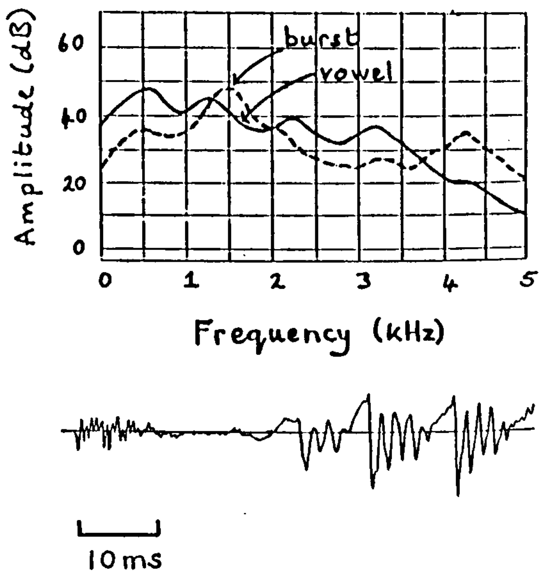

Fig. 1 Waveform (bottom) and spectra (top) sampled near the release of the syllable /ga/. Spectra are smoothed Fourier transforms sampled in the burst (dashed line) and at the onset of the vowel (solid line).

This paper describes some preliminary work in a planned series of studies of the acoustic characteristics of velar stops. We asked two questions, both of which focus on *spectral* rather than on temporal properties. First, can we synthesize an acceptable velar stop simply by manipulating the spectrum of the burst alone, and if so, what are the critical acoustic characteristics of such bursts? And second, to what extent are compact characteristics observable in *naturally*-spoken syllables? The focus of interest is the release burst and the first few milliseconds of the following vowel in syllable-initial stops.

2. Perceptual experiment

The stimuli for a perceptual experiment consisted of a series of synthetic consonant-vowel syllables. We constructed acoustic continua of bursts such that, when these bursts are followed by minimal vowels, we hear velar stops at one end of the continua, and either alveolars or bilabials at the other end, depending on the continuum. The various

bursts were synthesized manipulating the amplitudes of noise-excited formants in parallel synthesis.

Figure 2 shows short-time smoothed spectra of bursts at the extremes of the two continua—the velar-bilabial in the upper panel, and the velar-alveolar in the lower panel. In the velar-alveolar set of bursts, the classically compact shape of the spectrum labelled /g/ contrasts with the diffuse, rising spectrum, /d/, that is typical of alveolars. The variations in spectrum shape were achieved by changing the amplitude of excitation of $F2$. For the velar-bilabial set, formant amplitudes were altered so that the compact /g/ spectrum was made flatter and slightly falling, as for a bilabial.

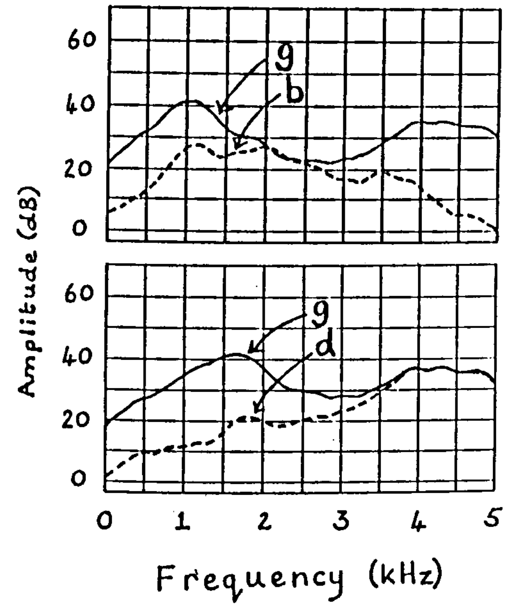

Fig. 2 Spectra of the bursts in the synthetic velar-bilabial continuum (top) and the velar-alveolar continuum (bottom). The spectrum at the velar end of each continuum is shown as the solid line and the spectrum at the bilabial or alveolar end is the dashed line.

Stimuli intermediate between these extreme pairs were made by changing the particular formant amplitudes in equal dB steps, resulting in two continua of 10 stimuli each. Listeners heard the bursts either in isolation, or with short, transitionless vowels following them, with or without aspiration. Since there were no transitions, the formant frequencies in the vowels were the same as those in the bursts (except for $F1$, for which there was a 20-ms rise at the onset from 250 to 500 Hz). $F2$ and $F3$ were lower in the velar-bilabial than in the velar-alveolar stimuli. For the velar-alveolar continuum, the burst duration was constant at 15 ms, whereas the burst duration decreased from 15 to 5 ms in one-ms steps for stimuli on the velar-bilabial continuum. We report here examples of results for the synthetic consonant-vowel stimuli, in which listeners were asked to identify the initial consonants.

Figure 3 shows the results for the CV stimuli of the velar-bilabial continuum and the velar-alveolar continuum for nine subjects. Forced-choice categorization functions for each continuum had a reasonably sharp crossover between 100% velar and 100% bilabial or alveolar responses, indicating that most listeners could classify sounds in terms of place of articulation using only the burst spectrum, perhaps relative to certain characteristics of the vowel spectrum. There are some differences between the responses for the velar-alveolar continua with and without aspiration, presumably a consequence of the closer proximity of voicing onset to the burst for the voiced continuum.

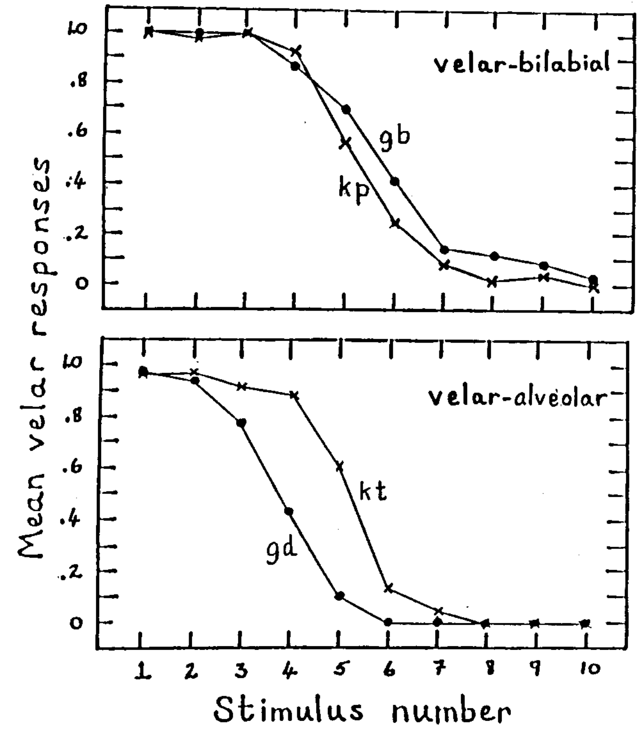

Fig. 3 Responses of nine subjects to stimuli on the velar-bilabial continuum (top panel) and the velar-alveolar continuum (bottom panel). The two functions represent the voiced (filled circles) and voiceless (crosses) continua.

In the upper part of Fig. 4 we show, for the velar-bilabial continuum, the spectrum of the vowel at the vowel onset (light solid line), together with burst spectra for two stimuli: the burst for the most extreme velar (heavy solid line) and the burst for the stimulus at which responses were closest to and not greater than 75% velar (dashed line). A similar display is given at the bottom for the velar-alveolar continuum. In both cases, velar responses for these synthetic stimuli are weakened when the midfrequency prominence drops 3-5 dB below the level of the second formant at the vowel onset. In some sense, the onset of the vowel might function as an anchor, or reference, against which the burst is evaluated.

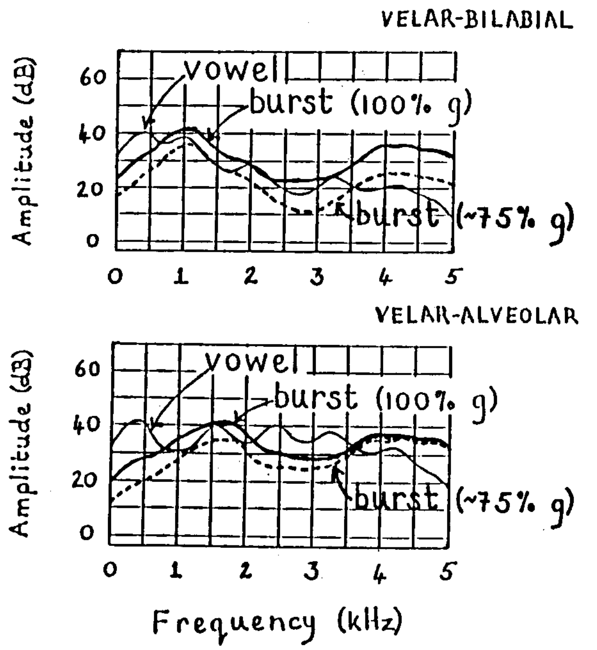

Fig. 4 The three spectra in each panel are the spectrum of the burst at the velar end of the continuum (heavy solid lines), the spectrum of the burst in the stimulus that elicited about 75% velar responses (dashed lines), and the spectrum near the onset of the vowel (light solid lines). The continuum corresponding to each panel is indicated.

Basically this experiment shows that when the midfrequency peak of the burst is sufficiently prominent in relation to the following vowel, listeners identify the consonant as velar even in the absence of transitions. *But* there are some puzzling aspects of these data, and of previous data obtained by others. First, four of our 13 subjects did not hear any velars at all on the velar-alveolar continuum, although their responses to the velar-bilabial continuum were basically the same as those of the other nine subjects. In fact, in presenting the results in Figs. 3 and 4 we have omitted data from the few subjects who heard no velars on the velar-alveolar continuum. These subjects may have focussed on the relation between the spectra of the burst and the vowel at high frequencies, presumably because the midfrequency spectral prominence was not sufficiently salient for them. (In subsequent experiments we increased the amplitude of the midfrequency spectral prominence by about 9 dB at the velar end of the continuum, and all subjects heard velars in this new continuum.) Second, several investigators have shown that velar responses can be obtained when the burst is completely absent, provided that a pair of adjacent formants (usually the second and third formants) are close together at the beginning of the transition into the vowel [3, 4].

These findings, together with published data on the analysis of velar consonants in real speech, have suggested to us that it is an oversimplification to describe a velar consonant as a burst of noise with uniform-amplitude over time, followed by a vowel with suitable transitions. Consequently, we have reexamined the acoustic properties of spoken velar consonants, particularly the fine structure of the short-time spectrum through the burst and into the onset of voicing.

3. Acoustic analysis of natural speech

We have looked at CV syllables spoken by several talkers of British or American English saying /gi/, /ge/, /ga/, /gu/. Several types of spectra were made of the burst and at least the first two periods of the vowel, including Fourier transforms, lpc spectra, and the output of certain auditory models. Spectra were made in successive 5 ms steps and additionally, for some syllables, in smaller steps.

Many of the spectra conform closely to the classical [compact] description for the burst. But as other investigators have also found [2], a substantial minority deviate from the classical picture. However, almost all of these so-called deviant or atypical utterances, have compact properties during at least *some* part of the burst or vocalic onset. Two of the most common types will be shown here.

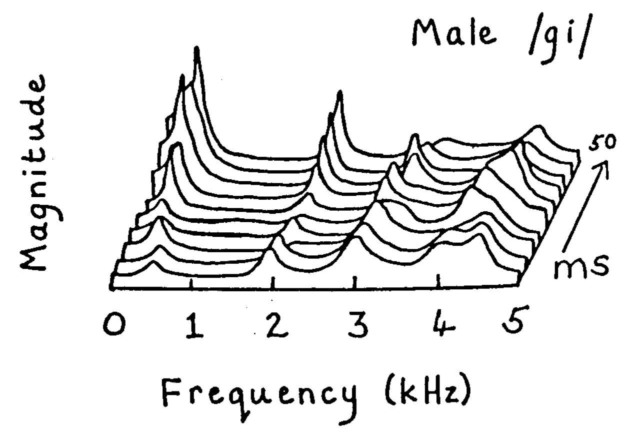

Fig. 5 Sequence of lpc spectra sampled at 5-ms intervals near the release of the syllable /gi/ produced by a male speaker. The amplitude scale is linear. Voicing onset occurs at about the 6th spectrum from the front. This sequence is an example in which spectral prominences appear intermittently in the burst.

Figure 5 shows lpc spectra at successive 5 ms intervals of a male talker's /gi/. Neither the burst nor the vowel onset appear particularly compact, but the peaks and valleys during the burst fluctuate somewhat in amplitude so that a more classical compact spectrum appears intermittently. Such fluctuating spectra will occur when there is a succession of transients at the release, although they can occur even when individual transients cannot be seen in the waveform. These fluctuations in compactness are more common in velars than other stops, presumably because of the longer constriction and slower release, and they may themselves contribute to the perception of velarity.

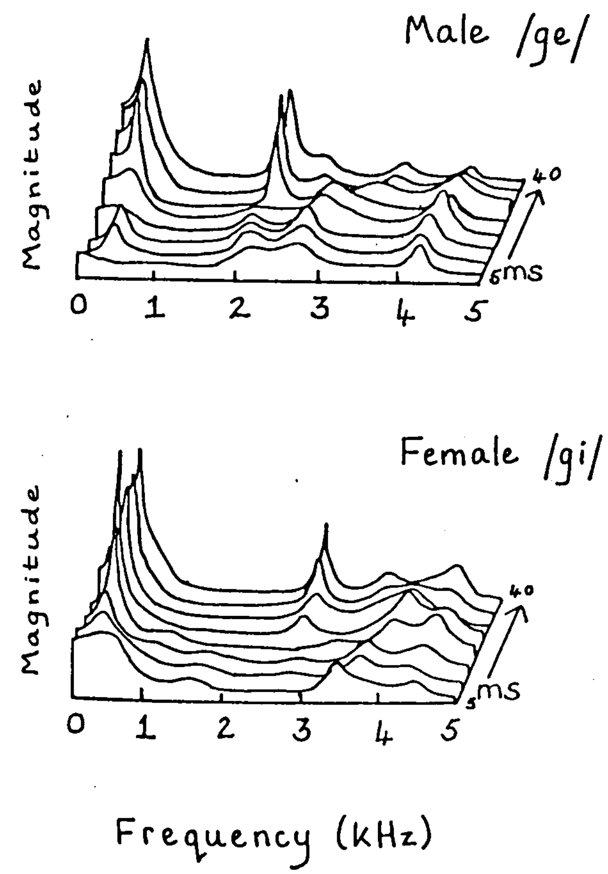

Fig. 6 These are examples of sequences of lpc spectra in which the burst does not show a compact spectral prominence, but a prominence appears following the onset of voicing.

Figure 6 illustrates a second type of nonclassical compactness in which the burst is not compact but there is a very sharp, narrow-bandwidth formant in the midfrequencies at the onset of the *vowel*. This sharply prominent peak appears abruptly and in relative isolation from surrounding peaks, and presumably arises because two formants are very close in frequency. This strongly compact vowel onset appears to be associated more with velars before front vowels, and possibly with weak bursts.

These two phenomena together—a spectrum that fluctuates in degree of compactness during the burst and a strongly compact midfrequency peak at the onset of the vowel—may each serve to enhance the compact percept. Rapid fluctuations in a spectrum may compensate for an otherwise weakly compact burst by somehow focussing attention by virtue of the fluctuating spectrum. And compactneses in the vowel onset may override any ambiguities of the burst. It is worth noting that we often found different types of compactness in different tokens of the same g-vowel syllable, or from the same talker. We are assessing the extent to which these "alternative forms of compactness" can strengthen the perception of velarity in synthetic consonant-vowel syllables.

4. Conclusions

In summary, we have seen in the perception experiments that velar stops in a CV syllable with steady-state formants are heard if the burst has a midfrequency spectral prominence with an amplitude at least as great as that of the corresponding peak at vowel onset. The analyses of natural speech show that the compact prominence is typically present in the burst spectrum, or it may be only intermittent, or it may be more evident in the vowel onset than in the burst. These data suggest that compactness should be defined not in terms of the prominence of a peak in the average burst spectrum, but rather in terms of the occurrence of prominence in the short-time spectrum in at least some region of the syllable onset, whether it be in the burst or in the onset of voicing. One possibility is that the perception of velarity in the consonantal release is enhanced if there are regions in the release phase where a compact spectral prominence is embedded in a context that has reduced compactness or prominence.

If these preliminary observations are confirmed on a larger dataset, then the next task is to begin to describe compactness more precisely through further preception experiments. If we can express compactness in terms of the amplitude, bandwidth, frequency range and time-course of a midfrequency peak relative to adjacent spectra, then we may be on the way to coming up with a description that subsumes burst and transition information under one umbrella.

5. Acknowledgements

This work was supported in part by grants from the National Institute of Neurological and Communicative Disorders and Stroke and from the National Science Foundation.

6. References

1. V.W. Zue. **Acoustic characteristics of stop consonants: A controlled study.** Ph.D. thesis, MIT, Cambridge MA USA (1976).
2. D. Kewley-Port. *Time-varying features as correlates of place of articulation in stop consonants.* **J. Acoust. Soc. Am., 73**, 322-335 (1983).
3. M.F. Dorman, M. Studdert-Kennedy, and L.J. Raphael. *Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues.* **Perception and Psychophysics, 22**, 109-122 (1977).
4. S.E. Blumstein and K.N. Stevens. *Perceptual invariance and onset spectra for stop consonants in different vowel environments.* **J. Acoust. Soc. Am., 67**, 648-662 (1980).