The stop–glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/

Susan Nittrouer and Michael Studdert-Kennedy[a)]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

(Received 19 February 1986; accepted for publication 18 June 1986)

Amplitude change at consonantal release has been proposed as an invariant acoustic property distinguishing between the classes of stops and glides [Mack and Blumstein, J. Acoust. Soc. Am. **73**, 1739–1750 (1983)]. Following procedures of Mack and Blumstein, we measured the amplitude change in the vicinity of the consonantal release for two speakers. The results for one speaker matched those of Mack and Blumstein, while those for the second speaker showed some differences. In a subsequent experiment, we tested the hypothesis that a difference in amplitude change serves as an invariant perceptual cue for distinguishing between continuants and noncontinuants, and more specifically, as a critical cue for identifying stops and glides [Shinn and Blumstein, J. Acoust. Soc. Am. **75**, 1243–1252 (1984)]. Interchanging the amplitude envelopes of natural /bV/ and /wV/ syllables containing the same vowel had little effect on perception: 97% of all syllables were identified as originally produced. Thus, although amplitude change in the vicinity of consonantal release may distinguish acoustically between stops and glides with some consistency, the change is not fully invariant, and certainly does not seem to be a critical perceptual cue in natural speech.

PACS numbers: 43.71.Es, 43.70.Fq

INTRODUCTION

Much research in speech perception has focused on discovering acoustic properties invariantly associated with phonological categories. For the most part, this work has not been successful. Rather, it has tended to show that the acoustic attributes associated with the perception of a particular category vary with phonetic context. However, there are exceptions to this trend. For example, Stevens and Blumstein (1978) found that the spectral tilt of the aperiodic energy immediately following syllable-initial stop release was invariantly related to the perception of place of articulation. Kewley-Port (1983) also reported invariant features distinguishing among places of articulation for syllable-initial stop consonants; these features were dynamic in nature.

Further, Mack and Blumstein (1983) reported a possible invariant cue for the stop–glide manner distinction: the ratio of the rms energy of a brief acoustic segment immediately following "release offset" to the rms energy of a brief acoustic segment at "release onset."[1] This ratio is large for stops and small for glides. In other words, stops have large and rapid increases in amplitude in the vicinity of consonantal release, while glides exhibit no such rapid amplitude change. The amplitude change itself largely results according to the acoustic theory of speech production from the narrowing of the pharyngeal cavity as the jaw lowers, and the consequent, correlated rise in frequency and amplitude of the first formant. The amplitude change in the first formant also, of course, spreads to the second and higher formants (Fant, 1960). In fact, Mack and Blumstein attribute their success in identifying an apparent invariant acoustic property to their use of a measure encompassing the whole spectrum of natural speech.

As a possible explanation for why earlier perceptual experiments failed to establish invariant relations between the acoustic signal and linguistic description, Mack and Blumstein (1983, pp. 1739–40) suggest:

> "... in their search for a particular attribute of the signal which categorizes the stop–glide phonetic contrast, researchers have typically used highly stylized and schematized test stimuli to the extent that these stimuli have shared all attributes save the particular dimension in question. The use of such stimuli may involve failing to manipulate or ignoring critical acoustic attributes for stops and glides present in natural speech."

While this criticism may be reasonable, it should be stressed that discovering an invariant acoustic property is not the same as demonstrating that the property has an invariant perceptual effect. If it were, perceptual studies would be supererogatory, and we could confine our attention to acoustic analysis. Mack and Blumstein's (1983) results therefore offer strong support for the well-known fact that differences in amplitude change in the vicinity of consonantal release are invariantly associated with the different productions of stops and glides, but do not address the question of whether these amplitude patterns invariantly lead to the corresponding percepts.

In an attempt to address this latter question, Shinn and Blumstein (1984) conducted a perceptual experiment using stimuli synthesized to approximate the natural tokens measured by Mack and Blumstein (1983). Two separate continua were constructed, one with formant values appropriate for /i/, and one with formant values appropriate for /a/. Three amplitude envelopes were given to both the /bi–wi/ and the /ba–wa/ continua. For one set, the amplitude envelope varied systematically between the /b/ end and /w/ end of the continua, according to values derived from Mack and

[a)] Also of Queens College and The Graduate School, The City University of New York.

 0001-4966/86/101026-04$00.80

Blumstein (1983). In a second set, all tokens were given a /b/ amplitude contour, and, in the third set, all tokens were given a /w/ contour. Combined results of forced choice and free identification tasks led to the general conclusion that the amplitude envelope invariantly distinguishes between continuants and noncontinuants, and is a critical perceptual cue for the specific distinction between stops and glides. This conclusion was summarized by the observation that the amplitude envelope was "... sufficiently strong to override the formant frequency, rate, and duration cues..." (Shinn and Blumstein, 1984, p. 1249).

One potential problem with this study is that, although the manipulated acoustic attributes of the stimuli were derived from analysis of natural speech, they were nonetheless synthetic. Whether they could be accurately characterized as "highly stylized and schematized" (Mack and Blumstein, 1983, p. 1739) is debatable, but clearly they were not natural. Accordingly, the possibility exists that some critical attribute for stops and glides present in natural speech may have been ignored (cf. Mack and Blumstein, 1983, p. 1740).

We therefore undertook a small study to determine, first, if we could replicate the acoustic measurements of Mack and Blumstein (1983), and second, to see what perceptual effect appropriate manipulation of the amplitude contours would have in natural samples.

I. ACOUSTIC ANALYSIS

Two male speakers were recorded reading CV syllables consisting of either /b/ or /w/ followed by one of the vowels /i,e,a,o,u/ using a TEAC recorder and Sennheiser microphone. Five samples of each syllable were obtained, yielding 50 tokens per speaker (25 each of /b/ and /w/). Tokens were digitized on a DEC Vax computer using a 10-kHz sampling rate and a 4.9-kHz low-pass filter setting. Release onset and offset were identified using the criteria described by Mack and Blumstein (1983). Amplitude ratios between rms values obtained immediately after release offset and at release onset were computed for each syllable. Table I reports mean values for samples for the two speakers in this study (column 3, "Original") as well as mean values and ranges obtained by Mack and Blumstein for four speakers (column 2).

TABLE I. Mean ratios of rms energy at release offset to rms energy at release onset. Data (means and ranges of four speakers) from Mack and Blumstein (1983), and from the present study (means of five tokens from each of two speakers). "Original" refers to mean values of original syllables. "Modified" refers to syllables with amplitude envelopes transposed to match those of appropriate templates.

Syllable	Mack and Blumstein	Present study	
		Original	Modified
/bi/	4.28	6.62	1.06
	(3.39–5.98)	3.76	1.35
/be/	4.22	5.54	1.05
	(2.62–6.02)	2.63	1.28
/ba/	3.47	7.62	1.14
	(1.70–6.93)	2.03	0.86
/bo/	3.77	5.48	1.01
	(2.15–5.33)	2.54	1.31
/bu/	2.49	6.77	1.10
	(1.37–3.21)	1.45	0.97
/wi/	1.12	1.24	7.07
	(1.05–1.16)	1.64	5.35
/we/	1.16	1.00	5.64
	(1.09–1.31)	1.31	3.30
/wa/	1.13	1.09	10.70
	(1.06–1.20)	1.11	7.46
/wo/	1.20	1.31	9.47
	(1.06–1.32)	1.00	1.96
/wu/	1.10	1.12	4.47
	(1.05–1.11)	1.70	2.48

It can be seen that, in large part, we replicated Mack and Blumstein's findings. All /b/ amplitude ratio means are above the criterial value of 1.37 proposed by Mack and Blumstein, and all but two /w/ amplitude ratio means are below the criterial value of 1.36. More precisely, 96% of speaker 1's tokens and 66% of speaker 2's tokens were correctly discriminated by Mack and Blumstein's metric. (Percentages ranged between 86% and 98% for individual speakers in their experiment.) Thus, although our second speaker does not exhibit perfect invariance,[2] the data confirm the fact that stops generally exhibit large amplitude increases in the region of consonantal release while glides generally exhibit little or no amplitude increase in this region.

II. PERCEPTUAL EXPERIMENT

The general method used for the perceptual study was based on the hypothesis that if the amplitude envelope in the vicinity of the release is an invariant cue to either the continuant–noncontinuant or the stop–glide distinction, then changing this characteristic should have a pronounced effect on perception. Specifically, if stops and glides (or stops and continuants) are distinguished by the amplitude contours described by Mack and Blumstein, then transposition of a stop-vowel amplitude contour onto a glide-vowel syllable should shift the listeners' judgments from glide to stop, and vice versa. To accomplish this manipulation, an amplitude-matching program was used to modify the amplitude contour of a syllable to match that of a template. Tokens to serve as templates were selected by picking syllables closest to the mean ratio values for a given context. For example, the /bi/ syllable for speaker 1 with an amplitude ratio closest to the mean of the speaker's five /bi/ tokens was the template imposed on all /wi/ syllables for speaker 1. The two exceptions to this method were the glide templates for speaker 2's /bi/ and /bu/ tokens: Since the mean amplitudes of the /wi/ and /wu/ syllables for this speaker exceeded Mack and Blumstein's criterial value (see Table I), the glide-vowel syllables with the lowest amplitude ratios were used as templates in these cases. The template and the syllable to be modified were aligned at release onset to preserve the critical amplitude characteristics in the vicinity of the release. Figure 1 displays samples of a /bV/ and /wV/ syllable produced by speaker 2 both before and after amplitude modification. The vertical lines indicate points of release onset and release offset, as defined by Mack and Blumstein. Comparing samples diagonally illustrates that modified tokens had amplitude

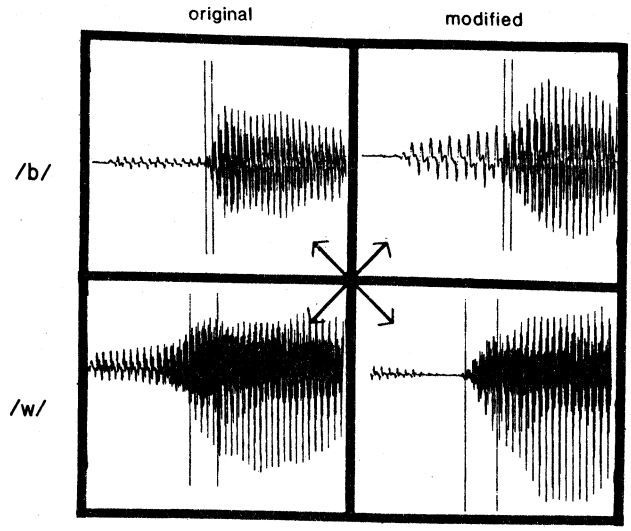

FIG. 1. Samples of a /bV/ and /wV/ syllable produced by speaker 2, both before and after amplitude modification. Vertical lines indicate release onset and release offset, as defined by Mack and Blumstein.

envelopes similar to original tokens of the alternative type. Modified tokens were the length of either the template or the original syllable, whichever was shorter. This manipulation did not create any systematic variation in length because there was no systematic length variation in the original tokens.

To check the effectiveness of the amplitude transposition, amplitude ratios were computed for the modified tokens using the spectrally defined acoustic segments of the original syllables.[3] Mean ratios for modified syllables are given in column 4 of Table I. In all cases, modified /b/ tokens exhibited ratios appropriate for glides (i.e., <1.36), and modified /w/ tokens exhibited ratios appropriate for stops (i.e., >1.37).

A test tape was made consisting of all the original and modified tokens in random order. Using a free identification task (Shinn and Blumstein, 1984), 19 undergraduate students listened to the tape under headphones and wrote the consonant they heard at the beginning of each syllable. The total number of responses for each syllable type—e.g., original /b/ syllables—was 2 speakers $\times$ 25 tokens $\times$ 19 listeners $= 950$. A response was scored "correct" if it matched the consonant of the original syllable. Therefore, if amplitude contour was an invariant cue to the stop/glide distinction, a 100% *error* rate would be predicted for modified syllables. Results are listed in Table II as percentages of total responses identified as the original syllable.

TABLE II. Percentages of responses ($N = 950$) given as *original* syllable, by 19 subjects.

	/b/		/w/	
	Original	Modified	Original	Modified
/i/	100.0	100.0	99.5	97.9
/e/	100.0	98.9	97.4	93.2
/a/	98.9	99.5	95.3	81.6
/o/	99.5	98.9	99.5	90.0
/u/	98.9	99.5	100.0	91.1
$\bar{X}$	99.5	99.4	98.3	90.8

These results indicate that, overall, amplitude transposition had little effect on perception. The mean score across all subjects and tokens was 97% correct recognition. Modifying /b/ syllables to match the amplitude contour of glides had essentially no effect: Five errors were made to original /b/ syllables, and six were made to modified /b/ syllables. The /w/ syllables were identified less accurately overall. Sixteen errors were made to original /w/ syllables (2% of the 950 responses). All errors involved the identification of the consonant as /b/, errors were evenly divided between speakers, and all but two of these errors were made to syllables ending with /a/ or /e/. Thus these 16 errors cannot be attributed to speaker 2's /wi/ and /wu/ tokens which demonstrated amplitude ratios above the 1.36 criterial value. The highest error rate was on the modified /w/ syllables with 88 incorrect responses (9% of total). Of these 88 errors, 35 (40%) occurred for the vowel /a/ and another 36 (41%) were evenly spread between the back vowels /o/ and /u/. Seventy-nine of these errors to modified /w/ syllables (90% of the 88 incorrect responses) consisted of /b/ responses. The other nine errors were spread between /h/ responses and "no consonant heard."

It might be suggested that modifying the amplitude envelopes as we did failed to provide critical information in the glide prevoicing and, therefore, was not an appropriate test of the hypothesis. That is, because speaker 1 did not prevoice his /b/ syllables when these tokens were given /w/ contours, they lacked the prevoicing information normally present in glides; conversely, when his /w/ tokens were given /b/ contours, prevoicing information was destroyed. However, this suggestion can be refuted on both theoretical and empirical grounds. Theoretically, Mack and Blumstein's hypothesis contains no provision for glide prevoicing. Empirically, if this concern were valid, then speaker 1's modified /w/ tokens should strongly bias responses toward "b" since they contain no prevoicing, and speaker 2's modified /w/ tokens should bias responses toward "w" since they retain prevoicing (speaker 2 prevoiced both /b/ and /w/ in his original utterances). In fact, 10% of reponses to speaker 1's modified /w/ tokens and 7% of responses to speaker 2's modified /w/ tokens were "b." Thus there seems to be little difference in responses as a function of prevoicing.

III. DISCUSSION

To a great extent, the acoustic results reported here replicated Mack and Blumstein's (1983): Stop-vowel syllables generally demonstrated large amplitude changes in the vicinity of consonantal release, while glide-vowel syllables demonstrated little or no amplitude change. However, the pattern was not fully invariant for one of two speakers.

In the perceptual experiment, the results were very different from those of Shinn and Blumstein (1984). In the present study, the amplitude envelope certainly did not override other perceptually relevant acoustic attributes of the signal. Consequently, perception of both the continuant–noncontinuant and the stop–glide distinction was affected only slightly by modifying the amplitude envelope. Perhaps the different results of the two studies reflect their different stimulus materials. While we manipulated natural tokens,

Shinn and Blumstein used synthetic stimuli and this may have "...involve(d) failing to manipulate or ignoring critical attributes for stops and glides present in natural speech" (Mack and Blumstein, 1983, p. 1740). As to what these attributes may be, the present study is silent. However, much previous work (e.g., Liberman *et al.*, 1956; Miller and Liberman, 1979; Schwab *et al.*, 1981; Miller and Baer, 1983; Pisoni *et al.*, 1983) suggests that critical information for the syllable-initial stop–glide (/b/ vs /w/) distinction is carried by the rate and extent of formant frequency change over the first 10 to 100 ms. In general, these results support the conclusion of Shinn *et al.* (1985): Many experimental effects disappear when stimuli are made to resemble natural speech. Thus it would seem that natural speech provides the listener with a complex of information for making phonological decisions, any one of which may be severely degraded or inappropriately manipulated without seriously affecting perception.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories, by NIH Grant NS-07237 to the first author (S.N.) through Haskins Laboratories, and by a Spencer Fellowship to the second author (M. S.-K.) through The Center for Advanced Study in the Behavioral Sciences. We are grateful to D. H. Whalen and Philip Rubin for their help with programming, to Mindy Sirlin and her students at Iona College for serving as listeners, and to Bruno Repp, Richard McGowan, Sheila Blumstein, and an anonymous reviewer for their comments on earlier drafts.

[1]Mack and Blumstein computed the rms energy for 15-ms segments, except at stop releases where 5-ms segments were used. Relevant segments were defined as follows:

stops

(1) release onset: that point, clearly visible in the waveform, where aperiodic noise occurs.

(2) release offset: that point, also visible in the waveform, where glottal pulsing begins.

glides

(1) release onset: the portion of the utterance characterized by a "visually perceptible increase in waveform amplitude and/or complexity" (Mack and Blumstein, 1983, p. 1741).

(2) release offset: the point where F2 appears as distinct from F1 in the LPC analysis.

The use of different definitions for stops and glides raises problems which we have ignored. We simply followed the procedures of Mack and Blumstein.

[2]The term "invariance" may be used slightly differently in this paper and in that of Mack and Blumstein. We have used the term in its literal sense to refer to something that "...remains the same irrespective of the context, something that is free of coarticulatory effects..." (Suomi, 1985, p. 268). This may not be the sense applied to the term by Mack and Blumstein. (See Suomi, 1985, for a complete discussion.)

[3]Because the acoustic segments used in computation of amplitude ratios for the original syllables were spectrally defined, it was impossible to *redefine* these segments for the modified syllables in exactly the same way. Modified syllables maintained the spectral characteristics of the original syllables. However, if amplitude ratios were computed for modified syllables using the same absolute time locations as the templates, these ratios obviously would be the same as those of the templates. Thus, no matter how amplitude ratios were computed, they met the criterion of Mack and Blumstein.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Kewley-Port, D. (**1983**). "Time-varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 322–335.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (**1956**). "Tempo of frequency change as a cue for distinguishing classes of speech sounds," J. Exp. Psychol. **52**, 127–137.

Mack, M., and Blumstein, S. E. (**1983**). "Further evidence of acoustic invariance in speech production: The stop-glide contrast," J. Acoust. Soc. Am. **73**, 1739–1750.

Miller, J. L., and Baer, T. (**1983**). "Some effects of speaking rate on the production of /b/ and /w/," J. Acoust. Soc. Am. **73**, 1751–1755.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonants and semi-vowels," Percept. Psychophys. **25**, 457–465.

Pisoni, D. B., Carrell, T. D., and Gans, S. J. (**1983**). "Perception of the duration of rapid spectrum changes in speech and nonspeech signals," Percept. Psychophys. **34**, 314–322.

Schwab, E. C., Sawusch, J. R., and Nusbaum, H. C. (**1981**). "The role of second formant transitions in the stop-semivowel distinction," Percept. Psychophys. **29**, 121–128.

Shinn, P., and Blumstein, S. E. (**1984**). "On the role of the amplitude envelope for the perception of [b] and [w]: Further support for a theory of acoustic invariance," J. Acoust. Soc. Am. **75**, 1243–1252.

Shinn, P., Blumstein, S. E., and Jongman, A. (**1985**). "Limitations of context conditioned effects in the perception of [b] and [w]," Percept. Psychophys. **38**, 397–407.

Stevens, K., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Suomi, K. (**1985**). "The vowel-dependence of gross spectral cues to place of articulation of stop consonants in CV syllables," J. Phonet. **13**, 267–285.