Integration of Acoustic Cues in Spanish Voiced Stops

Sergio Feijóo, Santiago Fernández and Ramón Balsa

Departamento de Física Aplicada, Universidad de Santiago de Compostela, Santiago 15706, Spain

Abstract: In Spanish, voiced stops in initial position usually show the presence of a voice bar before the release burst. In this study we examined the perception of word-initial voiced stops when the voice bar is removed. Removing the voice bar caused 37% of /b/ tokens to be classified as /p/; 41% of /d/'s to be classified as /t/'s ; and 19% of /g/'s to be classified as /k/'s. The change in voicing characteristics was significantly more marked for women. The results indicate that the integration of the voice bar with the burst and the vowel plays a relevant role in the perception of voicing in voiced stops, specially in women's speech, smoothing the dynamic frequency characteristics of burst and vowel.

INTRODUCTION

Numerous studies have shown that the perception of voicing in utterance-initial stop consonants is primarily related to voice onset time (1). In Spanish that difference is characterized by positive VOT's for voiceless stops and negative VOT's for voiced stops. In other languages such as English, positive VOT's are common in voiced stops (2). Negative VOT's imply the presence of a voice bar before the release burst, since the onset of voicing precedes the release of the consonant. In this study the perception of word-initial voiced stops when the voice bar is removed is examined, in order to asses its importance in the perception of voicing and its relation with acoustic features that are associated with the voicing contrast in English, namely VOT, first formant transition and onset value and the fundamental frequency contour and its onset value (3).

PERCEPTUAL AND ACOUSTIC ANALYSIS

The subjects were 17 University students with normal hearing who had volunteered for the experiments. Two conditions were considered: 1) In the first condition subjects heard the whole voice bar + 102.4 ms of signal from burst onset; 2) In the second condition the voice bar was removed and subjects were presented only with the 102.4 ms of signal starting at burst onset. In both experiments subjects were instructed to identify the consonant they heard as one of the 6 stops, /p/, /t/, /k/, /b/, /d/, /g/, including the option "another sound" in case they could not recognize the consonant as any particular stop. The tokens were a series of natural Spanish words whose initial syllable was formed by the combination of /b/, /d/ or /g/ with /a/, /e/, /i/, /o/, /u/. 9 men and 9 women served as speakers. The confusion matrices for both conditions can be seen in Table 1. In condition 1 94.51% of the tokens were correctly recognized as /b/, /d/ or /g/. In condition 2 only 54.44% of the tokens were recognized as a voiced stop. Condition 2 shows a change in the voicing characteristics of the tokens as perceived by the listeners: 36.73% of the /b/ tokens are perceived as /p/'s; 41.24% of /d/'s are perceived as /t/'s; and 19.28% of /g/'s are identified as /k/'s. In condition 2, the change in the perceived voicing characteristics is more marked for women: only 43.2% of their tokens are perceived as voiced, against a 65.7% for men. Responses were analysed separately for each condition. For condition 1 the ANOVA showed no significant effects. In condition 2 the ANOVA showed significant main effects of stop ($F(2, 267) = 52.86, p < 0.0005$); vowel ($F(4, 265) = 4.90, p < 0.001$); and sex ($F(1, 268) = 70.09, p < 0.0005$); plus a significant stop $\times$ vowel interaction ($F(8, 261) = 4.42, p < 0.0005$). The perceptual analysis also revealed that in condition 2 the tokens were perceived as the voiced stop, as its voiceless counterpart or as ambiguous signal.

It is well known that contiguous segments in the speech signal are combined to produce a certain perceptual effect in speech. In the case of Spanish voiced stops, the voice bar is combined with the following burst and vowel in order to produce a certain percept of the voicing characteristics. The overall effect of adding the voice bar is the presence of a strong low frequency component. Since the signals used in the experiments of

TABLE 1. Confusion matrices (in percent). Top: voice bar plus 102.4 ms from burst onset, bottom: 102.4 ms from burst onset.

	/b/	/d/	/g/	/p/	/t/	/k/	Other
/b/	94.57	2.61	0.59	0.72	0.06	0.00	1.44
/d/	1.70	93.86	1.70	0.06	1.57	0.06	1.05
/g/	1.76	0.52	95.10	0.33	0.06	1.11	1.11
/b/	42.16	0.59	2.22	36.73	0.98	0.06	17.25
/d/	1.76	47.32	3.59	3.20	41.24	1.63	1.24
/g/	0.85	0.65	73.86	1.44	1.31	19.28	2.61

condition 2 seem to have different voicing characteristics, we decided to study whether those differences could be explained by the acoustic variables. The variables measured were: burst duration (from burst onset until the vocalic onset), F1 of the first vocalic pulse (onset F1), difference between the value of F1 corresponding to the first pulse and fifth pulses (F1 transition), difference between the fundamental frequency of the first and fifth pulse (F0 contour). For condition 2 signals were divided into groups according to their perceived voicing characteristics. A histogram of the number of correct responses shows two extreme peaks and a central normal distribution. Those cases into the lower distribution were considered "voiceless", those in the higher distribution were considered "voiced", and those within one standard deviation of the central distribution were considered "ambiguous". A oneway analysis of variance was performed with each variable as the depending factor and voicing condition ("voiced", "voiceless", "ambiguous") as the independent factor. The analysis showed a significant effect for burst duration ($F(2,206) = 8.44, p < 0.0003$): burst duration of "ambiguous" signals is shorter than burst duration of the "voiced" and "voiceless" signals. The effect of the onset F1 was also significant ($F(2,232) = 6,47, p < 0.002$): onset F1 frequency is higher for "voiceless", intermediate for "ambiguous" and lower for "voiced" signals. The effect of F0 contour was also significant ($F(2,232) = 4.53, p < 0.01$): F0 contour of "ambiguous" signals is different from those of "voiced" and "voiceless" signals. Only for onset F1 the effect was maintained for each stop separately.

DISCUSSION

The perceptual results show that in Spanish the presence of the voice bar before the release of voiced stops helps in the perception of voicing, especially for women's speech, for which the spectral characteristics are shifted upwards in comparison with those of men. The perceptual data also point to differences in the acoustic realization of voiced stops, since for some of them voicing is correctly perceived even in the absence of the voice bar, while others are perceived as the corresponding voiceless stop or as an ambiguous signal. Differences among those three types of voiced stops across the three places of articulation are signaled especially by the F1 onset frequency. It is evident from the present data that the the voice bar is integrated with the following signal to produce a clear voicing percept, introducing a low frequency component that, in some way, compensates the frequency characteristics of the onset F1 frequency, which seems to be the stronger acoustic correlate of voicing when the voice bar is absent.

REFERENCES

1. Whalen, D.H., Abramson, A.S., Lisker, L. and Moody, M., "F0 gives voicing information even with unambiguous voice onset times," *J.Acoust.Soc.Am.* **93**, 2152–2159 (1993).
2. Lisker, L. and Abramson, A.S., "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 385–422 (1964).
3. Repp, B.H., "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," *Psychological Bulletin* **92**, 81–110 (1982).