From: Wolf, Jared J. & Dennis H. Klatt (eds.), 1979. Speech Communication Papers. Presented at the 97th Meeting of the Acoustical Society of America. Massachusetts Institute of Technology, Cambridge, Mass., 12-16 June 1979. 18

PASSIVE VOCAL TRACT ENLARGEMENT DURING VOICED STOPS

John J. Ohala and Carol J. Riordan, Phonology Laboratory,
Department of Linguistics, University of California, Berkeley, CA 94720

Voiced stops should become devoiced within 5 to 10 msec of stop closure if there is no vocal tract enlargement to delay the inevitable reduction in transglottal pressure drop. Since stops may be voiced longer than this, some cavity expansion must take place, either passive through tissue compliance or active through larynx lowering and the like. To estimate the duration of voicing in stops when only passive enlargement occurs, subjects produced isolated nonsense words of the form VC:V, where V was one of a variety of English and C: was an artificially prolonged /b/, /d/, or /g/. Oral air pressure was vented through a catheter leading from the pharynx to the atmosphere via the nasal cavity. At unpredictable times, a solenoid-activated valve closed the catheter and the consequent build-up of oral pressure extinguished the voicing. Voicing continued after this closure longer during /b/ than /d/ or /g/, and longer (in most cases) when coarticulated with high vowels than with low vowels. These results can most plausibly be explained by reference to differences in compliance of the surfaces on which oral pressure impinges during the various consonants and vowels.

Introduction

One of the prerequisites for voicing is an adequate transglottal pressure drop, estimated at approximately 2 to 3 cm H_2O (Isshiki 1959, Catford 1977:29). Obstruents, by their very nature, require a build-up of oral pressure which effectively reduces the transglottal pressure drop. Accordingly, as has long been known, voiced obstruents have a tendency to devoice. There is phonological evidence, however, that the tendency to devoice is not equal across all types of obstruents. Thus long obstruents have a greater tendency to devoice than short obstruents (Passy 1890:161). One of the ways this tendency manifests itself in languages is the greater incidence of voiceless long obstruents (so-called 'geminates') as opposed to voiced long obstruents (Jaeger 1978). Japanese, for example, admits in its segment inventory /p:, t:, k:, s:/ but not /b:, d:, g:, z:/, even though as short segments all are admissable, i.e., /p, t, k, s, b, d, g, z/.

Apparently, the place of articulation of the obstruent also makes a difference. In Nubian, according to Bell (1971) and personal fieldwork by the first author, regular grammatical processes that cause stem-final stops to become long, change final /g, d͡ʒ, d/ to /k:, t͡ʃ:, t:/ not /g:, d͡ʒ:, d:/. Final /b/, however, appears voiced when made long, i.e., as /b:/ not /p:/. See Table 1.

Table 1. Phonological data from Nubian.

English equivalent	Noun stem	Stem + 'and'
father	/fab/	/fab:ɔn/
scorpion	/sɛgɛd/	/sɛgɛt:ɔn/
donkey	/kad͡ʒ/	/kat͡ʃ:ɔn/
dog	/mʊg/	/mʊk:ɔn/

Related to this is the well-known constraint affecting voiced implosives and (plain) voiced stops, such that they are favored at forward places of articulation and disfavored at back places of articulation (Chao 1936, Haudricourt 1950, Greenberg 1970, Gamkrelidze 1975, Sherman 1975, Pinkerton 1979).

Finally, certain sound changes in Bantu languages suggest that the vocalic environment which the voiced stop is produced in may affect its tendency to devoice (or at least to retain both of the simultaneous features [+voice] and [+stop]). The data in Table 2 show that in certain modern Bantu languages a Proto-Bantu */b/ remained intact in all vocalic environments, the PB */d/ did so only before high vowels, and PB */g/ was not retained in any vocalic environment.

The effect of place of articulation upon the chances of a stop to retain its voicing has previously been attributed to the different volumes of the oral cavities involved (Martinet 1955). Smith (1977) and Smith and Westbury (1975) applied the same argument to variations found as a function of the quality of the vowel in the environment of the stops (in their study of variations of voice onset time in English stops), assuming that the vocal tract configuration

Table 2. Data from Bantu languages. From Guthrie (1967-1970); transcription simplified.

Proto-Bantu stem	Reflex in Duala	Reflex in Ngɔm	English equivalent
*-bi		-ɓe	*bad*
*-bua	ɗi-ɓua		*nine*
*-beede	ɗi-ɓɛ	re-ɓɛð	*breast*
*-bab	-babisɛ		*singe*
*-bod	-ɓɔ-	-ɓoð	*become rotten*
*-dib-	-ɗiɓ-	-ɗiɓ	*shut*
*-dug	-ɗu-	-ɗuk-	*paddle*
*-dob-	-ɔɓ-	-ðɔɓ-	*fish with line*
*-daad-		-ðað-	*lie down*
*-giko	ɗ-iɔ		*fireplace*
*-godi	m-ɔɗi	ŋ-kɔli	*string*
*-geda	e-yɛi		*iron*
*-gag-		-kak-	*go bad*
*-gomb-	-ɔmb-		*scrape*

appropriate to the vowel is coarticulated during the production of the stop closure. They noted that the vocal tract is larger during the production of high vowels as opposed to low vowels, primarily because the high vowels /i, u/ have a greatly enlarged pharyngeal cavity.

Mathematical models of speech aerodynamics show, however, that vocal tract volume *per se* cannot be a significant factor in the maintenance of voicing during stops (Rothenberg 1968, Ohala 1975, 1976, Catford 1977:74). Given reasonable estimates of vocal tract volume (30 to 150 cm^3), glottal air flow rate (70 to 120 cm^3/sec), and sublglottal air pressure (5 to 15 cm H_2O), voicing should be extinguished within 10 to 15 msec of stop closure onset in any case. With more conservative estimates of the above parameters, voicing should stop within 5 to 10 msec.

Since there are many cases where voicing during stops lasts longer than this, there must be some other way the oral cavity accomodates glottal air flow without raising oral air pressure past the point where voicing would be threatened. This is done by vocal tract expansion (Chao 1936, Javkin 1977). This expansion can be done passively through tissue compliance or actively by jaw lowering, velum raising, larynx lowering and the like.

It would be of some interest to know approximately how long voicing could persist in stops if only passive enlargement of the vocal tract were involved, i.e., without the speaker having to actively employ physiological mechanisms which compensate for the aerodynamic constraints. It would presumably be these limits which would account for the cross-language tendencies cited above. We undertook the following study to get data on this question.

Method

We sought some way of having subjects produce voiced stops coarticulated with various vowels, where active vocal tract enlargement could not play a role in the maintenance of voicing. We therefore devised the following procedure.

Subjects produced utterances of the sort $V_1C{:}V_2$ where $V_1=V_2$ was any of the American English vowels /i, u, ej, a/ and C: was an abnormally long /b, d, g/. Oral air pressure during the stop was vented to atmosphere by a nasal catheter (40 cm long, 3 mm inner diameter) positioned in the upper pharynx. The outer end of the catheter could be quickly closed by a solenoid-activated valve. During the stop closure the valve would be closed at unpredictable moments (by a second person). In no case, however, was the over-long stop closure held for more than 5 sec before the valve was shut.

The catheter itself added less than 3 cm^3 to the effective volume of the vocal tract and had neglible compliance.

Putting the vowels in intervocalic position and having the speakers release the consonant into the second vowel after the valve was closed helped to insure that the stop was coarticulated with the vocal tract configuration of the surrounding vowels.

Oral air pressure was recorded on an FM tape recorder along with the signals from a throat microphone and a second microphone placed approximately 10 cm from the speaker's mouth. See Figure 1.

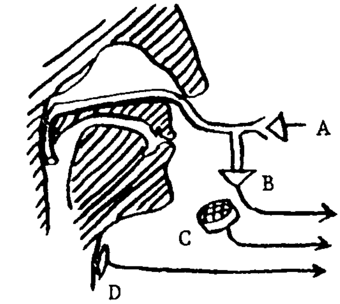

Figure 1. Schematic representation of experimental set-up. A: solenoid-activated valve, B: pressure transducer, C: microphone, D: throat microphone.

Measurements were done by hand on an oscillographic representation of the pressure signal and a rectified, integrated (time constant = 10 msec) version of the signal from the throat microphone. The duration of voicing (DV) was defined as the time from the abrupt rise in oral pressure to the last detectable glottal pulse in the voice signal. See Figure 2.

Results

We report here the results from one adult male subject. Processing of data from further subjects is underway.

There was some back pressure during the stop even when the valve was open due to the air resistance offered by the nasal catheter. This averaged 1.7 cm H_2O and was not significantly different between consonant and vowel conditions. This could mean, however, that the resulting DV would be slightly but consistently underestimated. Offsetting this was the finding that peak oral pressure during the stops, once the valve was closed and voicing had ceased, ranged from 13 to 16 cm H_2O. We can assume peak oral pressure equaled subglottal pressure in this situation. This was a higher-than-normal range for subglottal pressure and possibly points to this speaker trying to partially compensate for the steady back pressure felt during the valve-open portions of the stops. (This does not indicate, though, that the speaker employed active compensatory mechanisms to maintain voicing once the valve was closed and, we think, the relatively short DV's actually found are consistent with only passive tract enlargement taking place.)

The distribution of the DV's in most conditions was positively skewed, i.e., there were a few very long DV's (of the order of 200 msec). Consequently, the reported standard deviations must be read with caution. Because of the skewing we report the median values in addition to the mean. These data are given in Table 3. Summary statistics for the individual consonants and vowels are given in Table 4.

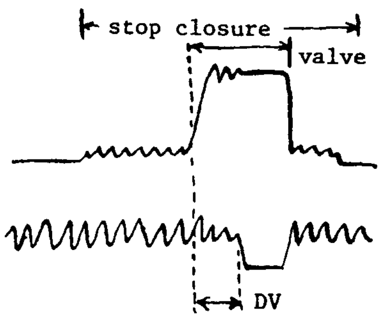

Figure 2. Schematic representation of how 'Duration of Voicing' (DV) was determined.

Discussion

Two main effects are evident in the data. First, the DV was longer the more forward the place of articulation of the stop. Second, the DV was longer the higher the vowel the stop was coarticulated with. There were two exceptions to this second generalization, viz., the DV's for /aba/ and /ada/ which were quite long in spite of the low vowel condition.

The data parallel quite strikingly the phonological data cited above. The data are also in agreement with the observations of Smith and others and with the claim that duration of voicing in stops is proportional to the volume of the oral cavity between the consonantal closure and the glottis. We noted above, though, that volume *per se* cannot account for most of the systematic variation actually found. The more likely candidate, we believe, for the factor which determines DV in stops is the yielding or compliance of the walls of the vocal tract. Of course, vocal tract surface area will be monotonically related to vocal tract volume. Of some importance, though, is the nature of the surfaces exposed to the impinging air pressure. Some surfaces (the cheeks, tongue, velum, pharyngeal walls) have very high compliance; other surfaces (hard palate, teeth) have very low or essentially zero compliance. Reliable quantitative data on the areas of these various surfaces and the values of their respective compliances (in an "average" speaker) are not available to us. Qualitatively, however, we can explain the trends in the data as follows. Labial stops have the greatest vocal tract surface area and thus have the greatest DV's. Only during labial stops coarticulated with low vowels, though, are the highly compliant lateral (as opposed to anterior) cheek walls exposed to the oral pressure (or at least so it seems to us upon kinesthetic introspection). This would account for /aba/ having the longest median DV. In the case of the other places of articulation, the cheeks are not involved and the DV should be proportional to vowel height since, as noted by Smith, the higher the vowel, the more enlarged is the pharyngeal cavity. The greater the pharyngeal volume, the greater its surface area.

Table 3. Statistics on Duration of Voicing in Various VCV Utterances (values in msec).

Utterance Type	Sample Size	Mean	Median	Standard Deviation
ibi	20	101.14	86.36	35.00
idi	19	65.27	67.55	8.86
igi	19	62.18	59.64	18.82
ebe	19	79.64	69.73	33.91
ede	19	64.82	60.91	19.00
ege	20	50.45	47.73	13.73
ubu	18	93.18	88.64	32.91
udu	20	68.91	61.36	23.64
ugu	20	52.73	53.82	15.09
aba	18	97.91	90.91	45.73
ada	20	75.45	61.36	42.73
aga	20	50.91	46.18	19.00

Conclusion

More data is needed, of course, but we can tentatively conclude that inductive and deductive evidence exists which shows that duration of voicing in stops, in the absence of any active vocal tract enlargement, will average about 70 msec with the exact value being proportional to the total compliance of the vocal tract walls which, in turn, will be equal to the product of the areas of the separate parts of the vocal tract surfaces exposed to the impinging oral pressure and their respective compliances. Many cross-language sound patterns involving voicing and devoicing of stops can be explained by reference to these relations.

Acknowledgements

This research was supported by the National Science Foundation and by the National Institutes of Health via a post-doctoral fellowship to the second author.

Table 4. Summary statistics (values in msec).

Condition	Sample	Mean	Median	Stan. Dev.
/i/	58	76.64	69.81	29.45
/a/	58	73.82	59.85	41.00
/u/	58	70.82	63.64	29.27
/e/	58	64.73	60.12	26.18
/b/	75	92.91	81.82	36.91
/d/	78	68.73	63.18	26.64
/g/	79	54.00	51.89	17.19
All	232	71.51	64.02	32.09

References

Bell, H. 1971. The phonology of Nobiin Nubian. African Lang. Rev. 9.115-139.
Catford, J. C. 1977. Fundamental problems in phonetics. Indiana Univ. Press.
Chao, Y. R. 1936. Types of plosives in Chinese. Proc. 2nd Int'l. Cong. of Phon. Sci. Cambridge Univ. Press. Pp 106-110.
Gamkrelidze, T. V. 1975. On the correlation of stops and fricatives in a phonological system. Lingua 35.231-262.
Greenberg, J. H. 1970. Some generalizations concerning glottalic consonants, especially implosives. IJAL 36.123-145.
Guthrie, M. 1967-1970. Comparative Bantu. Gregg.
Haudricourt, A. G. 1950. Les consonnes préglottalisées en Indochine. Bull. Soc. Ling. de Paris 46.172-182.
Isshiki, N. 1959. Regulatory mechanisms of the pitch and volume of voice. Oto-rhino-laryngology Clinic, Kyoto. 52.1065-1094.
Jaeger, J.J. 1978. Speech aerodynamics and phonological universals. Berk. Ling. Soc. 4.311-329.
Javkin, H. 1977. Towards a phonetic explanation for universal preferences in implosives and ejectives. Berk. Ling. Soc. 3.559-565.
Martinet, A. 1955. Economie des changements phonétique. Berne: Editions A. Francke.
Ohala, J. J. 1975. A mathematical model of speech aerodynamics. In: G. Fant (ed.) Speech Communication. Vol. 2. Stockholm: Almqvist & Wiksell. Pp. 65-72.
Ohala, J. J. 1976. A model of speech aerodynamics. Rep. of the Phonology Lab. 1.93-107.
Passy, P. 1890. Etude sur les changements phonétiques. Paris: Librairie Firmin-Didot.
Pinkerton, S. 1979. Quichean (Mayan) glottalized and non-glottalized stops: a phonetic study with implications for phonological universals. 9th Int'l. Cong. Phon. Sci.
Rothenberg, M. 1968. The breath-stream dynamics... Bibliotheca Phonetica 6.
Sherman, D. 1975. Stops and fricative systems. Working Papers in Lang. Universals 17.1-31.
Smith, B. L. 1977. Effects of vocalic context... West Coast Phonetics Symposium, Santa Barbara.
Smith, B. L. & Westbury, J. R. 1975. Temporal control of voicing... 89th ASA meeting, Austin.