MULTIPLE FUNCTIONS OF F0 IN SPECIFYING THE VOICING, QUANTITY AND ASPIRATION OF EAST BENGALI STOPS AND AFFRICATES.

Henning Reetz
Simone Mikuteit
Aditi Lahiri

reetz@em.uni-frankfurt.de, simone@simonemikuteit.com, aditi.lahiri@ling-phil.ox.ac.uk

ABSTRACT

Bengali exhibits a four way phonemic contrast in voicing and aspiration, allowing distinctions like e.g. [p, pʰ, b, bʰ]. These distinctions cut across stops as well as affricates in five places of articulation (labial, dental, retroflex, velar for stops and palate-alveolar for affricates), within singletons (in all word positions) and geminates (only word medial). The voicing distinction of stops and affricates is primarily manifested by the presence or absence of glottal pulsing during oral closure, which is generally sustained throughout the entire length of the closure or at least continues for a considerable duration of the whole closure length (cf. [4]). This paper investigates East Bengali stops and affricates, focusing on the correlation of F0-perturbations with respect to the voice, quantity (singleton~geminate) and aspiration contrast in different prosodic environments for sentence (level) and question (rising) intonation.

Keywords: F0, Tonogenesis, VOT, Aspiration, Bengali

1. INTRODUCTION

It has been known for a long time that micro-prosodic F0-perturbations support the voice(less)-ness of consonants [3] and that the presence of a higher or lower F0 after voiceless and voiced stops lead to the development of tones in certain languages, first suggested by Haudricourt [1]. Given that F0 was of primary importance in our investigation, two different speech modes were examined: declarative sentences (level intonation) and yes/no questions. Consequently we could compare F0 differences across different intonation types. In Standard Colloquial Bengali (SCB) spoken in Calcutta, India, the Yes-No question is known to end with a rising-falling LHL (low-high-low) contour [2]. One of the issues, therefore, would be to see how the intonation contour was realized in East Bengal (EB) but more importantly how it interacted with F0 realisations in a voicing contrast.

In a production study we analysed the F0 micro-contours of two female and two male speakers after the release under the condition of level and yes/no question intonation contours of 172 words. The results provide evidence for a distribution of F0 onsets that contrast voiceless being higher than voiced ones, geminates higher than singletons, and aspiration was found to have a general lowering effect on F0.

2. METHODOLOGY

2.1. Stimuli

Five different words were chosen for each of the 35 phonemes listed in Table 1. (However, for the voiced aspirated retroflex stops only two words could be found.) A total of 172 words we tested with four speakers using two intonation patterns. The material containing the crucial segments in intervocalic environment (CVC) consisted of 145 disyllabic, 22 trisyllabic and 5 four-syllable Bengali words. In 166 of the cases the vowel following the crucial segments was located in the second syllable of the word and in 6 of the cases it was in the third syllable. The preceding vowels were /i, u, o, ɔ, e, æ, a/. The following vowels on which we performed the F0 measurements were /o, ɔ, e, æ, a/, but not /i/ or /u/ to reduce confounding with the intrinsic higher F0, which is generally attributed to these high vowels. The two contrastive prosodic environments were (i) neutral intonation (level intonation), which was tested in four repetitions and (ii) yes/no question intonation, which was tested in three repetitions. Thus the whole data comprised 4816 words.

Table 1: Stop and affricate inventory of EB.

labial	dental	retroflex	palato-alveolar	velar
p / b	t̪ / d̪	ʈ / ɖ	tʃ / dʒ	k / g
- / bʰ	t̪ʰ / d̪ʰ	ʈʰ / ɖʰ	tʃʰ / dʒʰ	kʰ / gʰ
pː / bː	t̪ː / d̪ː	ʈː / -	tʃː / dʒː	kː / gː
- / bːʰ	t̪ːʰ / d̪ːʰ	- / ɖːʰ	tʃːʰ / dʒːʰ	kːʰ / -

2.2. Participants

Four paid native speakers of East Bengali (Dhaka dialect, which is usually considered to be the standard in Bangladesh) – two male and two female (henceforth M1, M2, F1, F2) aged between 20 and 30 with no report of speech disorders – participated in the experiment. The recordings were conducted in Dhaka/Bangladesh with graduates from the university of Dhaka who had learnt English as a second language.

2.3. Procedure

The stimuli were presented on four differently randomized lists on paper, corresponding to four repetitions. During the first four repetitions the subjects were instructed to read each word at normal utterance speed in a flat (monotone) intonation. However, in a few cases a slight effect of list reading was not avoidable, which resulted in a minor upward movement of pitch on the last syllable of each word. The repetitions from list one to three were presented again but this time subjects had to pronounce each word with a yes/no question intonation. They were instructed to imagine a situation where someone said a word that they did not understand and they would repeat the word they thought they heard, to find out if they did hear it correctly. All subjects uttered the said intonation contour very consistently. In the rare case of an improperly produced contour the respective word was noted down by the experimenter and presented again for a separate reading at the end of a repetition.

In order to measure F0, pulse marks for each glottal pulse were added (manually set with a semi-automatic assistance) at the relevant points of the signal – starting at the onset of regular glottal pulsing after ACT. The duration of each pulse was converted into frequency. By connecting the frequency values, a curve is obtained which shows the changes of fundamental frequency from one glottal pulse to the next.

3. ANALYSIS

3.1. Design

The F0 perturbations were analysed for the first six glottal pulses after ACT. In order to ensure the L-H intonational contour of the yes/no question across the crucial segments, six words that did not show the appropriate prosodic pattern were excluded from the set with question intonation. An ANOVA of F0 was performed with the following factors in a complex model: *voice* (voiced, voiceless), *quantity* (singleton, geminate), *intonation* (level, question), *aspiration* (aspirated, unaspirated), *glottal pulse* (1, 2, 3, 4, 5, 6 – treated as a nominal variable), *place* (labial, dental, velar, palato-alveolar/affricate) and *participants* (M1, M2, F1, F2) as a random factor.

In terms of place of articulation, an initial analysis had shown that the F0 curves of the retroflex stops exhibited irregular and inconsistent patterns with respect to voice and consonant quantity. These irregularities could not be attributed to a particular feature or to a particular subject. Since none of the obstruents in the other places of articulation showed such an inconsistency, the retroflexes were excluded from the calculations. Further were items excluded, where no F0 was computed reliably, items that were erroneously pronounced, and those whose F0 trajectory was recognized as an outlayer (e.g. octave jumps). The remaining number of tokens, which went into the analysis, was 3841.

3.2. Results

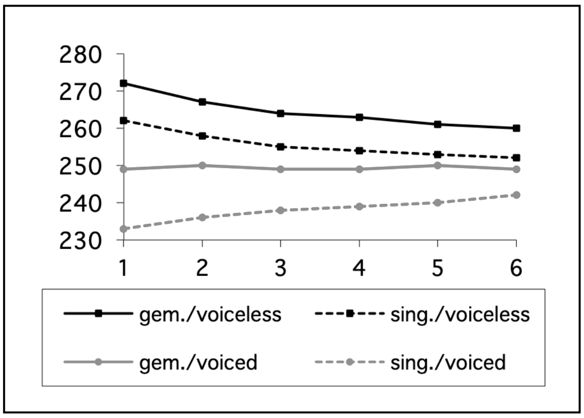
There was a significant overall effect for *voice* $F(1,3)=46.9$; $p\leq 0.006$, *quantity* $F(1,22804)=1025.0$; $p\leq 0.001$, *intonation* $F(1,3)=58.8$; $p\leq 0.005$, *aspiration* $F(1,3)=11.6$; $p\leq 0.042$ and *place* $F(3,9)=11.6$; $p\leq 0.002$. F0 was found to be higher in voiceless than in voiced obstruents, it was higher in geminates as compared to singletons, it was higher for the unaspirates than for the aspirates. In terms of place of articulation, a post hoc test showed that F0 was highest in the palato-alveolar affricates as compared to the stops which did not differ significantly from each other (labial vs. dental *n.s*; labial vs. velar *n.s*; labial vs. palato-alveolar affricates $p=0.0008$; dental vs. velar *n.s*; dental vs. palato-alveolar affricates $p=0.0006$; velar vs. palato-alveolar affricates $p=0.0086$). There was a significant interaction between aspiration and intonation $F(1,22804)=322.6$; $p\leq 0.001$ indicating that the aspirates had a significantly lower F0 in the question condition ($p=0$) but not in the neutral condition ($p=0.41$). A post hoc test on the interaction between aspiration, intonation and pulse (which was at the border of significance at $F(5,22804)=2.14$; $p\leq 0.058$) showed that the aspirates had a significantly lower F0 when produced with question intonation along all six glottal pulses, but in neutral intonation, significance was met only at pulses 1 and 2. Another significant interaction occurred between place and voice. In the voiced condition, a post hoc test of the dental and labial stops against the velar stops revealed a significant difference, as well as the stops (labial, dental and velar) against the affricates (palato-alveolar). In the voiceless condition the post hoc comparison revealed only a significant difference for the stops against the affricates but not for the within stop comparison.

4. DISCUSSION

4.1. Voice and Quantity

The curves corresponding to the voiced obstruents as well as to the singletons had a lower fundamental frequency than the related counterparts. This suggests, that the F0 perturbations in Bengali have multiple functions. They serve as an acoustic parameter for the distinction of voice and additionally as a parameter for the distinction of consonant quantity. This issue becomes transparent if these two parameters are displayed in one graph (Fig. 1). The highest F0 is found with the voiceless geminates, next is the curve of the voiceless singletons, located below are the voiced geminates and the lowest F0 is found with the voiced singletons. This four-way division of F0 perturbations in respect to voice and quantity was found in principle across all subjects. The individual realisations, however, varied between the four speakers.

Figure 1: F0 contours of the first 6 glottal pulses after the ACT of voiced and voiceless singleton and geminate obstruents.



4.2. Place and manner

We found higher F0 values for the voiced and voiceless palato-alveolar affricates than for any stops. The voiced velars had slightly higher values than the labials and dentals but in the voiceless series they showed a strong tendency to merge with the other stops.

4.3. Aspiration

Aspiration had the tendency to lower F0. While this lowering effect disappeared after the second glottal pulse in neutral intonation it was present throughout in question intonation. In a similar fashion the lowering induced on F0 by aspiration was stronger in the voiced series than in the voiceless one. Although the overall results showed significance in both cases, a comparison along each of the six glottal pulses revealed that in the voiced series significance was met from pulse 1 trough to pulse 4, while in the voiceless series only the first two glottal pulses had a significant difference. This behaviour can be accounted for by the strong pitch depressing influence in the production of stops that are both voiced and aspirated.

In respect to intonation a closer inspection of the curves seems to indicate that the lowering basically consists of two distinct types. (i) There is a trajectory downward found mainly on the first glottal cycle, which appears in neutral as well as in question intonation. (ii) In question intonation the whole curves of the aspirates are generally lowered and shifted downwards in relation to the unaspirates, while in neutral intonation they tend to merge with the unaspirates shortly after the aforesaid downward fold during the first or the first two glottal cycles.

4.4. Intonation

Speakers seem to strengthen the high-low relations of F0 of the voice, quantity and aspiration distinctions in question intonation (or at high pitch level) by making the high-low onset differences much more expressed. At the same time, speakers do not produce this spreading of F0 onsets for the place/manner distinctions. The difference between the highest and lowest F0 onset was always much smaller in neutral intonation than in question intonation (average for voice and quantity: 24 Hz~ 2.15 semitones (st), for neutral intonation, 53 Hz~ 2.95 st for question intonation).

The same effect of spreading could be observed in respect to aspiration. The highest versus lowest F0 onset difference in neutral intonation was 30 Hz (2.67 st) as opposed to 62 Hz (3.23 st) in question intonation. For place no comparable spreading effect is found in both intonation conditions.

It is well established that listeners' sensitivity to frequency changes with the absolute value of F0. Even after transferring the distances from Hertz to semitones, there was still a significant spreading effect in the question intonation for the voice, quantity and aspiration contrasts, but not for the place contrasts. In the first three cases it is possible that the speakers allow for this effect by actively enlarging the differences of the F0 onsets at higher utterance pitch. A second and more profound reason for this spreading is, that the speaker needs to establish a clear high or low distinction in reference to the intended normal or unperturbed pitch that is to be expected after the 'silent rise' during the closure. In other words: A difference of 10 Hertz can be detected rather well if the intonation is more or less kept at level. However, if the overall pitch contour is rising, and especially if the main part of the rise takes place during the silent gap of the closure, the speaker needs to establish an unambiguous high or low reference in relation to the intended contour, since the listener does not know what pitch *exactly* will be reached at the point after release. The speaker assists the listener by setting the F0 onsets well above or below the approximately expected pitch.

These results support our hypothesis that F0 perturbations are at least partially governed by planned laryngeal adjustments and are not mere products of automatic processes. Our results are consistent with the view that the F0 spreading in question intonation is triggered by the speakers' inherent knowledge about perceptual factors. In the same way, we can only explain the effects of quantity distinction upon F0 by the assumption of an active adjustment. As described above the same articulatory gesture differing only in timing produced different values of F0. There is no reasonable acoustic explanation why F0 should be higher for geminates, both voiceless and voiced ones. The argument that geminates produce a higher F0 because there is a higher level of tension of the vocal folds due to the longer closure duration could only hold for the voiceless ones. In order to maintain voicing during oral closure in the voiced stops, the pressure drop across the glottis must be held a certain level in order to allow the vocal folds to continue to vibrate. This is achieved by expanding the volume of the vocal tract. One of the means to obtain such an expansion is to lower the larynx. This movement generally causes a downward tilt of the cricoid cartilage, which in turn causes the vocal folds to shorten and to slacken. The carrying over of the slack vocal folds condition into the initial part of the following vowel will result in a lowered F0 [5]. If we imagine that the longer closure of a geminate even reinforces this strategy we would expect to find – as a natural consequence – an F0 that is lower in voiced geminates than in the corresponding singletons. This, however, is opposite to what was found in the results. We must therefore infer that the speakers manipulate F0 height in one way or the other in order to establish a uniform high-low F0 relation in respect to quantity.

4. SUMMARY AND CONCLUSION

The F0 excursion after voiceless obstruents are generally higher than after voiceless ones, geminates are higher than singletons, and aspiration is a general F0 depressor (see Table 2). We argued that the F0-depression by the aspiration is probably based on automatic consequences of aerodynamic conditions and a specific laryngeal setup, whereas the other effects might be under active speaker's control.

The correlation with place of articulation was that the F0 curve of the palatal-alveolar affricates was generally higher and separate from the curves of the stops. Within the stop category there was a weak tendency for velars to exhibit a higher F0 than the labials and dentals.

In terms of intonation, there was a much larger separation of F0 onsets for the contrasts of voice, quantity and aspiration at high pitch as compared to low pitch. We attribute this effect to a controlled adjustment by the speakers. As for the voicing contrast, a comparison of closure durations showed that this spreading effect cannot be an automatic consequence of the longer closure that is generally found with voiceless stops.

Table 2: Average F0 values of first glottal pulse after the ACT.

Voicing	Quantity	Aspiration
Voiceless (267 Hz)	Geminate (270 Hz)	Unaspirated (282 Hz)
		Aspirated (261 Hz)
	Singleton (261 Hz)	Unaspirated (274 Hz)
		Aspirated (251 Hz)
Voiced (241 Hz)	Geminate (250 Hz)	Unaspirated (257 Hz)
		Aspirated (242 Hz)
	Singleton (233 Hz)	Unaspirated (245 Hz)
		Aspirated (221 Hz)

7. REFERENCES

[1] Haudricourt, A.–G. 1954. De l'origine des tons en vietnamien, *Journal Asiatique* 242, 69–82.

[2] Hayes, B., Lahiri, A. 1991. Bengali intonational phonology, *Natural Language and Linguistic Theory* 9, 47–96.

[3] House, A. S., & Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels, *The Journal of the Acoustical Society of America* 25, 105–113.

[4] Mikuteit, S., Reetz, H. 2007. Caught in the ACT: the timing of aspiration and voicing in East Bengali, *Language & Speech* 50, 247–277.

[5] Stevens, K. N. 1998. *Acoustic Phonetics*. Cambridge MA: MIT Press.