THE ROLE OF INTENSITY IN BREATHY VOICED STOPS: A CLOSE LINK BETWEEN PRODUCTION AND PERCEPTION

LIESELOTTE SCHIEFER

Institut für Phonetik und Sprachliche Kommunikation
der Ludwig-Maximilians Universität München, FRG

ABSTRACT

The acoustic analyis of the intensity difference between the breathy and steady vowel portions following a breathy voiced stop in Hindi revealed a significantly lower intensity in the breathy portion compared with the steady one with a clear influence of the vowel. The perceptual importance of that acoustic cue was tested with 4 CV combinations. The results reflect an interaction between the intensity and vowel on the one hand, and an interaction between intensity, amplitude of H1/H2, and F1 on the other hand.

INTRODUCTION

Over the last few years several studies on the difference between breathy voiced or murmur phonation and normal voicing were carried out with languages such as Gujarati [1], !Xõo [1, 3], Hmong [2], Hindi [4]. All studies confirmed that the amplitude of the first harmonic (H1) is always higher in murmur (henceforth breathy) phonation compared to the second harmonic, resulting from the sinusoidal glottal source waveform. The perceptual importance of this acoustic parameter was shown by Bickley [1]. On the other hand, breathy voiced stops in Hindi are always accompanied by a drop in overall intensity after the release of the stop. We have already examined the perceptual importance of this acoustic cue using the method of speech editing for stimulus generation and a naturally produced syllable /ḍho/ as point of departure. The results were in agreement with the assumptions of categorical perception [4]. As these results were obtained with only one CV combination, it seemed interesting to test the perceptual load of the acoustic cue with other vowels, too, and to compare these results with acoustic data.

MATERIAL AND INFORMANTS

For the acoustic analysis a list of words was prepared, which contained the breathy stops /bh dh ḍh gh/ in word-initial position. Each stop was followed by the phonemically long vowels /a e i o u/ and occurred in 10 different words, which consisted of either one, two, or three syllables. The material was not controlled for the consonant following the initial CV syllable. Twenty lists were prepared, each containing a subset of the words in randomized order. The lists were read by three informants (1 female, 2 males, aged 23 to 40), all native speakers of Hindi, originating from New Delhi or Uttar Pradesh. The recordings were made in New Delhi in the language lab of the Centre of German Studies, School of Languages of the Jawaharlal Nehru University using a Uher Report and a Senheiser MD421N microphone. The distance to the microphone was set at about 50 cm. The same word list was recorded from another informant (35 years, female) in Munich in the soundproofed room of the Institute using a Telefunken M15 tape recorder and a Neumann U87 microphone. This recording served as font for the manipulation and generation of the stimuli employed in the perception tests.

PROCEDURE

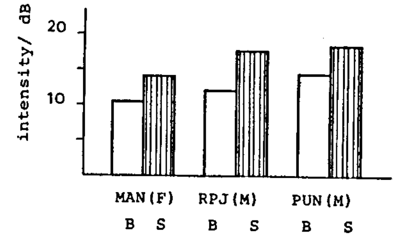
The material was digitized on a PDP11/50 with a sample rate of 20 kHz, filtered with a cut off frequency of 8 kHz, and stored for further analysis. The material was manually segmented with the help of a segmentation routine (for further information cf. [5]). Four different parts in the initial CV syllables were defined: voicing lead, burst + vóiceless aspiration, breathy part of the vowel, and steady part of the vowel. All periodic portions were segmented into single pitch periods [5]. For the acoustic analysis the intensity was calculated for each pitch period and averaged over all periods of the breathy and steady vowel portion, respectively, for each speaker. Separate analyses of variance were applied to all comparisons of means for all speakers. The normality of distribution was checked by a chi-square procedure, while homogeneity of variance was controlled for by applying the chi-square statistics for independent measurements. The level of significance was set to $p < .05$. Multiple comparisons of means were calculated by the use of an a priori F statistic.

For the perception tests 4 CV combinations /dha dho dhu dhi / were selected for manipulation in order to test the interaction between intensity and the vowel. A set of programs was used to generate the stimuli. The procedure has been described in detail elsewhere [4, 5]. The first CV syllable was separated from the rest of the words and the breathy portion of the vowel was eliminated totally. The resulting syllable consisted of voicing lead and burst (which remained unchanged), and the steady portion of the vowel. The fundamental frequency of the vowel was adjusted to 210 Hz for all CV combinations, with a rise over the first five pitch periods and a fall over the last five periods. The first stimulus of each continuum was generated by superimposing a quasi-linear intensity curve on the first 21 pitch periods, the first period being adjusted to 25 dB, the 21st to 55 dB. The intensity was kept constant for the rest of the vowel, with a decrease over 5 periods at the end of the contour. The other 6 stimuli of each continuum were derived from the first stimulus by increasing the intensity onset in the 1st period by 5 dB. For each CV condition identification and discrimination tapes were prepared. In the identification test each stimulus occurred 5 times in randomized order with a pause of 3.5 sec after each stimulus and a pause of 10 sec after a block of 10 stimuli. For discrimination the AX paradigm was used with the step size = 2. Both presentation orders AB and BA as well as AA occurred. The interstimulus interval was 500 ms, pairs were separated by 3.5 sec, blocks of 10 pairs by 10 secs. Each pair occurred 3 times in randomized order. Answer sheets were prepared to allow responses for either breathy or voiced stops in a forced-choice paradigm in the identification task, whereas in the discrimination task subjects had to decide whether the stimuli within a pair sound the "same" or "different". All perception tests were run in the language lab of the Centre of German Studies in New Delhi using a Telefunken language trainer with head phones. The tests were run at a comfortable listening level. About 15 subjects participated in the tests. All were staff or students of the School of languages and were paid for their participation.

RESULTS

Acoustic analysis. Fig. 1 displays the acoustic results for the three speakers averaged over all stops and vowels. Fig. 2 to 4 display the results for the vowel, Figs. 5 to 7 for the stop conditions separately for the three speakers. It is obvious that the intensity of the breathy vowel portion differs significantly from that of the steady portion in all speakers. On the other hand, the amount of difference between both portions is not the same in all subjects: it is large for RPJ (11.1 dB), and smaller for PUN (7.9 dB) and MAN (7.22 dB). The influence of the vowel on the intensity is large for all informants: MAN $F(4,177 = 18.821$; $p < .001$; RPJ : $F(4,185) = 7.65$; $p < .001$; PUN: $F(4,184) = 14.711$; $p < .001$).

Fig. 1: Intensity of the breathy and steady vowel portions in dB averaged over all stops and vowels; plotted separately for the speakers MAN(F), RPJ(M), and PUN(M)



But the vowels do not contribute in the same way to the intensity difference between the breathy and steady portion as can be seen from the following diagrams, which show the significance between the single vowels.

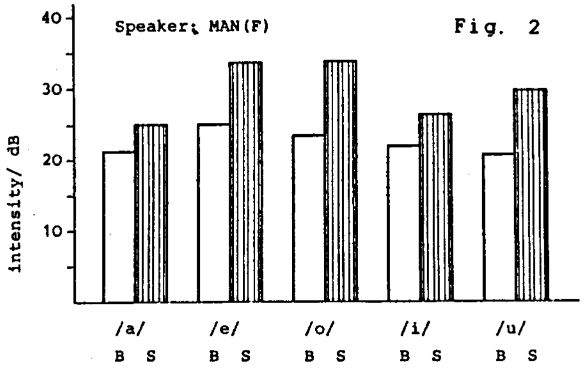
```
MAN(F)   a   i   e   u   o
         x...x   x   x...x
RPJ(M)   u   a   i   o   e
         x.......x   x...x
PUN(M)   i   u   e   a   o
         x...x       x   x
             x...x
```

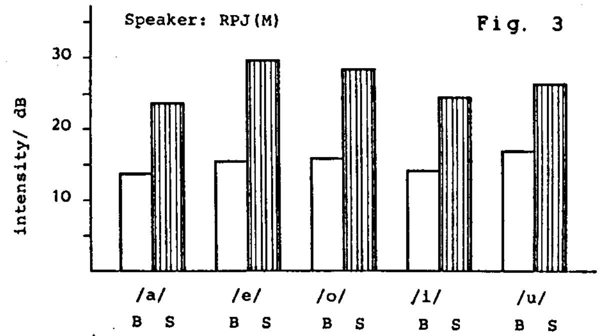
(The diagrams should be read as: vowels underlined by a common dotted line do not differ significantly, whereas vowels not underlined by a common line do.)
MAN shows more influence of the tongue position on the intensity difference as it is largest for the back vowels /o u/, whereas the influence of the tongue height plays an important role in RPJs productions: mid vowels have the larger differences. The results from PUN are not clear, as /o/ and /a/ produce the largest, high vowels the smallest intensity difference. If summarized over all informants the following rank order appears:

i	u	a	e	o
6	7	7	11	14

In other words: the intensity difference between the breathy and steady portion (of the vowel) is a function of the tongue height of the vowel: /i u/ < /a/ < /e o/. The influence of the stop's place of articulation is less compared with the vowel, as only MAN shows a significant influence: MAN: $F(3,195) = 10.322$; $p < .001$; RPJ: $F(3,208) = 2,1$; $p > .05$; PUN: $F(3,184) = .854$; $p > .05$.

Perception tests. The results from the identification task are plotted in Fig. 8. The number of participants is given in the figure. It is obvious that subjects did divide the continuum into two parts only in the /dho/ condition, which resembles best





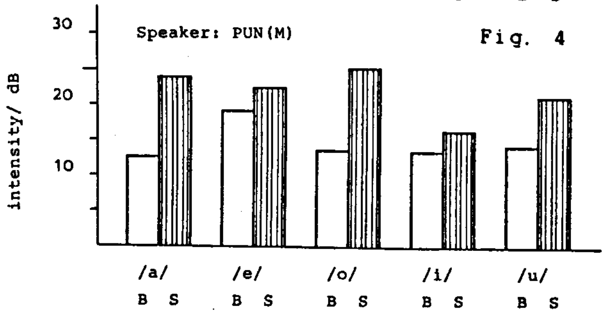
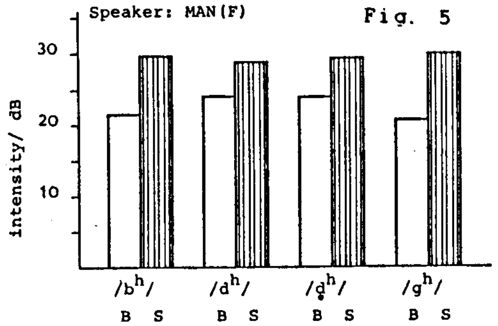
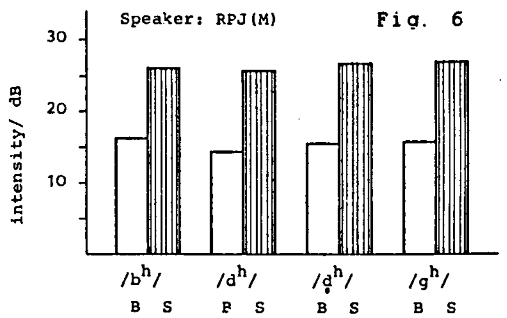


Fig. 2 to 4: Intensity of the breathy and steady vowel portions as a function of the vowel plotted separately for the three speakers





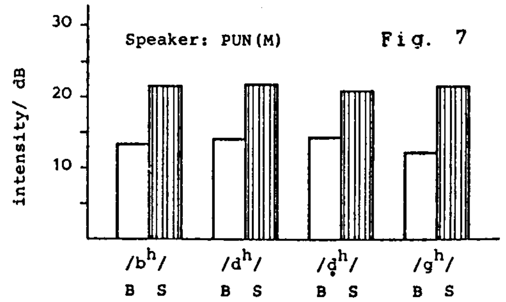


Fig. 5 to 7: Intensity of the breathy and steady vowel portion as a function of the stop plotted separately for the three speakers

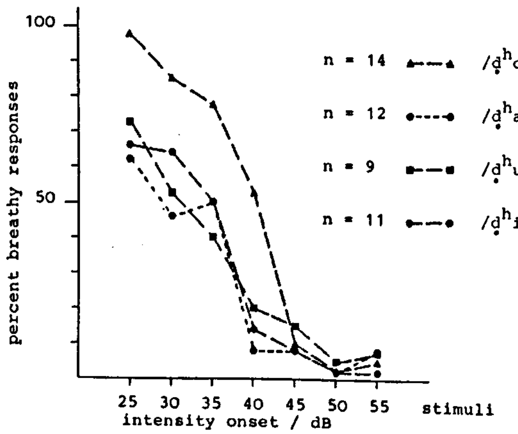


FIG. 8: Percentage breathy responses displayed for the four vowel conditions /ḍha ḍho ḍhi ḍhu/

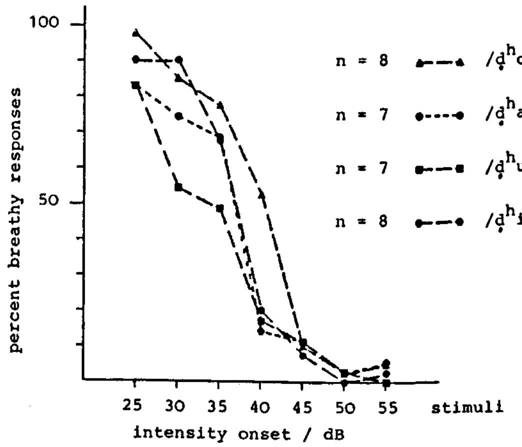


Fig. 9: Percentage breathy responses displayed for the four vowel conditions /ḍha ḍho ḍhi ḍhu/

the assumptions of categorical perception. In the other tests only the first stimulus was assigned to the breathy category. We asked if these results may be due to an interaction between intensity and the vowel, or if they reflect a difference in subjects ability to make use of that special acoustic cue. Therefore we reanalyzed the results and included only those subjects in the analysis who unambigously assigned two categories to the continuum. These results are given in Fig. 9. This time, the identification function improoved for all conditions. The boundary between breathy and voiced occurs latest in the /ḍho/ condition (cf. Table 1), earlier in /ḍha/ and /ḍhi/, earliest in /ḍhu/ where the boundary is less steep. The main effect of continua is significant ($F(3,26) = 3.644$; $p < .05$), but on the other hand, the continua do not differ significantly from each other as shown in the diagram:

```
ḍho   ḍhi    ḍha   ḍhu
x.......x    x.......x
        x...x
```

TABLE 1: Points of intersection between the identification function and the 50% line.

/ḍho/	/ḍhi/	/ḍha/	/ḍhu/
4.06	3.34	3.26	2.68

The results from the discrimination tasks correspond well with those predicted from the identification task using the Haskins formula. In all tests the discrimination peak of the obtained discrimination function correspond in location and height with the calculated one. On the other hand, subjects could discriminate slightly better than predicted, but the difference was not significant.

DISCUSSION

The results from the acoustic analysis confirm that vowels contribute in different degrees to the intensity difference between the breathy and steady vowel portions after breathy voiced stops. This difference is largest for the mid vowel /o/ in all speakers examined. On the other hand, the perception tests gave best results in the /ḍho/ condition which showed the steepest boundary and the latest intersection between the identification function and the 50% line. This means, that the back vowel /o/ needs less intensity difference in perception than do the high and low vowels. At first glance, one would like to explain these results by a close link between the production and perception: the acoustic cue "intensity" is most powerful when it is applied to mid vowels, less powerful with other vowels. In order to make sure that no other acoustic cue was involved, we reexamined the acoustic structure of the syllables used for manipulation with regard to the relationship between the amplitude of the first and second harmonic in the so called steady portion of the vowel, where no breathiness could be detected auditorily. The results are as follows:

/ḍho/ : H1 < H2
/ḍhi/ : H1 > H2
/ḍha/ : H1 < H2
/ḍhu/ : H1 > H2

This means, that the amplitude of H1 is higher in the high vowels /i u/, whereas H2 exceeds the value of H1 in either /o/ or /a/. These relationsships are undoubtedly due to the formant structure of the vowels, where F1 interacts with H1 in the high vowels, and with H2 in /o/. No interaction between F1 and H1/H2 occurs in /a/. We believe that our results reflect an interaction between the overall intensity, the amplitude of H1 and H2 as well as F1. As these results were rather unexpected, further investigations are needed to explain the extent of that interaction.

REFERENCES

[1] Bickley, C.: Acoustic analysis and perception of breathy vowels. Working Papers Vol. 1, MIT Speech Communication, pp. 73-83 (1980)

[2] Huffman, M.K.: Measures of phonation types in Hmong. University of California Working Papers in Phonetics 51:1-25 (1985)

[3] Ladefoged, P.: The linguistic use of different phonation types. University of California Working Papers in Phonetics 54: 28-39 (1982)

[4] Schiefer, L. - Kotten, K.: Amplitude envelope and the perception of breathy stops in Hindi. Proc. 10th Int. Congr. Phon. Sci., Utrecht, pp. 459-463, (Cinnaminson, Dordrecht, 1984)

[5] Schiefer, L.: Fo in the production and perception of breathy stops: evidence from Hindi. Phonetica 43: 43-69 (1986)

Acknowledgment

This research was supported by German Research Council (DFG) grant Schi 181/2-1. I am grateful to Prof. R.P. Jain and Prof. P. Talgeri for their help in selecting the participants for the perception tests and the permission to use the Language Lab, as well as Swaran Thakur-Weifenbach, Manorama, R.P. Jain, and Puneet who served as informants.