Physiological study of whispered speech in Moroccan Arabic


1. Université Sidi Mohamed Ben Abdellah, Faculté Polydisciplinaire de Taza, Taza-Morocco.
2. Laboratoire de Phonétique et Phonologie, Sorbonne-Nouvelle/CNRS-UMR7018, Paris-France.
3. Department of Linguistics, University of Victoria, Victoria-Canada.
Chakirzeroual@yahoo.fr; esling@uvic.ca; lise-buchman@numericable.fr

Abstract

The analysis of pictures extracted from endoscopic video images filmed during isolated Moroccan Arabic /i1C1i2C2/ items (where C1 = fricatives + plosives), showed that:

- All the segments of a whispered item are produced with an anterior-posterior epilaryngeal compression, an addition of the ventricular bands and of the membranous part of the glottis, and an opening at the level of its cartilaginous part.
- Glottal aperture differences observed in normal speech between voiceless consonants tend, in the whispered speech, to be maintained between the fricatives, and to be neutralised between the plosives.
- We did not observe regular laryngeal articulatory differences between all the whispered consonant pairs corresponding, in normal speech, to the linguistic "voicing" contrast.

1. Introduction

Segments of a whispered sequence are characterized by the lack of voicing and the presence of friction noise except during the closure phase of plosives. Several perceptive studies show, however, that the segmental contrasts observed in normal speech, such as "voicing" [11, 14], tend to be maintained in whispered speech. Even suprasegmental patterns such as accent [13] and tonal contrasts [1] can also persist in whispered speech.

"Whisper", with "voice" and "breath", are simple phonation modes, i.e. sound sources produced in the larynx and modulated by supralaryngeal articulatory adjustments. Canonical "voice" is a periodic laryngeal sound which results from regular vocal fold vibrations. "Whisper" and "breath" are aperiodic sounds generated by turbulent airflow. "Whisper" is more productive than "breath"; this latter is observed nearly exclusively during /h/ or during the (oral) release of an aspirated plosive (the "aspiration" phase). According to Catford and mainly to Laver [2, 9], the fact that the airflow during "whisper" is more strident than during "breath" constitutes the most prominent difference between these two modes. We will show below that the intensity is not the only phonetic parameter that distinguishes between these two phonation modes.

Several studies have shown that the vocal folds are fully adducted during "voice" and fully open during "breath" [2, 9]. On "whisper", we have relatively fewer studies whose descriptions are not always similar. In spite of this, it appears that "whisper" is produced when the cartilaginous part of the glottis is open and its membranous part moderately or much adducted [7]. According to Catford and especially to Laver, the main articulatory difference between "whisper" and "breath" is attributed to the size of the glottis, which is narrower during the former than during the latter mode. For Laver [9: 191], "the sounds that are normally voiceless... remain voiceless, and it is only the sounds that would normally be voiced, such as the pronunciations of the vowels, that become whispered." Our Moroccan Arabic (MA) data will show, however, articulatory differences between whispered segments and their non-whispered cognates, even though the latter are voiceless in normal speech.

"Whisper" has another articulatory characteristic that is not always mentioned in the literature (Laver and Sonesson [9, 12]), or very marginally (Catford [2]). Our MA data agrees, in fact, with Esling's [5] voice model, that goes farther than Hirose's [7] characterization of whisper. Esling predicts that the whispered segments are always produced with the involvement of the supraglottic level, and more precisely with posterior-anterior aryepiglottic constriction (Fig. 1, and [5]).

Although we consider this constricted articulatory property as intrinsic to whispered segments, we do not imply that all the laryngeal adjustments stay the same through a whispered sequence. Indeed, our data will also show laryngeal articulatory differences even between whispered segments. These results partially agree with other previous EMG [14], endoscopic [15] or transillumination [10] studies.

2. Method and material

A naso-endoscope (Olympus ENF-P3) was inserted through the nostril of one MA native speaker (the first author, 36 years), and placed immediately over the larynx. Its external tip was connected with a micro camera (Olympus CJH 250) whose frame rate was 25 frames/second.

The items are MA words, and some non-words, having the /i1C1i2C2/ shape and where C1 = /t d T D s z/ and where C2 = /t d T D s z x f l/. /T D/ are "emphatic" coronals, or uvularized correspondents of /t d/.

To limit the variability of the naso-endoscope position, the nine items were placed in only one list that was pronounced several times. Within each list, each item was repeated twice with normal speech followed by two others with whispered speech, without being put in a carrier phrase (in an isolated manner). The segmentation was made while combining spectrographic and auditory indications.

The video sequences were digitised with Adobe Premiere 6.0 and the acoustic (limited here to VOT) and statistical measures with, respectively, TF32 and StatView. The pictures were analysed with the help of Adobe Photoshop and Acrobat 6.0, which allowed us to make three articulatory measures for each of the nine consonants (Fig. 1, Table 1).
3. Results and discussion

The segments have been analysed in /i1C1i2C2/ items. Those pronounced with normal speech are represented, in general, by their standard phonetic symbols (example: /t/, /i1si2b/). /i1/ and /i1si2b/ designate their whispered correspondents. In /i1si2b/, the four segments are whispered. Our analyses will be done by taking into consideration the results of two previous aerodynamic studies [16, 17]. These latter were made on the MA consonants produced with normal speech by the same speaker who participated in this present study.

Three separated one-factor ANOVA tests show that, in whispered speech Da [F(8, 27) = 12.81, p<0.001], Wg [F(8, 27) = 12.77, p<0.001], and S [F(8, 27) = 32.85, p<0.001] vary significantly with the nature of the consonant. The same parallel significant effects are also observed in normal speech: Da [F(8, 27) = 17.58, p<0.001], Wg [F(8, 27) = 76.55, p<0.001], and S [F(8, 27) = 117.57, p<0.001]. Another paired t-test shows that Da of the nine consonants is statistically larger in normal speech than in whispered speech [df = 35, t = 21.64, p<0.001].

3.1. /i/ vs /i/w

During /i/ (Fig. 2), the glottis is fully adducted, the ventricular bands very separated, the epilaryngeal space very open, and only the vocal folds vibrate. The epilaryngeal space of /i/w (Fig. 2) is very compressed: the base of the epiglottis, the arytenoids, and aryepiglottic folds are brought closer. The supralaryngeal cavity as during /i/ (Fig. 2). However, during the closure of /i/, the glottis opens progressively and reaches a maximal glottic opening (Wg) immediately before or at the beginning of its oral release, and then it closes progressively while going toward /i/. At the beginning of /T/, the glottis opens very slightly and keeps this state throughout almost the whole length of the closure. The resolution of the camera does not permit determining exactly, during /T/, the onset of its very fast vocal fold adduction regarding its oral release. It is not excluded that the glottis was already closed at the onset of the oral release of /T/. The arytenoids are drawn apart during /i/ and remain in contact during /d/ and even /T/ (Fig. 2). Note that Wg (p<0.001) and S (p<0.001) are larger during /i/ than during /T/. The noise release (NR) is substantially longer during /i/ (80msec) than during /T/ (17msec), /d/ (17msec) and /D/ (19.25msec) (Fig. 3).

The complete adduction of the vocal folds, at the moment or immediately after the release of /d D T/, prevents the aspiration phase. The burst (transient+frication [8]) of these consonants is produced mainly by the discharge of the air compressed, during the closure, in the oral cavity [4: 227]. However, the larger opening of the glottis and the slower adduction of the vocal folds during the release of /i/ permits a discharge of the air compressed in the supralaryngeal and sublaryngeal cavities [4], thus a longer NR. Spectrographic analyses (Fig. 3) show that the NR of /i/ is dominated by the "frication" phase. /i/ is therefore an affricated and not an aspirated consonant. The observations in the whispered speech confirm this analysis.

Although the vocal folds are adducted during /d D/ and slightly adducted during /T/, the NR of /T/ is very similar and even shorter than the one for /D/. According to Dart [3], when the vocal tract walls are tense, the intraoral pressure (Po) and oral airflow (U) are respectively higher and lower than when they are lax. This tension prevents passive vocal tract expansion and increases the Po. "This decrease in elasticity of the cavity walls also contributes to a lower peak flow by decreasing the amount of elastic recoil of the walls and thereby slowing down the initial flow velocity at release" [3: 143]. Previous analysis [16] showed that /T/ has a higher Po and lower U than for /D/! These observations suggest that /T/ is produced with a more important tension of its supralaryngeal and even laryngeal structures. This hypothesis can also explain why the arytenoids remain adducted during /T/ and are separated during /i/. The laryngeal cavity during /d T D/ remains very similar to that of /i1i2/ (Fig. 2). During /d/, the glottis opens slightly compressed to its posture during /i1i2/. In spite of this, /d T D/ seem to possess Wg and S that are not very different. Indeed, Wg and S magnitude differences between /i1i2/ vs /d1i2/ (p=0.31, p=0.22), /i1i2/ vs /D1i2/ (p=0.12, p=0.34) and /i1i2/ vs /T1i2/ (p=0.12, p=0.06) are not statistically significant. /i1i2/ continues to have a very long burst which is clearly distinct from /i1i2/ onset (Fig. 3). The burst of /T/ d D/ appears short (Fig. 3), even though the glottis remains open during their release (Fig. 2). The very significant length of the /i/ and /i1i2/ "frication" phase is mainly attributed to the slower rate of their oral release. Indeed, a more distributed anterior part of the tongue seems to be maintained much longer, actively and/or passively, very close to the dento-alveolar region. This hypothesis is strengthened by previous observations that showed that /i/ is laminal whereas /d D T/ are apical [18].

Note that /D/ and /T/ are separated during /T/. During /T/ d D/ the glottis opens progressively and reaches a maximal glottic opening (Wg) immediately before or at the beginning of its oral release, and then it closes progressively while going toward /i/. At the beginning of /T/, the glottis opens very slightly and keeps this state throughout almost the whole length of the closure. The resolution of the camera does not permit determining exactly, during /T/, the onset of its very fast vocal fold adduction regarding its oral release. It is not excluded that the glottis was already closed at the onset of the oral release of /T/. The arytenoids are drawn apart during /i/ and remain in contact during /d/ and even /T/ (Fig. 2). Note that Wg (p<0.001) and S (p<0.001) are larger during /i/ than during /T/. The noise release (NR) is substantially longer during /i/ (80msec) than during /T/ (17msec), /d/ (17msec) and /D/ (19.25msec) (Fig. 3).

The complete adduction of the vocal folds, at the moment or immediately after the release of /d D T/, prevents the aspiration phase. The burst (transient+frication [8]) of these consonants is produced mainly by the discharge of the air compressed, during the closure, in the oral cavity [4: 227]. However, the larger opening of the glottis and the slower adduction of the vocal folds during the release of /i/ permits a discharge of the air compressed in the supralaryngeal and sublaryngeal cavities [4], thus a longer NR. Spectrographic analyses (Fig. 3) show that the NR of /i/ is dominated by the "frication" phase. /i/ is therefore an affricated and not an aspirated consonant. The observations in the whispered speech confirm this analysis.

Although the vocal folds are adducted during /d D/ and slightly adducted during /T/, the NR of /T/ is very similar and even shorter than the one for /D/. According to Dart [3], when the vocal tract walls are tense, the intraoral pressure (Po) and oral airflow (U) are respectively higher and lower than when they are lax. This tension prevents passive vocal tract expansion and increases the Po. "This decrease in elasticity of the cavity walls also contributes to a lower peak flow by decreasing the amount of elastic recoil of the walls and thereby slowing down the initial flow velocity at release" [3: 143]. Previous analysis [16] showed that /T/ has a higher Po and lower U than for /D/! These observations suggest that /T/ is produced with a more important tension of its supralaryngeal and even laryngeal structures. This hypothesis can also explain why the arytenoids remain adducted during /T/ and are separated during /i/.

The laryngeal cavity during /d T D/ remains very similar to that of /i1i2/ (Fig. 2). During /d/, the glottis opens slightly compressed to its posture during /i1i2/. In spite of this, /d T D/ seem to possess Wg and S that are not very different. Indeed, Wg and S magnitude differences between /i1i2/ vs /d1i2/ (p=0.31, p=0.22), /i1i2/ vs /D1i2/ (p=0.12, p=0.34) and /i1i2/ vs /T1i2/ (p=0.12, p=0.06) are not statistically significant. /i1i2/ continues to have a very long burst which is clearly distinct from /i1i2/ onset (Fig. 3). The burst of /T/ d D/ appears short (Fig. 3), even though the glottis remains open during their release (Fig. 2). The very significant length of the /i/ and /i1i2/ "frication" phase is mainly attributed to the slower rate of their oral release. Indeed, a more distributed anterior part of the tongue seems to be maintained much longer, actively and/or passively, very close to the dento-alveolar region. This hypothesis is strengthened by previous observations that showed that /i/ is laminal whereas /d D T/ are apical [18].

Figure 1: (1) Da (cm): distance between the arytenoids and the epiglottis. (2) Wg (cm): width of maximal glottal opening. (3) S (cm²): Section of the glottis. These measures correspond to relative and not absolute values. (a): arytenoid, (b): vocal fold, (c): ventricular band, (d): aryepiglottic fold, (e): base of the epiglottis, (a+d+e): epilaryngeal space or aryepiglottic sphincter.
3.3. /s z /b/ vs /s z w/

During /s/, the glottis opens and reaches a maximal width and section that are significantly larger in /z/ than in /s/ (Wg and S: p<0.001). However, during /z/ w/, the glottis remains as adducted as during /i 1/ z w/. Some previous aerodynamic analyses have shown that, in MA, Umax during /z/ is substantially higher than during /s/, even though their Po are not significantly different [17]. This regularity shows that the supralaryngeal opening during /z/ is larger than during /s/ (based on the Orifice Equation). The glottis opens widely during /s/ to permit generation of a transglottic airflow which is sufficiently high to achieve a more elevated value of Po. This latter condition is necessary to initiate, during /s/, a turbulence through its large supralaryngeal opening. Another hypothesis also may explain these /z/ vs /s/ laryngeal differences. Indeed, it is widely accepted that the spectral properties of a fricative depend mainly on the place of its narrowest constriction in the vocal tract [Phil Hoole: personal communication]. If the supralaryngeal opening of /z/ is larger than during /s/, the glottis remains similar or opens slightly compared to /i 1/ w. Wg and S of /z/ w are significantly larger than those of /s/ w and /s z w (p<0.001).

These two hypotheses also can explain why Wg and S are larger during /z/ w than during /s/ w (p<0.001). The glottal adjustment seems to be similar during /s/ w and /z/ w: Wg (p=0.33); S (p=0.47), but different while comparing /z/ w and /s/ w: Wg (p<0.001); S (p<0.001). On the other hand, the epilaryngeal constrictor is significantly more compressed during /s z w than during /z/ w (p<0.01).

/s/ w develops a laryngeal cavity shape that is very similar to that of /z/ w (p=0.33); its Wg and S are not significantly different from those of /s/ w (p=0.19, p=0.11).

4. Conclusion

The configurations of the laryngeal cavity (glottic and supraglottic levels) for all the segments within a sequence, pronounced with normal speech, become different when this same sequence is whispered. Therefore, all the segments of a whispered sequence must be considered as whispered, and not only those that are "voiced" in normal speech, as implied by Laver [9]. Whispered segments always have an anterior-posterior epilaryngeal compression (between the base of the epiglottis, the aryepiglottic folds and the arytenoids), the glottis remains open mainly at the level of its cartilaginous part, and the ventricular bands advance toward the median line.

Although /t T d D/ w-stops seem to have very similar laryngeal adjustments, they tend to maintain the same differences regarding the length of their noise release as in normal speech. We deduce that these differences are controlled, in whispered speech, not by the coordination between laryngeal and supralaryngeal gestures as in normal speech, but mainly by supralaryngeal adjustments.

/s z b/ w-stops show some laryngeal adjustment differences which are mainly due to the particular behaviour of /z/ w. Indeed, the maximal opening of the glottis as well as its whole section are significantly larger during /z/ w than during all the rest of the consonants. We demonstrated that this regularity, also observed in normal speech, is bound to the aerodynamic properties of /z/ w.

We did not observe, in whispered speech, clear laryngeal articulatory differences between all the Moroccan Arabic consonant pairs related, in normal speech, by the linguistic "voicing" contrast. However, other acoustic, aerodynamic and perceptual investigations are necessary to determine if this contrast is maintained or not in whispered speech.

Table 1: Mean values (and standard deviation) of:

| Da (cm): distance between the arytenoids and the epiglottis, Wg (cm): width of maximal glottal opening, and S (cm²): section of the glottis. Each measure is the average of 4 repetitions (see Fig. 1). |
|---|---|---|---|---|---|
| **Normal speech** | **Whispered speech** |
| Da | Wg | S | Da | Wg | S |
| /s/ | 2.11 | 0.61 | 1.01 | 0.91 | 0.77 | 0.95 |
| /z/ | 1.70 | 0.00 | 0.00 | 0.88 | 0.84 | 1.05 |
| /i/ | 2.53 | 0.99 | 2.36 | 1.51 | 1.27 | 2.56 |
| /s/ | 1.75 | 0.00 | 0.00 | 1.45 | 0.75 | 1.15 |
| /t/ | 2.06 | 0.56 | 0.77 | 0.82 | 0.71 | 0.85 |
| /d/ | 1.93 | 0.00 | 0.00 | 0.66 | 0.63 | 0.68 |
| /t/ | 1.60 | 0.27 | 0.18 | 0.89 | 0.82 | 1.12 |
| /d/ | 1.89 | 0.00 | 0.00 | 0.87 | 0.71 | 0.99 |
| /i/ | 1.96 | 0.13 | 0.05 | 0.89 | 0.86 | 1.18 |
| | 0.04 | 0.12 | 0.05 | 0.06 | 0.03 | 0.14 |
Figure 2: State of the glottic and supraglottic levels during Moroccan Arabic /i1/ and /C1/ = /s z/ pronounced in /i1C1i2C2/ items with normal speech (top row), and whispered speech (bottom row).

Figure 3: Spectrograms of /-ti/, -Tib, -dir, -Dim/ sequences extracted from /i1ti2/, i1Ti2b, i1di2r, i1Di2m/ items pronounced with normal speech (top row) and whispered speech (lower row) by a Moroccan Arabic speaker.

5. References