

Interrelationship between vocal effort and vocal tract acoustics: a pilot study

Maëva Garnier¹, Joe Wolfe¹, Nathalie Henrich², John Smith¹

¹ Music Acoustics Group, School of Physics, University of New South Wales, Sydney, Australia

² Département Language and Cognition, GIPSA-Lab, Grenoble, France

mgarnier@unsw.edu.au

Abstract

How and why do vocal tract resonances and articulation change when shouting? Vocal tract resonances R1 and R2, Open quotient, Fundamental frequency, voice intensity, larynx height, lip aperture and spreading were simultaneously recorded for a female native speaker of French, on 7 French vowels and for 3 different conditions of increasing vocal effort: “normally” (S1), when keeping a pitch constant (S2) and when keeping both pitch and articulatory position constant (S3). R1 depends not only on articulation, but also on open quotient and larynx height. Results suggest that raising R1 and hyper-articulating in loud speech may serve to tune R1 to the first or the second voice harmonic, and thus help in increasing voice intensity.

Index Terms: vocal effort, resonances, articulation.

1. Introduction

Loud speech involves glottal, acoustic and articulatory changes in vowel production [6, 8, 11, 17, 18]. In natural speech, an increase of vocal intensity is often accompanied with an increase of fundamental frequency [16]. Several other glottal parameters may vary with increased vocal intensity [11]: speed of closure increases, duration of glottal open time over fundamental period (open quotient, Oq) decreases, and asymmetry of glottal pulses increases. Articulatory movements are amplified, especially mouth aperture, associated with an increase of the first formant frequency (F1). Variation of the second formant frequency (F2) is less significant and depends on languages and vowel categories [17, 18].

It is not clear whether all these speech modifications are related to the increase of vocal intensity, whether they may be produced in order to increase speech intelligibility, or whether they may be considered as compensating for effects of the increase in vocal intensity [8].

Loud speech implies an increase in vocal effort. In the context of linear source-filter theory [5] which prevails in the speech community, vocal effort is mostly considered at the glottal level and described by the variations of glottal flow parameters [11]. Within this framework, formant modifications which accompanies an increase of vocal effort are commonly related to articulatory aspects [13]. Thus, the increase in F1 is commonly assumed to be primarily a consequence of mouth aperture, as both are strongly related in natural speech [14] and always vary together with vocal effort. However, this is a simplification. There is an acoustical interaction between source and filter which is neglected as a first approximation in the source-filter theory [5]. In particular, the first formant is produced by the first acoustic resonance of the vocal tract (R1), whose frequency does not depend only on the vocal-tract articulatory shape, but also on the boundary conditions at the glottis [2], [19]. A recent study has demonstrated that the first two resonance frequencies of a simple tube excited at one end and open at the other end would rise with increased glottal width and

increased glottal open quotient [2]. Thus, if increasing vocal effort decreased Oq with no change in opening amplitude, then this effect, on its own, would lower R1 instead of increasing it, as observed in shouted speech.

In this article, we explore the different contributions of Oq, F0 and articulation to the variation of R1 with vocal effort, in order to better understand the causes and effects of formant and articulation changes in loud speech.

2. Material and method

2.1. Experimental protocol

The corpus consisted of 7 French vowels ([a], [ɛ], [e], [y], [u], [o] and [œ]). Each vowel was first sustained for 4 seconds in normal phonation, then for 4 more seconds in loud phonation, in the same breath. 3 sessions were recorded: In the first (S1), the subject was only told to produce the vowel at normal, then loud level. In the second session (S2), the speaker was told to keep a constant pitch when increasing vocal effort. In the third session (S3), the subject was told to keep both pitch and articulation constant. Each vowel was produced 5 times in every session, at both normal and loud level.

2.2. Measurements

In this pilot study, we simultaneously recorded electroglottographic (EGG) and articulatory signals from a female native speaker of French, as well as acoustic resonances of her vocal tract.

The method used to measure vocal tract resonances involves injecting a carefully synthesised broad-band acoustic current at the mouth, while the subject is phonating, and recording the vocal tract response to that excitation with a microphone also at the mouth [3]. In this experiment, we focus on the first two resonances. Therefore we have excited the vocal tract with a broadband signal from 200 to 3000 Hz. The vocal tract response to that excitation was recorded with a pressure field microphone (B&K 1/4" 4944A) at the speaker's lips, with a Nexus conditioning amplifier, and then digitised at 44.1 kHz with a MOTU 828 audio interface. The resonance frequencies may be measured with a resolution approaching the frequency spacing of the excitation (11 Hz = 44.1 kHz / 4096). Unlike formant estimation using LPC, it is not limited by the fundamental frequency. This is an advantage for high pitched voices, such as women's shouted speech, for which we can considerably improve the accuracy of resonance estimation. The cost of high frequency resolution is inevitably poor resolution in time (1/11 Hz = 91 ms). Further, several measurements must be averaged to obtain some clear maxima of the impedance curve, which constrains us to study sustained vowels instead of continuous speech. We have here averaged the 30 last measurements of each vowel (i.e. about the 3 last seconds), and manually detected the maxima from this average impedance curve.

EGG signal and larynx vertical movements have been recorded with an electroglottograph EG2 (Glottal Enterprise). Both signals were digitised as for the microphone, except that the laryngeal movements signal was amplitude modulated at

10 kHz and then demodulated for analysis. F0 was measured from the derivative of the EGG signal, by detecting glottal closing times [9]. Glottal opening times were directly detected from the EGG signal, using a threshold method at 3:7 [12]. Oq was then measured as the ratio of the glottal opened phase duration over the fundamental period. F0 and Oq mean values were measured on the same 3s time period of the vowel as for impedance measurements. Vertical displacement of the larynx was calibrated prior to the recordings, using the same method as Pabst et al. [15], to convert signal changes into millimeters.

Articulatory data were extracted from video recordings (25 images/s) of the speaker's lips, using a semi-automatic detection method of lip contour. The subject did not move, since she had to keep the microphone and the noise source on her lower lip. The pixels to millimeters conversion was calibrated prior to the recordings. Mean lip spreading and aperture were measured over the same 3s time period of the vowel as for acoustic and glottal measurements.

3. Results

3.1. Global observations

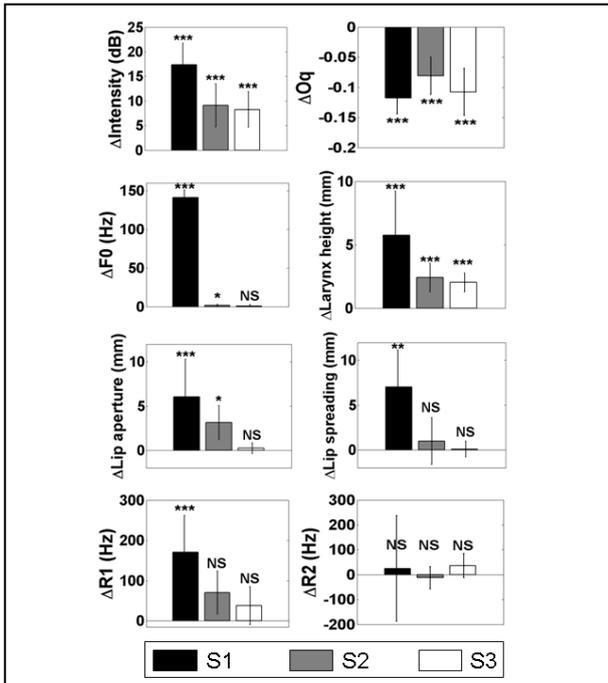


Figure 1: Change in glottal, acoustic and articulatory parameters between normal and loud phonation. Three different conditions were studied: without instruction (S1), maintaining constant pitch (S2), maintaining constant pitch and articulatory position (S3). Histograms and error bars represent the mean value and the standard deviation of the difference between normal and loud phonation over all the vowels¹.

The first six graphs of Figure 1 show how the speaker complied with the experimental instructions, i.e. increased vocal effort for all sessions, constant F0 for S2, and constant F0 and lip aperture for S3:

First, vocal intensity significantly increases in the three sessions from normal to loud phonation, more in S1 (+17 dB) than in S2 and S3 (+9 dB and +8 dB). This corresponds to a

significant decrease in Oq for every vowel of S1 (-0.12) and S3 (-0.11), and a little less in S2 (-0.08 for all vowels except [u] and +0.04 for [u]). Averaged over all three sessions, Oq is observed at about 0.66 in normal phonation and 0.56 in loud phonation.

Second, the rise of F0 is large in S1 (+141 Hz) and negligible in S2 and S3 (+2 Hz and +1 Hz), as instructed.

Last, the variation of articulatory parameters depends on the vowel category (see paragraph 3.2). Nevertheless, a global and significant amplification of lip aperture movements is observed in S1 and S2 for almost all the vowels (except [e] in S2), while it does not change for any vowels in S3, as instructed. Lip spreading also significantly increases in S1 (except for [a]), as well as for some vowels in S2, but again not in S3. The larynx rises significantly in every session, although this change is rather greater in S1 (+6 mm) than in the two last sessions (+2 mm). Thus, lip articulation has been kept constant in S3 but the subject has not been able to avoid laryngeal movements.

The two last graphs of Figure 1 show how these experimental conditions have affected the variation of R1 and R2 with vocal effort. As expected, we observe a significant increase of R1 (+171 Hz) with increasing vocal effort in the “natural” speech condition (S1). This increase is smaller and not significant in both S2 and S3 (+62 Hz and +38 Hz).

No significant change is observed in the mean of R2, because of changes of different sign for different vowels.

3.2. Observations by vowel category

On Figure 2, we can see how R1 and R2 conjointly vary with vocal effort for different vowels, and for the three sessions.

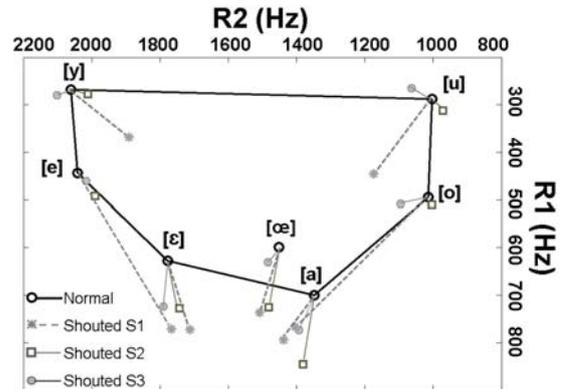


Figure 2: Variation of the first two vocal tract resonances (R1 and R2) from normal to loud phonation, for 3 different conditions: without instruction (S1), maintaining constant pitch (S2), maintaining constant pitch and articulatory position (S3). Each point represents the mean value over 5 productions of the same vowel. Standard deviations (omitted for clarity) are typically from 10 to 30Hz.

First, we observe in “natural” shouted speech (S1) a similar increase in R1 (+120Hz on average) for open ([a], [ɛ], [œ]) or closed vowels ([y], [u]), but much more (+298 Hz) for mid-open vowels ([e], [o]). On Figure 3, we similarly observe that for [o], [e], [œ], [ɛ] and [a], lip aperture significantly increases when shouting (from 4 to 13 mm). The increase is greater for vowels whose aperture is smaller in normal speech. This results in a smaller distinction in R1 and lip aperture among these vowels when they are shouted. For the vowels [u] and [y], however, lip aperture barely increases for [u] and [y] (+1 mm). We also observe that R2 increases significantly for back vowels and decreases significantly for front vowels, which leads to a reduction of the vowel system (i.e reduced distinction among vowels).

In S2, for open vowels ([a], [ɛ], [œ]: +114 Hz), the increase in R1 is significant and comparable with that in S1. For [e]

¹ The following notation has been adopted for indicating statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and NS (not significant) $p > 0.05$.

the change is slight (+40 Hz), and there are no significant changes for [y], [u] and [o] (+17 Hz). Lip aperture also increases more for open vowels (+5 mm) than others (+2 mm). There are no significant R2 variations for any vowel. This results in an expansion of the vowel system in the R1 dimension (see Figure 2).

In S3, R1 still increases significantly for extreme open vowels ([a] and [ɛ]: +97 Hz), and for [œ] (+39 Hz), but less than in S1 or S2. It does not change significantly for the other vowels (+14 Hz) and even tends to decrease for [u] (-11 Hz). This cannot be related to lip articulation, which does not change for any vowel. R2 increases somewhat for every vowel except [e]. This slight variation (+48 Hz on average) is even significant for [o], [u] and [œ]. The vowel system transform then consists in an expansion along the R1 dimension and a slight shift towards higher R2 frequencies.

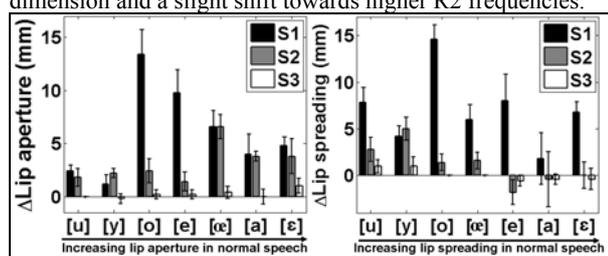


Figure 3: Variation of lip aperture and spreading from normal to loud phonation, for the 3 different conditions. Histograms and errorbars represent the mean value and the standard deviation of this variation over 5 productions of the same vowel.

4. Discussion

4.1. Contributions to the increased R1 in shouted speech

When the speaker is asked to keep constant pitch and mouth position while shouting (S3), a decrease in R1 is observed only for the vowel [u], and that is very small. For all other vowels, R1 either does not change, or increases significantly for [a] and [ɛ], even though Oq decreases. This opposes the global trend for F1 to decrease with Oq, reported by Barney et al. [2] for *in vitro* measurements on a simple tube.

Could this discrepancy come from the rising of the larynx that we observe in shouted speech, even when F0 and mouth articulation are controlled (S3), and which contributes to shorten the vocal tract? Using a simple two-tube resonators model ([5], p.66), we estimate that a 2 mm larynx rise would increase R1 on the order of 0 to 14 Hz, according to the vowel, which cannot totally explain the observed increase in R1 for open and mid-open vowels (in S3), especially if this effect is combined with the expected decrease in R1 with Oq. In addition, we do not observe a very strong correlation between R1 variation and larynx displacement from normal to shouted speech ($R=0.48$).

An alternative explanation is that the speaker may have changed the pharyngeal area or moved the tongue forwards a little when increasing her vocal effort, which could explain the small shift of the vowel system along the R2 dimension (in S3), and which could also have slightly affected R1. But again, it could hardly explain the large variation of R1 observed in [a] and [ɛ]. Another explanation may be found in [2] again. Although these authors report a global trend for F1 to decrease with Oq, they also show that there are some particular cases for which F1 can increase up to +40 Hz with a lowering of Oq, for a constant glottal width, and up to +160 Hz for a simultaneous increase of glottal width with a lowering of Oq. Thus, R1 changes in S3 could be well explained by a combined effect of larynx rising and Oq

lowering, on condition that glottal width would have increased with vocal effort in this case. Such an evolution of laryngeal parameters could occur for a phonation mode called “flow voice” by Gauffin and Sundberg [7], which would be consistent with the perceived voice quality in S1. Yet, the perceived voice quality in S2 and S3 seems closer to “pressed” than to “flow” voice. Such a phonation is related to a decrease of glottal width and open quotient with increased vocal effort [7].

The R1 rise in S2 is less than in S1, where pitch was allowed to rise. However, controlled pitch was not the only factor to change in S2: the rise of the larynx, decrease in open quotient and amplification of articulatory gesture were all also smaller than in S1 and can all explain this weaker increase in R1. The observation that the larynx rises with vocal effort, but especially with F0 (see Figure 1) is consistent with the reports of Pabst and Sundberg [15]. This does not seem to be highly controllable by this speaker. We also notice that vocal intensity increases much more with a simultaneous increase of F0, whereas the decrease in Oq is similar across all sessions (although a little weaker in S2), and that articulatory changes between S2 and S3 do not contribute a lot in the increase of vocal intensity. These different observations, coupled with the more pressed voice quality perceived in S2 and S3, may indicate that the speaker has played on vocal folds tension to increase vocal intensity when she was told to keep a constant pitch (in S2 and S3), while in the “natural” shouting condition (S1), she may have only used subglottal pressure to increase her vocal intensity, resulting in a simultaneous increase in pitch [16]. When looking at voice range profiles, this seems quite logical that the most intense sounds cannot be produced without increasing pitch. Although rising pitch may be part of a global vocal effort gesture rather than a deliberate strategy on its own, it is not only a consequence of the increase in vocal intensity but can also influence it in return [1].

Lastly, R1 increases more when the subject is “allowed” to move her articulators, and especially to open the lips (S3→S2), which confirms the well-known influence of jaw and lip aperture on the first vocal tract resonance [14].

To summarize, these different arguments seem to indicate that R1 mainly depends on lips and jaw articulation. There are no strong evidence of an influence on R1 rise with vocal effort of laryngeal factors which are directly or indirectly related to the increase of vocal intensity (decrease of Oq [11], larynx elevation related to the increase in F0 [15]). These factors may not be controlled individually and intentionally by the speaker but rather participate in a global vocal effort gesture.

4.2. Effect or reason for hyper-articulating and raised R1 in shouted speech

Schulmann [18] suggested that mouth opening increases in shouted speech to prevent the turbulent noise that could hypothetically be created if the speaker increased expiration pressure without widening the vocal tract constriction. Little turbulent noise was noted in S3.

Another hypothesis is that formant and articulatory modifications could serve to increase or preserve speech intelligibility. On the one hand, increasing jaw opening is considered as a speech perturbation in many studies [8] so it may seem paradoxical to propose that it could improve intelligibility. On the other, the variable (F1-F0) has been shown to be more relevant to vowel height perception than F1 itself [4]. In S2, the speaker reduces her articulatory gestures whereas she is only told to keep a constant F0. This supports the hypothesis that articulatory modifications in shouted

speech aim at preserving a constant R1-F0. In “natural” shouted speech (S1), this is indeed the case for [y], [u], [œ] and [ɛ] but not for [a], [o] and [e] (see Figure 4). Variations of R1 and R2 in shouted speech consist not of a simple “shift” of the vowel system, which would preserve the contrast between vowel categories along this R1-F0 dimension, but rather of a speech reorganisation, and even a vowel reduction, inasmuch as [e] and [ɛ] become almost confused, as do [a], [o] and [œ] (see Figure 2).

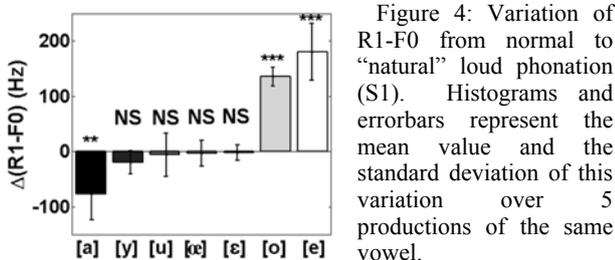


Figure 4: Variation of R1-F0 from normal to “natural” loud phonation (S1). Histograms and errorbars represent the mean value and the standard deviation of this variation over 5 productions of the same vowel.

Lastly, one might expect that increased lip opening increases radiation and therefore vocal intensity [6]. A first argument in favour of that hypothesis is that lip spreading in S1 increases as much as lip aperture (see Figure 3), although lip spreading is not as directly related to the first two formants and may be increased for reasons other than phonetic purposes. Another argument comes from the observation that the increase of lip aperture in “natural” shouted speech (S1) is smallest for vowels with small apertures in normal phonation (see Figure 3), which results in a similar R1 and lip aperture in shouted speech for every open and mid-vowels. This saturation effect might come from a physiological limitation in mouth aperture. More probably, it could also indicate that the speaker tries to reach an “articulatory target”, which appears to be close to the second voice harmonic for normal shouted speech (S1) (see Figure 5). Such a tuning of the first resonance to the second harmonics has already been reported in Bulgarian singing, in which it helps the singers to produce very loud sounds [10]. This tuning is neither observed in normal speech (where it is not needed) nor in S3 (where it is impossible because articulation change is forbidden). In S2, however, R1 appears to be tuned close to 2F0 for [o] and [e], and to 3F0 for [œ] and [ɛ]. Resonance tuning has previously been reported for trained singers [10] but perhaps it might also be an intuitive strategy more widely used by untrained speakers. If so, R1 rises not only as a consequence of vocal intensity, but also as a strategy to increase it.

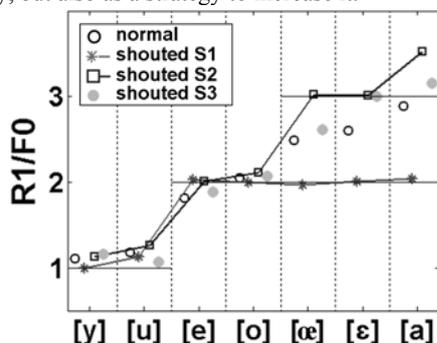


Figure 5: Representation of mean R1/F0 value in normal phonation and 3 different conditions of increasing vocal effort: without instruction (S1), maintaining constant pitch (S2), maintaining constant pitch and articulatory position (S3).

5. Conclusion

In this pilot study of shouted speech, R1 varies with articulation, but also varies with vocal effort when pitch and articulation are kept constant. We have discussed the extent to

which this variation may be related to a rise in the larynx position, a lower open quotient, and an increase of glottal width. The observations are consistent with the hypothesis that the rise of R1 in shouted speech, associated with hyper-articulation, have the effect of tuning R1 to harmonics of the voice (F0 for closed vowels and 2F0 for open and mid-open vowels), which could help in increasing vocal intensity. We have presented the results of a pilot study. We are currently recording some other speakers to verify the generality of these observations. In a future work, it would also be interesting to record tongue movements, in order to explore their contribution to resonances variation in shouted speech, and to verify that such a resonance tuning also happens in continuous speech.

6. References

- [1] Alku, P., Vinturi, J. and Vilkmann, E., "Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation", *Speech Communication*, 38: 321–334, 2002.
- [2] Barney, A., De Stefano, A. and Henrich, N., "The effects of glottal opening on the acoustic response of the vocal tract", *Acta Acustica United With Acustica*, 93: 1046-1056, 2007.
- [3] Epps, J., Smith, J. R. and Wolfe, J., "A novel instrument to measure acoustic resonances of the vocal tract during speech", *Measurement Science and Technology*, 8: 1112-1121, 1997.
- [4] Eriksson, A. and Traunmuller, H., "Perception of vocal effort and distance from the speaker on the basis of vowel utterances", *Perception and Psychophysics*, 64(1): 131-139, 2002.
- [5] Fant, G., "Acoustic theory of speech production." Mouton, La Haye, 1960.
- [6] Garnier, M., Bailly, L., Dohen, M. et al., "An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance", in *Proceedings of ICSLP*, Pittsburgh, PA, USA, 17-22, 2006.
- [7] Gauffin, J. and Sundberg, J., "Spectral correlates of glottal voice source waveform characteristics", *Journal of Speech and Hearing Research*, 32(3): 556-565, 1989.
- [8] Gueumann, A., Kroos, C. and Tillmann, H. G., "Are there compensatory effects in natural speech?" in *Proceedings of ICPHS*, San Francisco, US, 399-402, 1999.
- [9] Henrich, N., d'Alessandro, C., Doval, B. et al., "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation", *Journal of the Acoustic Society of America*, 115(3): 1321-1332, 2004.
- [10] Henrich, N., Kiek, M., Smith, J. et al., "Resonance strategies used in Bulgarian women's singing style: a pilot study." *Logopedics Phoniatrics Vocology*, 32(4): 171 - 177, 2007.
- [11] Holmberg, E. B., Hillman, R. E. and Perkell, J. S., "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice", *Journal of the Acoustic Society of America*, 84(2): 511-529, 1988.
- [12] Howard, D. M., "Variation of electroglottographically derived closed quotient for trained and untrained adult female singers", *Journal of Voice*, 9(2): 163-172, 1995.
- [13] Lienard, J. S. and Di Benedetto, M. G., "Effect of vocal effort on spectral properties of vowels", *Journal of the Acoustic Society of America*, 106(1): 411-422, 1999.
- [14] Lindblom, B. E. and Sundberg, J. E., "Acoustical consequences of lip, tongue, jaw, and larynx movement", *Journal of the Acoustic Society of America*, 50(4): 1166-1179, 1971.
- [15] Pabst, F. and Sundberg, J., "Tracking multi-channel electroglottograph measurement of larynx height in singers", *STL-QPSR*, 33(2-3): 67-78, 1992.
- [16] Plant, R. L. and Younger, R. M., "The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech", *Journal of Voice*, 14(2): 170-177, 2000.
- [17] Rostolland, D., "Phonetic structure of shouted voice", *Acta Acustica*, 51: 80-89, 1982.
- [18] Schulman, R., "Articulatory dynamics of loud and normal speech", *Journal of the Acoustic Society of America*, 85(1): 295-312, 1989.
- [19] Swerdlin, Y., Smith, J. and Wolfe, J., "The effect of whisper and croak vocal mechanisms on vocal tract resonances", submitted to the *Journal of the Acoustical Society of America*