Tuning of vocal tract resonances in shouted voice

[1,2]Maëva Garnier, [1]Marine Barbé, [2]John Smith, [2]Joe Wolfe

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes
[2]School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia

Introduction: Shouting is not just a louder mode of speaking, just as running is not just a faster way of walking. Shouted speech is characterized by spectral transformations and time reorganizations, compared with comfortable speech. In particular, it is produced at higher pitch and with amplified articulatory movements [1].

The relationship between vocal pitch and intensity is well understood and comes from the fact that both parameters are strongly influenced by variations in subglottal pressure [2]. However, fundamental frequency (f_0) can also be controlled by additional parameters such as the contraction of laryngeal muscles, so that pitch and intensity can, however, be varied independently to some extent, which is especially the case in singing [3].

The relationship between vocal intensity and hyper-articulation is less clear, and several phenomena may be involved at the same time: First of all, the sound intensity produced by brass musical instruments is much influenced by the radiation at the bell, which increases sound radiation and produces loud sounds. A larger mouth opening, as observed in loud vocalizations, can then be compared to such a horn and partly explain variations in vocal intensity [4]. A second phenomenon that may be involved is the non-linear source-filter interaction, whereby the vocal tract may influence vocal-fold vibration. Theoretical bio-mechanical models [5] as well as the observation of expert operatic singers [6] have shown that tuning the frequency of one (or both) of the first two vocal tract resonances to one of the first voice harmonics improved efficiency of the vocal fold vibration and "boosted" vocal intensity. The larger mouth and jaw opening movements observed in shouted speech may be considered in that framework, as possible articulatory strategies to tune the first resonance frequency to voice harmonics. A third phenomenon that we envisage would be more related to phonetics and speech intelligibility. It has been established that formant frequencies (F1, F2), considered in isolation, are not simply the acoustic cues on which listeners rely to perceive vowels. The acoustic cue to vowel height, in particular, may be F1-f_0 rather than F1 [7]. In shouted speech, the pitch rise that accompanies the increase of vocal intensity might then degrade the acoustic cue to vowel height if F1 were not also modified. We therefore hypothesize than the larger mouth and jaw opening movements observed in shouted speech may also tend to compensate for this perturbation and maintain the F1-f_0 distance.

The goal of this study was to test these hypotheses in order to better understand the contribution and effect of hyper-articulation in shouted speech.

Material and methods: Eight native speakers of Australian English (4M, 4F, from 22 to 65 years old) participated in the study. The corpus consisted of six vowels of Australian English: [a], [Ɛ], [y], [u], [o] and [oe]. Each vowel was first sustained for 4 seconds in normal phonation, then for 4 seconds in loud phonation, in the same breath. 3 sessions were recorded: In the first (S1), the participants were only told to produce the vowel at normal, then loud level. In the second (S2), they were told to keep a constant pitch when increasing vocal effort. In the third session (S3), they were instructed to keep both pitch and articulation constant. Each vowel was produced 5 times in every session, at both normal and loud level.

Five signals were simultaneously recorded: 1. the audio signal with a pressure microphone (Brüel and Kjær 4944-A) mounted on a stand and resting on the participant's lower lip. Sound pressure level (SPL) was calibrated using a measuring amplifier (Nexus, Brüel & Kjaer); 2. the vocal tract resonances, following the method presented in [6], with a frequency resolution of 11 Hz between 200 and 3000 Hz; 3. the electroglottographic signal (EGG) and 4. the larynx vertical movements, using with a two-channel Electro-Glottograph (EG2 Glottal Enterprise); 5. images of the participants at a rate of 25 images/s, using a video camera placed 30 cm in front of the speaker's face.

Acoustic and physiological descriptors were extracted from these signals: mean SPL, fundamental frequency (f_0), frequency of the first two vocal tract resonances (f_{R1} and f_{R2}), lip aperture (LA) and spreading (LS) and vertical displacement of the larynx (ΔLH).

The data were analyzed with mixed models using R software, considering the phonation mode (normal vs. shouted), the vowel ([a], [Ɛ], [y], [u], [o] and [oe]), the session (S1, S2 or S3) and the gender as fixed effects and the participant as a random effect.

To investigate resonance tunings, we searched for the closest voice harmonics if_0 and jf_0 to the first two resonance frequencies f_{R1} and f_{R2}. For each speaker, vowel and session, we examined the distribution of the frequency distances $(f_{R1} - if_o)/ if_o$ and $(f_{R2} - jf_o)/jf_o$ (defined on [-1,1]). We considered that the speaker was tuning f_{R1} and f_{R2} to a voice harmonic when the distribution showed a single peak located in the central [-0.25 0.25] interval. For each vowel, session and gender, we then counted the number of speakers who demonstrated such a resonance adjustment of f_{R1} or f_{R2}.

Results: We first ensured that participants complied with the experimental instructions in the three sessions: Indeed, they significantly increased their vocal intensity between the 4 first seconds and the 4 last ones of each sustained vowel. That increase, however, was much greater in S1 (+14.9 dB) compared with S2 and S3 (+8.4 dB). They were also able to maintain f_o constant in S2 and S3, whereas f_o increased by 104 Hz on average in S1. Lastly, they were able to maintain their lip articulation whilst increasing vocal intensity in S3, whereas LA and LS increased significantly in the first two sessions (both +6 mm in S1 and +3 mm in S2). None of the participants showed a significant variation in larynx height with the increase of vocal intensity in S3 either.

As expected, all the participants showed a significant increase of f_{R1} when increasing vocal intensity in the S1 (+118 Hz). That increase did not depend significantly on the vowel but was significantly reduced in the two last sessions (S2: +44 Hz; S3: +45 Hz). It is interesting to note that the increase of f_{R1} was still significant in S3, and comparable with that observed in S2, although lip articulation and larynx height did not vary. The second resonance frequency f_{R2} showed a similar trend in women. For men, however, it followed that tendency for [u] and [o] only, and still increased for [a] and [oe] in S1 only, but did not vary significantly in the other cases.

When looking at the frequency distance between f_{R1} and f_{R2} and their closest harmonics, we observed that resonance frequencies were not frequently found in the proximity of a voice harmonic for the vowels produced at comfortable level. At shouted level, however, f_{R2} was frequently observed close to a voice harmonic (whose rank varied with speaker's gender and vowel) for both genders and the two first sessions S1 and S2, and less frequently in the last session S3. Tunings of f_{R1} at shouted level were globally less frequent than those of f_{R2}, but still significantly more frequent than for comfortable speech. Their rate of occurrence was comparable in the three sessions for women, and greater in S1 and S2 than in S3 for men.

Finally, when looking at the frequency distance between f_{R1} and f_0, we observed for both men and women that this acoustic cue was maintained roughly constant from comfortable to shouted level in S1, but not in the last two sessions, where it increased significantly when shouting.

Discussion: The results of this experiment indicate that when a speaker shouts "normally", his/her amplified mouth and jaw movements indeed contribute not only to tune one of his/her first two vocal tract resonances to the frequency of a voice harmonic – which may be beneficial for vocal intensity and efficiency – but also to maintain the frequency distance between F1 and f_o – which helps preserve vowel intelligibility (at least, the vowel height feature). However, when speakers are instructed to maintain their pitch constant when shouting, they still amplify lip and jaw apertures. This invalidates our third hypothesis that hyperarticulation in shouted speech would primarily aim at preserving the F1–f_o phonetic cue to vowel height.

On the other hand, the fact that resonance tunings on voice harmonics were rarely observed at comfortable phonation level and less frequently observed when articulation was constrained (S3) compared to "natural" shouting (S1), supports our second hypothesis that lip and jaw hyper-articulation in shouted speech may primarily or partially have the effect of coupling the vocal tract with the glottal source, which may facilitate vocal fold vibration as well as boosting the energy of lower harmonics.

References:

[1] Schulman, R. (1989). Articulatory dynamics of loud and normal speech. The Journal of the Acoustical Society of America, 85(1), 295-312.

[2] 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.

[3] Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal vibratory mechanisms: The notion of vocal register revisited. Journal of voice, 23(4), 425-438.

[4] Flanagan, J. L. (1960). Analog measurements of sound radiation from the mouth. The Journal of the Acoustical Society of America, 32(12), 1613-1620.

[5] Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. The Journal of the Acoustical Society of America, 123(4), 1902-1915.

[6] Henrich, N., Smith, J., & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. The Journal of the Acoustical Society of America, 129(2), 1024-1035.

[7] Eriksson, A. and Traunmuller, H., "Perception of vocal effort anddistance from the speaker on the basis of vowel utterances", Perception and Psychophysics, 64(1): 131-139, 2002.