Progress Report – Extracting Vowel Quality from Violin Performance

Robert Mores

HAW Hamburg, 22081 Hamburg, E-Mail:mores@mt.haw-hamburg.de

Introduction

Vowel–quality (VQ), a feature of human voice, is used as a meaningful timbre descriptor for stringed sounds. When asked to imitate a violin sound with one's own voice, a listener will automatically adjust the pitch but also the VQ. The VQ is therefore a powerful feature as it represents a good piece of what is perceived by listeners. This feature is intelligible as it can be understood by everyone with normal hearing and speaking abilities. VQ can be extracted by simply taking notes in a listening session, without any further tool. Here, automatic extraction is applied to recorded sounds to gain a broader view of VQ in violins sounds in general. This paper summarizes the progress achieved so far as well as some results.

Method

Results have been achieved on the basis of this approach: a) a tool for extracting VQ from speech is developed and b) validated against trusted reference data, c) VQ results are represented in the chart of the International Phonetic Association (IPA), d) the analysis is extended to process violin sound instead of voice, e) this extension is validated through perceptual tests, f) a separate study investigates the impact of other variables in performances. Trusting this approach, VQ results for some Italian violins are presented.

a) Formant extraction forms the basis for VQ extraction, following established methods. After resampling to f_s = 11025 Hz analyses are done on 40 ms long Blackman–windowed sections of sound. F0 is extracted according to the YIN method [1], F1 to F2 are extracted by linear prediction of order N = 13, standard root-solving according to Atal [2] and formant separation along bandwidth criteria according to Dunn [3] but with wider bandwidths, 500 Hz for F1 and 600 Hz for F2, according to findings of Kim [4].

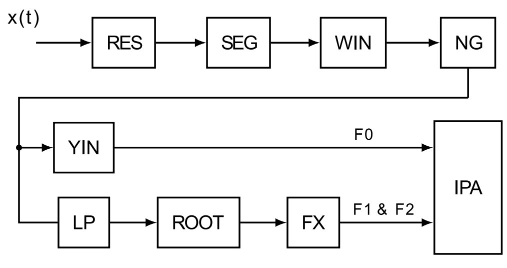

FIG. 1: Block diagram for VQ extraction from a time series x(t): resampling (RES), Blackman–windowed sound segmentation (SEG + WIN), noise gate (NG), F0 extraction (YIN), linear prediction (LP), root solver (ROOT), formant extraction (FX), IPA chart.

b) The performance of the analysis is validated against the Michigan vowel data corpus MVD [5] where extracted F1 (F2) frequencies deviated by more than 150 (250) Hz in only 1 (5) cases, respectively, for the 1077 sound samples of 12 different /hVd/ utterances presented by 45 men and 48 women. Another validation uses the TIMIT data corpus (61238 vowels, 13826 semivowels, 9981 nasals, and 6968 CVt and VCt sounds, transitions between vowels and consonants) in combination with ground truth data of the MSR-UCLA VTR-Formant Database [6]. F1 and F2 are as precisely identified as with the reference tool Wafesurfer.

c) Frequencies of the fundamental F0 and of formants F1 and F2 are finally transformed to height and backness in IPA chart representations using Pfitzinger's formula [7]:

height = 2.61 * ln(F0) - 9.03 *ln(F1) + 47.9

backn = -0.486 *ln(F0) + 1.74 *ln(F1) - 8.39 *ln(F2) + 59.2

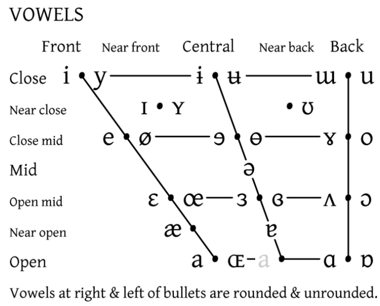

FIG. 2: IPA chart representations of vowel–quality

The IPA chart is preferred against F1/F2 diagrams, as it intuitively translates to tongue position and jaw opening, and a reader can directly imagine the sound.

d) Necessary tool extensions for a transition to strings are: (i) pitch range extension up to 850 Hz, (ii) pitch / resonator decomposition by deconvolution, (iii) modified bandwidth criteria for formant separation.

e) These extensions are verified by perceptual tests. Subjects have been asked to match violin sounds to IPA reference speech sounds. VQ from perceptual tests matches well with those extracted from violin sounds across the total IPA chart.

f) Extracting VQ from performance implies the challenge of working with additional independent variables: (i) the individual performance of a musician, (ii) the music, (iii) room and recording parameters. During investigations, musicians showed a considerable degree of repeatability even across various sessions [8]. Separate studies on the distribution of extracted formant frequencies show that varying the microphone distance has a much lower impact on distributions than the choice of violin.

Exemplary Analysis

The tone G4 is played on the "Willemotte" Stradivari with some vibrato. Figure 3 shows the spectrum of extracted LP coefficients (top) for individual 40 ms sections of sound and the PSD spectrum across the entire 1.8 seconds of sound (bottom). Formant F1 is relative strongly represented, in accordance with the existence of the C2 and C3 modes of a violin. F2 and F3 are situated in the frequency range from

1.5 to 2.5 kHz where there are enough supporting air and plate modes in the violin. Note that F4 and F5 are situated in the region of the bridge hill, in agreement with general findings for the Belcanto character. When listening to the sound sample, the perceptual match of VQ is convincing. However, not all violin sounds are that distinct. The note A3 on the same violin reveals LP and PSD data according to Figure 4, resulting in a non-distinct VQ, see Figure 5 on the right. When listening to such violin sounds, no distinct vowel comes to mind, but a range of vowels can be imagined. In a systematic search among eight violins across different levels of quality and ranging from old to new, 45 out of 120 random choice sound samples revealed a distinct VQ [9].

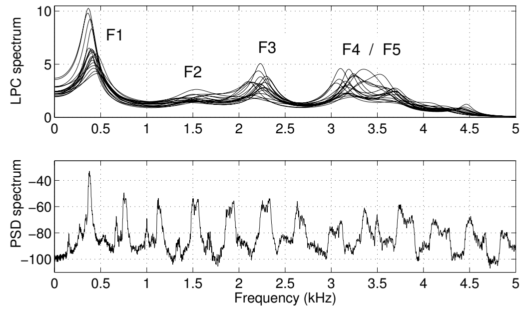

FIG. 3: G4 played on the "Willemotte" Stradivari, top: LPC spectrum for individual 40 ms sections, bottom: PSD spectrum of 1.8 sec. of sound

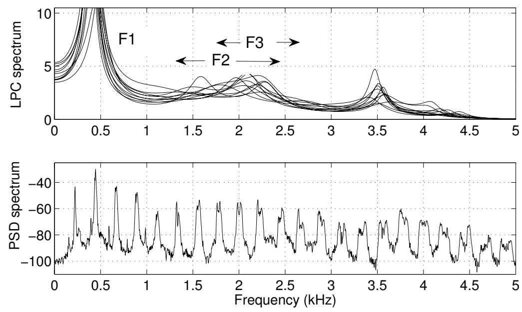

FIG. 4: A3 played on the "Willemotte" Stradivari, top: LPC spectrum for individual 40 ms sections, bottom: PSD spectrum of 1.4 sec. of sound

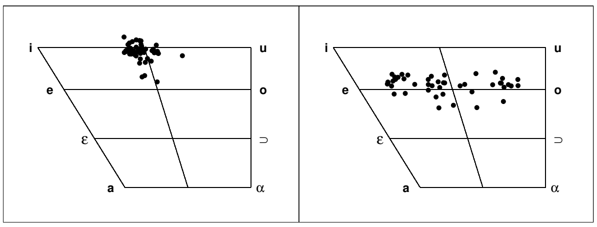

FIG. 5: IPA charts for G4 (left, corresp. to Figure 3) and A3 (right, corresponding to Fig. 4) played on the "Willemotte" Stradivari

Some Italian Violins

Figure 6 represents the VQ as extracted from recordings of the Bruch G minor violin concerto [10]. The violins analysed are: a violin from Andrea Amati dated between 1560 and 1570, Nicolo Amati from 1656, "Joachim" Stradivari from 1714, "Spanish" Stradivari from 1677, "Gibson" Guarneri del Gesù from 1734, "Ex.-Vieuxtemps" Guarneri del Gesù from 1739.

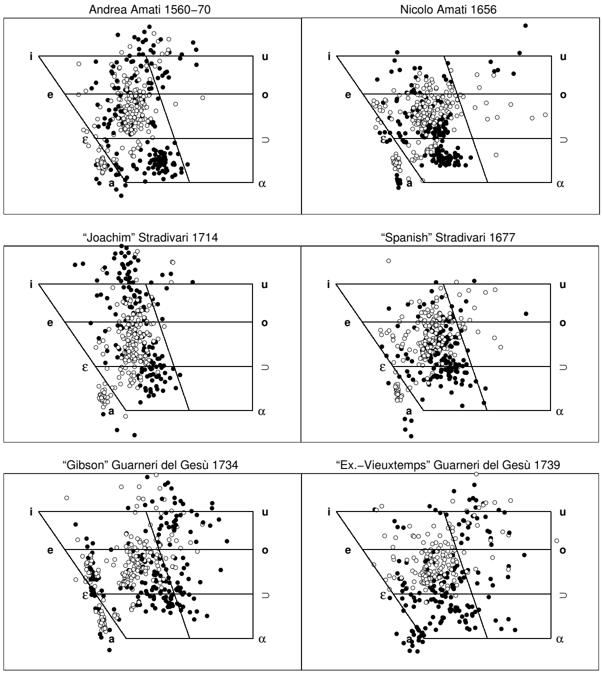

FIG. 6: IPA representations of VQ in violin sounds on an excerpt of Bruch's G minor concerto, black circles: note played between G3 and F#4, white circles: note played between G4 and G5

Observations are: (1) the identified VQ is well within the space of the IPA chart. The populations of VQ for strings match that of voice, (2) there are similarities and differences between violins: the two Amati violins have a rather distinct open /a/ character, the violin from Andrea Amati reveals a similar population as the "Spanish" Stradivari, both are from the end of the 17th century, both Stradivari violins have their VQ definitely in the front region, while both Guarneri violins clearly reveal a wider distribution from front to back.

References

[1] Cheveigne, A. Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music," JASA **111**, 1917-1930.
[2] Atal, B. S., Hanauer, S. L. (1971). "Speech analysis and synthesis by linear prediction of the speech wave," JASA **50**, 637ff.
[3] Dunn, H. K. (1961). "Methods of measuring vowel formant bandwidths," JASA **33**, 1737–1746.
[4] Kim, C., Seo, K., Sung, W. (2006). "A robust formant extraction algorithm combining spectral peak picking and root polishing." EURASIP Journ. on Appl. Signal Proc., Article ID 67960, 1-16.
[5] Hillenbrand, J., et al. (1995). "Acoustic characteristics of american english vowels," JASA **97**, 3099–3111.
[6] L. Deng, X. Cui, R. Pruvenok, J. Huang, S. Momen, Y. Chen, A. Alwan, "A database of vocal tract resonance trajectories for research in speech processing, in: Proc. ICASSP, 2006, pp. 60–63.
[7] Pfitzinger, H. (2005). "Towards functional modelling of relationship between the acoustics and perception of vowels," ZAS papers in Linguistics **40**, 133–144.
[8] Mores, R. (2010). "Measuring Perceived Distance of Violins - A Direct Scaling Test," Proc. 38th Int. Conf. Audio Eng. Soc., Audio Quality Evaluation, Pitea, Sweden, June 13-15.
[9] Mores, R. (2010). "Vowel Quality in Violin Sounds," In: R. Bader, C. Neuhaus, U. Morgenstern (eds.): "Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology", Peter Lang Verlag.
[10] Ricci, R. "The glory of Cremona", Decca.