Some Phonatory and Resonatory Characteristics of the Rock, Pop, Soul, and Swedish Dance Band Styles of Singing

*D. Zangger Borch and †Johan Sundberg, *Luleå, and †Stockholm, Sweden

Summary: This investigation aims at describing voice function of four nonclassical styles of singing, Rock, Pop, Soul, and Swedish Dance Band. A male singer, professionally experienced in performing in these genres, sang representative tunes, both with their original lyrics and on the syllable /pae/. In addition, he sang tones in a triad pattern ranging from the pitch Bb2 to the pitch C4 on the syllable /pae/ in pressed and neutral phonation. An expert panel was successful in classifying the samples, thus suggesting that the samples were representative of the various styles. Subglottal pressure was estimated from oral pressure during the occlusion for the consonant [p]. Flow glottograms were obtained from inverse filtering. The four lowest formant frequencies differed between the styles. The mean of the subglottal pressure and the mean of the normalized amplitude quotient (NAQ), that is, the ratio between the flow pulse amplitude and the product of period and maximum flow declination rate, were plotted against the mean of fundamental frequency. In these graphs, Rock and Swedish Dance Band assumed opposite extreme positions with respect to subglottal pressure and mean phonation frequency, whereas the mean NAQ values differed less between the styles.

Key Words: Formant frequencies–Closed phase–MFDR–NAQ–Inverse filtered–Nonclassical–Popular music–Phonation threshold pressure–Singing styles–Subglottal pressure.

INTRODUCTION

Voice usage may differ substantially between singing styles. Some styles, particularly those using strong glottal adduction and high subglottal pressures, henceforth P_{sub}, are commonly regarded as potentially harmful to the phonatory mechanism.[1] At the same time, several artists in the nonclassical styles of singing have had careers extending over several decades of years. Moreover, several voice pedagogues successfully train singers in such styles, thus suggesting that they can be produced in nonharmful manners.

Using electromyography and electroglottography (EGG) on a single subject, Estill[2] compared physiological characteristics of "belt" and operatic voice quality. They found a longer vocal fold contact phase, determined by means of an EGG signal, and greater vocalis activity in belt. Evans and Howard[3] made similar observations. Bestebreurtje and Schutte[4] analyzed resonatory and phonatory properties in a single subject performing in belt and in a "speech-like" style. They found that the contact phase, also in this case measured from the EGG signal, typically exceeded 50% in belt. However, Lebowitz and Baken[5] found the contact phase did not differ systematically between belt and legit styles, being near or lower than 50% in both styles. Schutte and Miller[6] compared voice source and spectrum characteristics of classical and nonclassical styles of singing and observed difference in P_{sub}, larynx position, and vocal fold adjustment.

The assumed potential risk of singing in some nonclassical styles concerns the phonatory function. Analysis of this function is feasible by means of the inverse filtering strategy, which derives the transglottal airflow from the radiated flow signal by elimination of the effects of the formants.[7] Using this technique, Sundberg and Thalén[8] found that phonation was more similar to deliberately hyperfunctional phonation in the Blues style than in the Pop, Jazz, and, particularly, the Opera styles. Björkner[9] reported that, at comparable relative P_{sub}, opera singers had a stronger voice source fundamental and a shorter closed phase than musical theater singers.

Most of the above investigations have studied single tones produced in different singing styles. In the present investigation, we applied a musically more realistic setting by analyzing phonatory (ie, voice source) and resonatory (ie, formant frequency) characteristics used by an experienced artist when singing real tunes in different nonclassical styles. More specifically, the question we ask in the present investigation is: What are an experienced performer's phonatory and resonatory voice properties when singing in different popular music?

METHOD

Recordings

Coauthor DZB, active as performer and voice teacher in the popular music styles for almost 25 years, served as the single subject. He performed two tasks:

1. At least three sequences of tones sung on the syllable /pae/ during a diminuendo on each of the pitches of a Bb major triad (fundamental frequencies $F0 \approx 117$, 147, 175, 233, and 294 Hz, approximately), see Figure 1. These sequences were recorded first in neutral phonation and then in pressed phonation.
2. A series of phrases representing the Rock, Pop, Soul, and Swedish Dance Band singing styles, see Table 1. The last mentioned is a low-effort style, typically used by vocalists who sing dance music melodies for several hours in hotels on Saturday nights. The examples were first

Accepted for publication July 27, 2010.
From the *Department of Music and Media, Luleå University of Technology, Luleå, Sweden; and the †Department of Speech Music Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden.
Address correspondence and reprint requests to Johan Sundberg, PhD, Department of Speech Music Hearing, School of Computer Science and Communication, KTH, SE-100 44 Stockholm, Sweden. E-mail: pjohan@speech.kth.se
Journal of Voice, Vol. ■, No. ■, pp. 1-6
0892-1997/$36.00

doi:10.1016/j.jvoice.2010.07.014

FIGURE 1. Tone sequences used for the subject's triad task. The sequences were sung three times on each of the pitches Bb2, D3, F3, Bb3, D4.

performed with the original lyrics, and thereafter with the syllables of the lyrics replaced by the syllable /pae/.

The recordings were made in the anechoic room of the Department of Linguistics, Stockholm University so as to eliminate possibilities of room reflections that might affect the result of the inverse filtering. Because the voice source is heavily influenced by variation of subglottal pressure, it was deemed advantageous to record the diminuendo task, allowing voice source analysis for different subglottal pressures. The /pae/ version of the song phrases allowed measurement of the subglottal pressures associated with observed voice source characteristics.

As illustrated in Figure 2, audio, EGG, and oral pressure were digitized and recorded on separate channels in computer files, using the *Soundswell* software (HiTech Development AB, Stockholm, Sweden). Figure 3 shows an example of the recorded files. Audio was picked up by a B&K condenser microphone (4003) located at a distance of 100 cm from the subject's mouth, EGG by a Glottal Enterprises (Syracuse, NY) EGG 2 two channel electroglottograph, and oral pressure by a GAELTEC S7b pressure transducer. Calibration of sound level was made by a Bruel & Kjaer (Naerum, Denmark) calibrator capsule producing a reference sound pressure level of 93.8 dB. Calibration of pressure was done by immersing the pressure transducer in a water tank, at a measured distance below the surface. Flow was calibrated by means of the Glottal Enterprises flow calibration gadget. All calibration signals were recorded in computer files together with announcements of the values recorded.

Analysis

Oral pressure during /p/ occlusion is a good approximation of P_{sub} (see eg, Ref. 10). It was measured by means of the *Soundswell* software package. Long-term average spectra (LTAS) were analyzed by the Spectrum Section program in *Soundswell*, using an analysis bandwidth of 300 Hz. Equivalent sound level was measured by means of the Soundswell Histogram program and fundamental frequency $F0$ extracted by means of the Soundswell Corr program.

TABLE 1.
Phrases Chosen as Representative of the Various Styles Analyzed

Rock: Verse 1 from *Crying in the rain*, lyrics and music: D Coverdale.
Pop: Verse 1 from *Your song*, lyrics: Bernie Taupin, music: Elton John
Soul: Verse 1 from *His eye is on the sparrow*, lyrics: Civilla D. Martin, music: Charles H. Gabriel
Swedish Dance Band: Chorus of *Inget stoppar oss nu*, lyrics and music: Ingela Forsman and Lasse Holm

Voice source analysis was performed by means of the custom-made *Decap* software (Svante Granqvist) described in detail elsewhere.[11] The program offers a choice between analysis of audio (ie, pressure) or flow waveforms. We analyzed the audio signal, because the flow signal contained some noise. For each pitch and style sample, a period from the quasi-steady state in the beginning of the vowel was selected. As P_{sub} varied between the tones, the P_{sub} values associated with the analyzed flow glottograms was somewhat different from the P_{sub} measure observed during the preceding /p/ occlusion. This error was estimated to be less than about 1 cm H_2O if the analysis was made in the beginning of the vowel, just after the waveform had stabilized after the /p/ consonant.

In the present application, the *Decap* program displayed the inverse filtered flow waveform and its derivative and the corresponding spectra before and after the inverse filtering, see Figure 4. In addition, the program also displayed the derivative of the EGG signal (dEGG). For tuning the inverse filters, a ripple-free closed phase and a smoothly falling source spectrum envelope were used as the main criteria.[11] After a delay of the dEGG signal that corresponded to the travel time of the audio signal from the glottis to the microphone, synchrony was obtained between the main negative peak of the dEGG and the final end of the closing phase of the flow glottogram.[12]

After tuning the inverse filters, the filtered signal, that is, the flow glottogram, was saved together with its derivative and the dEGG signal in a new file. In a separate file were saved the frequencies and bandwidths of the inverse filters used. These values offer particularly reliable formant frequency data, because they are determined on the basis of both spectrum and waveform information.

The flow glottograms were then analyzed with respect to different parameters that could be assumed to vary between the styles analyzed. These parameters were closed phase, pulse amplitude, and maximum flow declination rate (MFDR), that is, negative peak value of the derivative of the flow glottogram, as illustrated in Figure 4. The closed phase and the pulse amplitude are both affected by changes of glottal adduction and were hence assumed to be relevant.[13] MFDR represents the excitation of the vocal tract and is therefore closely related to the sound pressure level of the radiated vowel.[14] In addition, the normalized amplitude quotient (NAQ) was calculated, defined as the ratio between pulse amplitude and the product of period and MFDR. It has been found to correlate with the degree of perceived pressedness.[15]

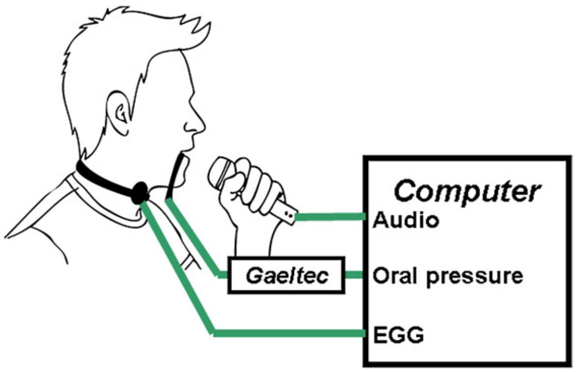

FIGURE 2. Recording setup.

The inverse filtering could mostly be performed without problems. However, the ripple in the closed phase could not be entirely eliminated in some cases. The cause of this ripple, observed mainly in the pressed samples, may be resonance in the flow mask.[16]

As only one subject produced all examples analyzed, it was relevant to find out to what extent the examples were typical for the respective styles. Hence, a listening test was carried out in which seven subjects, all experienced teachers in the non-classical styles of singing concerned, were asked to classify on response sheets the analyzed examples as Pop, Soul, Rock, or Swedish Dance Band. A test file was edited with tones taken from the /pae/ version of the songs, separated by 2-second long pauses. The durations of the examples varied between 0.12 and 1.5 seconds. Each example occurred twice in the file and all examples occurred in random order, one for each listener. The duration of the entire file was 3.25 minutes.

RESULTS

Listening test

Table 2 lists the percentages of consistent classification, that is, the occurrence of the same classification of both presentations of the same stimulus. On average across listeners, 65%

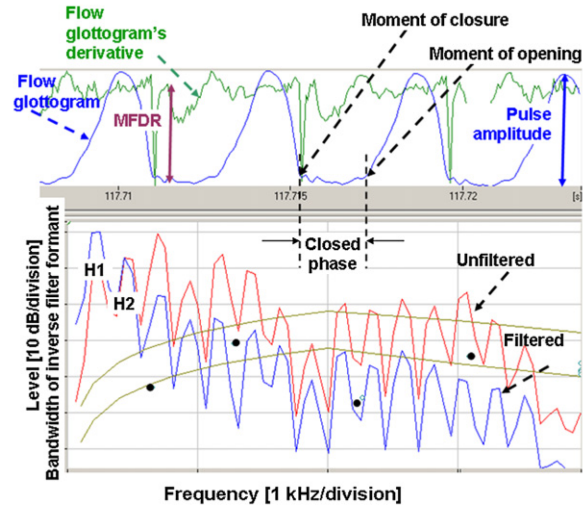

FIGURE 4. *Decap* display showing in the upper panel the waveforms of the inverse filtered flow and its derivative. The lower panel shows the spectra of the unfiltered and filtered flow signal. The *x*-coordinates of the small filled circles show the frequencies of the formants and their *y*-coordinates their bandwidths on an arbitrary scale. The two smooth curves represent the variation range of normal bandwidth values.

(standard deviation [SD] 19%) classifications were consistent. However, listener 5 produced consistent classification in no more than 33%. This seemed to be because of practical problems during the test. Therefore, her results were discarded. The percent of consistent classifications then amounted to 70% (SD 16%).

Results are shown in terms of a confusion matrix in Table 3. Swedish Dance Band was classified as Pop almost as frequently as Swedish Dance Band, and Pop was frequently classified as Swedish Dance Band. Soul and in particular Rock were mostly classified as the styles intended by the singer. The test thus showed that the examples of Swedish Dance Band and Pop were difficult to separate, whereas the examples of Soul and

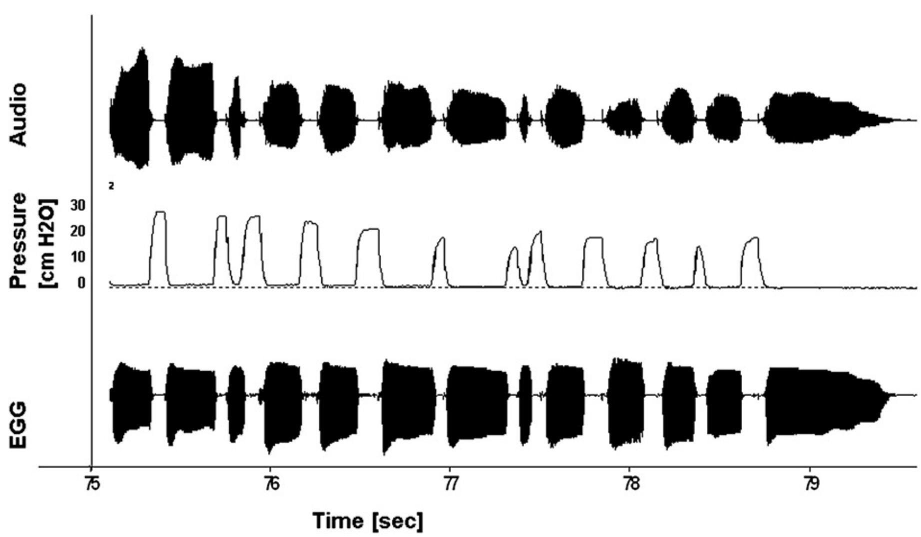

FIGURE 3. Example of recording showing from top audio, oral pressure, and EGG.

TABLE 2.
Percentage of the Experts' Consistent Classifications of Repeated Stimuli in the Listening Test

Listener	Consistent Classifications (%)
1	48
2	72
3	56
4	66
5	87
6	67
7	92

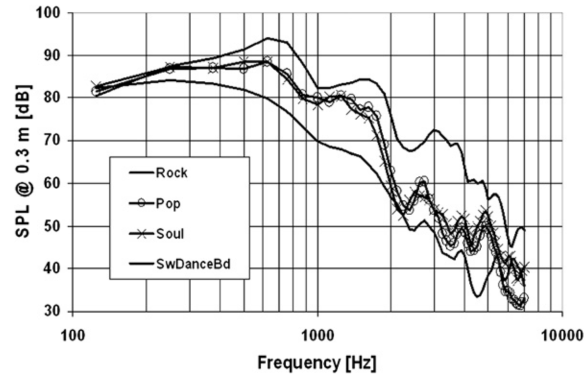

FIGURE 5. Long-term average spectra (LTAS) of the four songs representing the four indicated styles.

Rock were easy to identify as such and hence seemed quite representative.

LTAS

LTAS curves for the tunes sung in the four different styles are shown in Figure 5. Rock produced the highest and Swedish Dance Band the lowest curve, with Pop and Soul assuming intermediate positions. In the low-frequency region, Swedish Dance Band and Rock showed their highest levels near 250 and 700 Hz, respectively, whereas the curves for Pop and Soul peaked at 600 Hz. In the high-frequency region, Pop, Soul, and Swedish Dance Band showed a peak near 2700 Hz, whereas Rock showed one close to 3000 Hz. At 3000 Hz, the level difference between Swedish Dance Band and Rock was no less than 24 dB.

Formant frequencies

Averages of $F2$, $F3$, and $F4$ for the four styles and for neutral and pressed phonation are plotted as a function of mean $F1$ in Figure 6. $F1$ and $F2$ varied by 23% and 8%, respectively, between the styles, Rock and Soul showing the highest and lowest values of both $F1$ and $F2$, respectively. Soul, Swedish Dance Band, and neutral were produced with low $F1$ values, whereas Pop, Rock, and pressed were sung with high $F1$ values. The articulatory characteristics producing these differences would include larynx height and jaw opening.

Subglottal pressure

Figure 7 shows as a function of $F0$ the highest and lowest P_{sub} values observed in pressed and neutral phonation for the softest and loudest triad patterns. P_{sub} ranged between 2 and 53 cm H_2O and increased with $F0$, particularly in loud phonation. Both the lowest and the highest pressures were clearly higher in pressed than in neutral phonation, as expected. The figure also shows phonation threshold pressure (PTP_T) for male adults, calculated by means of Titze's equation using a mean speech $F0$ value of 120 Hz.[17] The graph also shows PTP_T multiplied by 1, 3, 6, and 9. In loudest pressed, the singer's pressures approached 9*PTP_T, while those used in loudest neutral were closer to 6*PTP_T.

The average P_{sub} observed in the examples of the four styles are shown by the centers of the ellipses, the axes of which correspond to ±1 SD in $F0$ and in P_{sub}, respectively. The averages fall between 3* PTP_T and 6* PTP_T. The Rock and Swedish Dance Band styles were extreme, with respect to both mean $F0$ and mean P_{sub}, while Pop and Soul assumed intermediate values. In Swedish Dance Band style, P_{sub} ranged between 8 and 16 cm H_2O, whereas Rock ranged between 28 and 53 cm H_2O. The corresponding mean $F0$ values were 8 and 20 semitones above A2 (110 Hz), respectively. The centers of the ellipses in the figure lie close to 3* PTP_T for the Swedish Dance Band, close to 4* PTP_T for the Pop and Soul styles, and at 5* PTP_T for Rock. This indicates that the $F0$ differences between the styles do not account for all of the P_{sub} differences between them.

Voice source

The closed quotient measured in the flow glottograms varied much less systematically with P_{sub} than has been found in operatic baritone voices.[18] The variation was similar to that reported for untrained subjects.[19] It was particularly great in the softest phonations in the Rock and pressed styles, presumably because these phonation types are rarely produced with low P_{sub}.

TABLE 3.
Confusion Matrix for the Classification of the Various Examples Included in the Listening Test

	Sum (%)			
	Swedish Dance Band	Pop	Soul	Rock
Swedish Dance Band	44,4	47,6	7,9	0,0
Pop	35,7	52,4	8,7	3,2
Soul	13,5	16,7	69,8	0,0
Rock	0,0	6,3	8,7	84,9

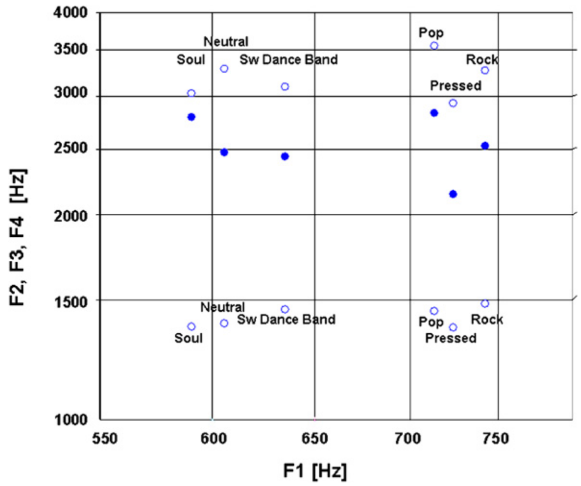

FIGURE 6. Average formant frequencies of the vowel /ae/ as sung in the indicated styles. Mean $F2$, mean $F3$, and mean $F4$ (open, filled, and open circles, respectively) are plotted as functions of mean $F1$. Also, plotted are the corresponding means measured in the neutral and pressed versions of the triad pattern shown in Figure 1.

Figure 8 shows mean NAQ ratio for the triads sung in neutral and pressed phonation together with the mean values for the songs performed in the different styles. As in Figure 6, the means are represented by the centers of the ellipses, the axes of which correspond to ±1 SD in $F0$ and NAQ, respectively.

NAQ varied between 0.1 and 0.17 in neutral and showed no systematic dependence of $F0$. In pressed, it was close to 0.1 for the two lowest pitches and increased slightly with frequency. This is an expected result, as tones produced with low NAQ values tend to be perceived as pressed.[15] The mean NAQ values for the styles were rather similar and all close to the values for neutral, even though the mean NAQ for Rock was clearly closer to but not as low as the values for pressed phonation.

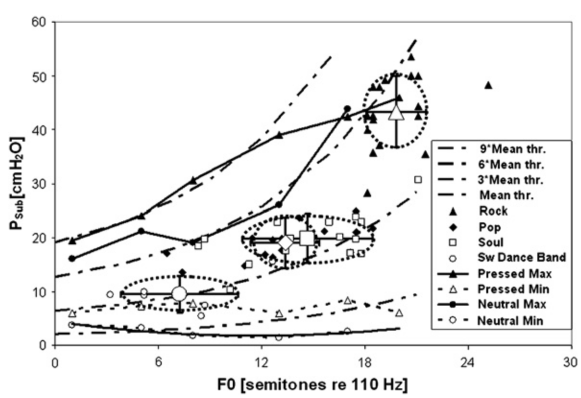

FIGURE 7. Subglottal pressures plotted as functions of log $F0$. Filled and open triangles and circles represent the maximum and minimum pressures observed in the subject's samples of pressed and neutral phonation. Chain-dashed curves show 1, 3, 6, and 9 times Titze's PTP, respectively. The large symbols and the axes of the ellipses show means and SD, respectively, of P_{sub} and $F0$ for the indicated styles, and the smaller versions of the same symbols represent the data.

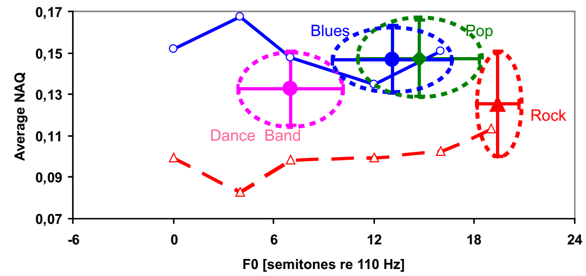

FIGURE 8. NAQ values plotted as a function of log $F0$. The solid and dashed curves refer to the values observed in the subject's neutral and pressed phonation. The large symbols and the axes of the ellipses show the mean NAQ and SDs observed for the vowel [æ] sung in the indicated styles.

DISCUSSION

This investigation analyzed a single subject's productions. Yet, the results can be assumed to be relevant given the fact that the subject has been performing professionally in concerts and studios for a long time without damaging his phonatory mechanism. The listening test showed that the examples of Rock and Soul were easy to classify and hence can be regarded as representative, while the examples of Swedish Dance Band and Pop were sometimes confused.

The NAQ average for the styles ranged between 0.13 and 0.18. This was somewhat surprising. NAQ has been found to decrease with increasing degree of phonatory pressedness,[15] but it has also been found to be lower at low than at high $F0$.[9] The Rock style is typically perceived as more pressed than the other styles, and phonation in the Swedish Dance Band samples sounded much more relaxed than in the Rock samples. Hence, one would expect that the mean NAQ of the Rock style would be lower than that of the Swedish Dance Band style. However, the average $F0$ of the Rock samples was about 12 semitones higher than that of the Swedish Dance Band style. It was also surprising that the mean NAQ of the styles were all similar to those for neutral phonation. This may reflect vocal technique; the singer may have learnt to avoid a hyperfunctional type of phonation for reasons of phonatory hygiene.

The Rock sample had the narrowest and highest pitch range. This is typical for the Rock style of singing. This would promote the impression of high energy, a main characteristic of this style.

Figure 7 compared the average P_{sub} observed in the samples of the four styles with multiples of Titze's PTP_T. This pressure reflects how P_{sub} typically varies with $F0$. The mean pressure of the Swedish Dance Band style was close to 3* PTP_T, whereas the mean pressure of the Rock style approached 6* PTP_T. Björkner[9] found that the highest P_{sub} values used by musical theater and operatic baritone singers were between 36 and 39 cm H_2O. Substantially higher P_{sub} were used by our subject in the Rock samples. At least in part, this would be a consequence of the fact that P_{sub} is typically raised with increasing $F0$. Björkner's data referred to an $F0$ of 278 Hz (pitch close to C#4), whereas the average P_{sub} in our Rock samples were close to 370 Hz (pitch close to F#4).

CONCLUSIONS

Placing different singing styles within a given subject's $F0$ and phonation type ranges showed that Rock and Swedish Dance Band singing styles were extreme about P_{sub} and $F0$. $F1$ and $F2$ were lowest in Soul, possibly reflecting a relatively low larynx position. $F1$ was high in Pop and Rock. The mean NAQ values suggested that phonation type in all styles was close to neutral, except Rock which was closer to, but not as low as pressed. This may be relevant from the point of view of phonatory hygiene.

Acknowledgments

The authors are indebted to the kind and efficient assistance of Peter Branderud and Hassan Djamshidpey during the recordings in the anechoic room of the Linguistics Department of the Stockholm University. This investigation is part of coauthor DZB's doctoral dissertation work in Musical Performance carried out at the Luleå University of Technology with coauthor JS as co-supervisor.

REFERENCES

1. Laukkanen A-M, Sundberg J. Peak-to-peak glottal flow amplitude as a function of F0. *J Voice*. 2008;22:614–621.
2. Estill J. Belting and classic voice quality: some physiological differences. *Med Probl Perform Artists*. 1998;3:37–43.
3. Evans M, Howard DM. Larynx closed quotient in female belt and opera qualities: a case study. *Voice*. 1993;2:7–14.
4. Bestebreurtje ME, Schutte HK. Resonance strategies for the belting style: results of a single female subject study. *J Voice*. 2000;14:194–204.
5. Lebowitz A, Baken RJ. Correlates of the belt voice: a broader examination. *J Voice*. In press.
6. Schutte HK, Miller DG. Belting and pop, nonclassical approaches to the female middle voice: some preliminary considerations. *J Voice*. 1993;7:142–150.
7. Rothenberg M. A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. *J Acoust Soc Am*. 1973;53:1632–1645.
8. Sundberg J, Thalén M. Describing different styles of singing. A comparison of a female singer's voice source in 'Classical', 'Pop', 'Jazz' and 'Blues'. *Logoped Phoniatr Vocol*. 2001;26:82–93.
9. Björkner E. Musical theater and opera singing—why so different? A study of subglottal pressure, voice source, and formant frequency characteristics. *J Voice*. 2008;22:533–540.
10. Hertegård S, Gauffin J, Lindestad P-Å. A comparison of subglottal and intraoral pressure measurements during phonation. *J Voice*. 1995;9:149–155.
11. Salomão G, Sundberg J. Relation between perceived voice register and flow glottogram parameters in males. *J Acoust Soc Am*. 2008;124:546–551.
12. Henrich H, d'Alessandro C, Doval B, Castellengo M. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J Acoust Soc Am*. 2004;115:1321.
13. Sundberg J. *The Science of the Singing Voice*. DeKalb: N. Ill. Univ. Press; 1987.
14. Fant G. *Speech Acoustics and Phonetics*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2004.
15. Sundberg J, Thalén M, Alku P, Wilkman E. Estimating perceived phonatory pressedness in singing from flow glottograms. *J Voice*. 2004;18:56–62.
16. Hertegård S, Gauffin J. Acoustic properties of the Rothenberg mask. *STL-QPSR*. 1992;33(2–3):009–018.
17. Titze I. Phonation threshold pressure: a missing link in glottal aerodynamics. *J Acoust Soc Am*. 1992;91:2926–2935.
18. Sundberg J, Andersson M, Hultqvist C. Effects of subglottal pressure variation on professional baritone singers' voice sources. *J Acoust Soc Am*. 1999;105:1965–1971.
19. Sundberg J, Fahlstedt E, Morell A. Effects on the glottal voice source of vocal loudness variation in untrained female and male subjects. *J Acoust Soc Am*. 2005;117:879–885.