Spectral characteristics of three styles of Croatian folk singing

Paul Boersma[a)] and Gordana Kovacic
Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016CG Amsterdam, The Netherlands

(Received 8 June 2005; revised 16 December 2005; accepted 31 December 2005)

This paper examines the differences between three Croatian folk singing styles, namely *klapa, ojkanje*, and *tarankanje*. In order to factor out singer-specific properties, each of the styles was performed by the same 12 professional male singers. The 36 performances were analyzed with a long-term average spectrum (LTAS) method from which direct effects of the pitch distribution were removed. After factoring out each singer's average, the 36 pitch-corrected LTAS contours were reduced to a two-dimensional representation in two ways: (1) a principal-component analysis and (2) a graphical plot of spectral slope versus speaker's formant strength. Both ways clearly separate the three styles. The spectrum of the klapa style turns out to be similar to that of speech. The ojkanje style is extremely loud and shows two spectral peaks: a sharp one tuned at twice the fundamental frequency and appropriate for long-distance communication on mountain slopes, and a broad one around 3.5 kHz, reminiscent of a speaker's formant. The tarankanje style has a very flat spectrum implemented by vocal pressedness and nasality, which is appropriate for blending into or imitating the timbral characteristics of the sopile folk instrument.
[DOI: 10.1121/1.2168549]

PACS number(s): 43.75.Rs, 43.70.Gr, 43.72.Ar [DD] Pages: 1805–1816

I. INTRODUCTION

Although there exist many more nonclassical than classical singers, most scientific information on the singing voice is based on studies of voices trained in the Western classical tradition, perhaps because its relative uniformity throughout the world allows a comparison of results across the various independent studies contributing to a deeper understanding of this single style. A broader understanding of the singing voice has to involve investigating the acoustics of many nonclassical styles, and some contributions have been made already on styles as diverse as pop (Schutte and Miller, 1993; Doskov *et al.*, 1995; Thalén and Sundberg, 2001; Borch and Sundberg, 2002), Broadway/musicals (Thalén and Sundberg, 2001; Stone *et al.*, 2003), country and western (Burns, 1986; Stone *et al.*, 1999; Sundberg *et al.*, 1999b; Cleveland *et al.*, 2001), jazz and blues (Thalén and Sundberg, 2001), Estonian folk (Ross, 1992), belting (Estill *et al.*, 1994), and overtone singing (Bloothooft *et al.*, 1992; Klingholz, 1993; Lindestad *et al.*, 2001; Van Tongeren, 2002).

The present investigation adds three more styles to the literature on acoustically investigated nonclassical singing: *klapa, ojkanje*, and *tarankanje*. These three very different styles all belong to the traditional music culture of Croatia. The relatively new klapa style (*klapa*="*group of friends*") originated in the 19th century in the Mediterranean part of Croatia (Dalmatia and Dalmatian islands), uses a Western European musical scale, and is usually performed *a cappella* with multiple parts in harmony, typically as soft, slow, serenadelike love songs (Rapanić, 1979; Ćaleta, 1997; Bezić, 1979). The tarankanje style is typical for the Istrian peninsula, the Kvarner islands, and the Croatian Littoral; it uses the Istrian musical scale, which has six narrowly spaced tones impossible to transcribe in the Western musical notation system (Bonifačić, 2001); it accompanies dance and is sung for a large part as strings of meaningless syllables (e.g., *tanana*) that can blend with or replace local wind instruments (Bonifačić, 1996). The ancient ojkanje style [the singers who participated in this study call it *dozivački*, from the verb *dozivati* "call (loudly)"], whose name refers to an *oj*-like syllable that is sung as a loud and tremolous "wild howl" before and/or after a loud short text (Dobronić, 1915; Bezić, 1968; Marošević, 1994), uses narrow non-Western intervals and tends to be perceived by outsiders as shouting or nonmusic (Dobronić, 1915; Ćaleta, 1999; Marošević, 2004); it is distinctive for mountainous Croatia, i.e., the Dinaric region and the Dalmatian hinterland.

The advantage of taking these three styles as the subject of investigation is that there exists a professional ensemble of folk singers that performs all of the three styles. Taking these singers as subjects for the present study allows us to reveal stylistic variation by analyzing the intrasubject differences across the performances of the three styles and factoring out any singer-specific characteristics.

II. METHOD

A. Subjects

A total of 12 male professional folk singers voluntarily took part in the investigation. All were members of LADO Folk Dance Ensemble of Croatia, which has been practicing song and dance from all regions of the country for over half a century. The singers had been performing Croatian folk music as LADO members for a period of 4 to 20 years, with an average of 10 years. Their ages ranged from 24 to 45, with an average of 33 years. None of the subjects had for-

[a)]Electronic mail: paul.boersma@uva.nl

mally studied singing before joining the LADO ensemble. During the time of participation in the study, all singers reported to be in good vocal and physical condition (singer 1 reported just having recovered from a common cold, which is still audible in his klapa performance).

B. Data collection

The recordings were performed in an anechoic chamber of the Department of Electroacoustics of the Faculty of Electrical Engineering and Computing of the University of Zagreb. The background noise level measured inside the chamber was 19 dB(A) as measured by an integrating sound level meter (Brüel & Kjœr 2231).

The subjects were recorded one by one. Each was asked to perform one traditional song from each of the three styles. The songs were selected by the artistic director of the LADO Ensemble, who also chose the "key" for each song. The klapa style was represented by the song *Zaspalo je siroče* from Dalmatia, performed in G-major, the ojkanje style by the song *Mi smo rekli zapivati ode* from the Dalmatian hinterland, and the tarankanje style by the song *Homo u kolo* from Istria. Each singer performed each song three times, but only one performance of each song, namely the one that was judged best both by the singer himself and by the second author of the paper, was selected for acoustical analysis. The criteria were authenticity, stable vocal quality, and the singer's overall satisfaction with his performance.

Each singer performed in a standing position and was instructed to keep a constant distance of 0.3 m between his mouth and the microphone. The signal was recorded with a Behringer ECM 8000 omnidirectional microphone and fed via a TOA D-4 microphone preamplifier to an AIWA HD-S200 digital tape recorder with a sample rate of 44.1 kHz. Electroglottographic data were obtained in a fourth (shorter) performance with a Laryngograph, but these data turned out to be unreliable because of large vertical larynx movements (especially in ojkanje) and are not included in this study.

For each individual singer, the recording was preceded by a test recording in which the gain of the preamplifier was set to the optimal level for that singer. The gain was then kept constant for the three styles in order to make sure that the sound levels of the styles could afterwards be compared for each singer, although the recordings were not calibrated for absolute sound pressure level (for the unexpectedly loud ojkanje performance by singer 2, the recording gain was decreased by 6 dB, which was later corrected by doubling the amplitude).

The recording sessions thus yielded 12 performances of each of the three songs.[1] The average durations of the songs turned out to be 57.16 s for klapa, 50.75 s for ojkanje, and 43.50 s for tarankanje.

C. Acoustic analysis: Pitch-corrected LTAS

The purpose of the recordings was to obtain information on the average and individual production (phonatory and articulatory) and spectral properties (e.g., the presence or absence of a singer's formant) of the three styles. To this end,

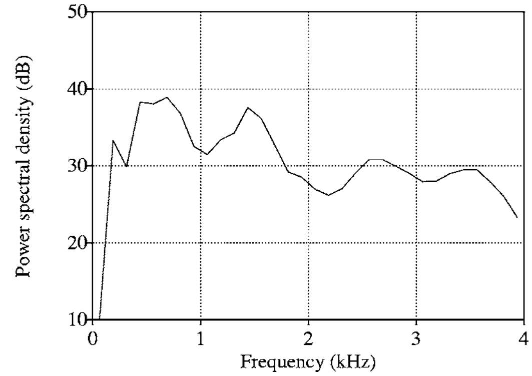

FIG. 1. Pooled LTAS for the 12 tarankanje performances, without pitch correction.

each of the 36 recordings was subjected to a long-term average spectrum (LTAS) analysis, performed with the phonetic analysis program PRAAT (Boersma and Weenink, 2005). Every LTAS was computed with a bin width of 125 Hz and a frequency range of 0–4 kHz. However, a simple LTAS may show undesirable F0-related phenomena. As an example consider Fig. 1, which shows the simple LTAS for tarankanje, pooled over all 12 singers.

The figure clearly shows a peaked spectrum, especially in the second bin (125–250 Hz; the peak is at the center, i.e., at 187.5 Hz), the fourth bin (375–500 Hz), and the sixth bin (625–750 Hz). The individual singers vary in the presence of the second and third peaks, but the first peak is present in the LTASs of all 12 singers. The three peaks could correspond to the first, second, and third harmonics of a fundamental frequency around 220 Hz. This is confirmed by a histogram of the 70 544 F0 values measured by PRAAT for the 12 performances of tarankanje. Figure 2 shows that the largest peak is in the bin between 200 and 210 Hz.

To annihilate the influence of F0, a pitch-corrected LTAS method was designed, and this was used for most analyses in this paper.

The pitch-corrected LTAS procedure is summarized in Fig. 3 and runs as follows. For the voiced parts of the recording (Fig. 3 shows three examples), PRAAT's cycle-to-cycle waveform matching procedure detects all the pitch periods. Each pitch period is subsequently excised (data windowing is not needed, and the phase of the glottal pulse within the excised period does not influence the result). The figure shows the excision of one example period for each of the three voiced parts; the three periods have different dura-

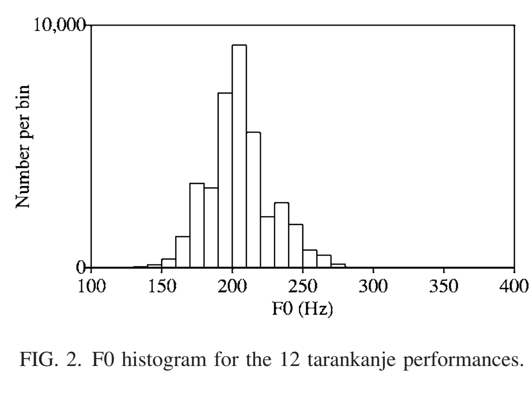

FIG. 2. F0 histogram for the 12 tarankanje performances.

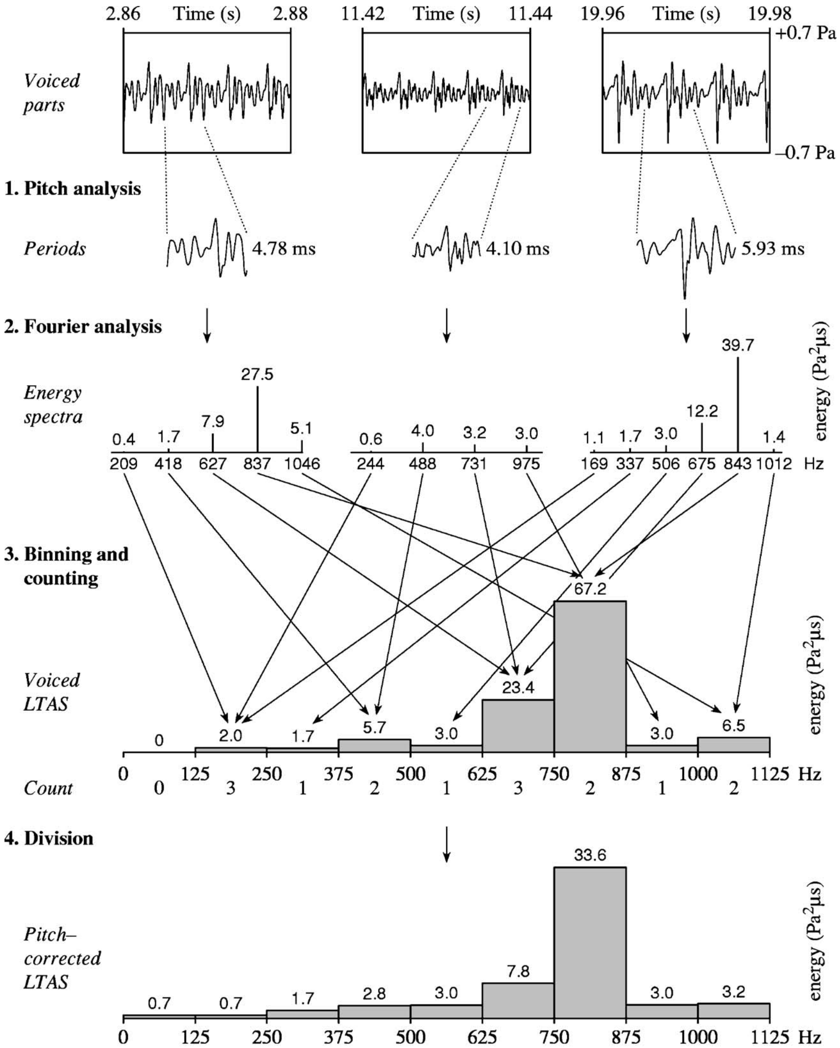

FIG. 3. The pitch-corrected LTAS method.

tions because the three voiced parts have different fundamental frequencies.

In the next step, each period is converted to a line spectrum by Fourier transformation. For instance, a period with a duration of 4.78 ms (one of the examples in Fig. 3) is transformed to a spectrum with a frequency spacing of 1/4.78 ms=209 Hz. This spectrum therefore contains information on energies at the harmonics, i.e., at 209, 418, 627, 837, and 1046 Hz and so on. For instance, the "energy"[2] of the third harmonic (at 627 Hz) is 7.9 $\mathrm{Pa}^2\mu\mathrm{s}$ (if the recording was indeed calibrated between −0.7 and +0.7 Pa, as the waveforms suggest).

The next step in the procedure is binning and counting. Each of the energies is put into the appropriate 125-Hz-wide bin of a LTAS. For the first period shown in the figure, the sixth bin, which runs from 625 to 750 Hz, receives 7.9 $\mathrm{Pa}^2\mu\mathrm{s}$ of energy, because the third harmonic at 627 Hz falls within this bin. Likewise, the high-energy peak of 27.5 $\mathrm{Pa}^2\mu\mathrm{s}$ associated with the fourth harmonic of 837 Hz is put into the seventh bin, which runs from 750 to 875 Hz. When the whole sound has been processed and all the energies have been added into their appropriate bins, the resulting histogram simply represents the LTAS of the voiced parts. However, while the energies in the harmonics of each period are put in their appropriate bins, the procedure also keeps track of how many harmonics leave their energies in each bin. For instance, after the processing of the three example periods there are three pieces of energy that have been put into bin 6 and two pieces of energy that have been put into bin 7. These counts can be read off Fig. 3 by counting the number of arrows that end in each bin and are also explicitly mentioned below the voiced LTAS histogram.

The last step is the actual pitch correction. The total energy in each bin is divided by the number of energies that had been put into that bin. The result is a LTAS in which each bin represents the average energy of the harmonics that entered it. For instance, the 23.4 $\mathrm{Pa}^2\mu\mathrm{s}$ that went into bin 6 is divided by 3 (the number of harmonics that contributed to the energy in this bin), yielding 7.8 $\mathrm{Pa}^2\mu\mathrm{s}$. The 67.2 $\mathrm{Pa}^2\mu\mathrm{s}$ that went into bin 7 is divided by only 2, because only two harmonics contributed to it. This enhances the peak in the seventh bin with respect to the value in the sixth bin.

To arrive at the final pitch-corrected LTAS, the procedure includes three more details. First, bins with a count of zero are subjected to linear interpolation between their neighbors or to constant extrapolation at the edge; in the figure this happens to the first bin. Next, in order to bring the histogram to the same scale as the uncorrected "voiced"

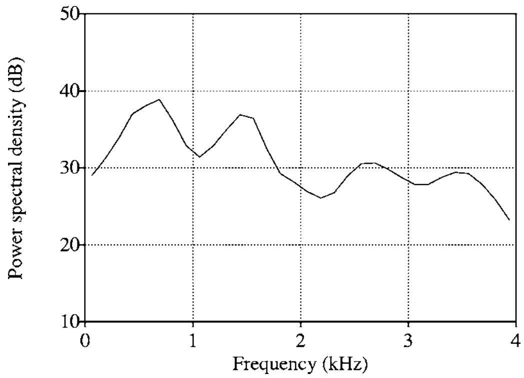

FIG. 4. Pooled LTAS for the 12 tarankanje performances, with pitch correction.

LTAS, the energy in every bin is multiplied by the average number of energies that went into a bin; in the example the multiplication factor would be 1.67, which is the total number of energies (15) divided by the number of bins (9). Finally, the energy in each bin is converted to an energy spectral density (in $Pa^2\,s/Hz$) by dividing it by the bin width (125 Hz), then converted to a power spectral density (in Pa^2/Hz) by dividing it by the total duration of the recording, then directly converted to a value in dB relative to $4.0 \times 10^{-10}\ Pa^2/Hz$. In this way, any calibration of the waveform in units of Pascal is faithfully preserved in the final pitch-corrected LTAS curve, and duration differences between the recordings are compensated for.

Figure 4 shows the final pitch-corrected LTAS curve for tarankanje. The curve is much smoother than the noncorrected LTAS of Fig. 1.

Comparable improvements apply to the other two styles, although the effect of the method is not always smoothing. Sharp spectral features can remain, as can be seen in the ojkanje curve in Fig. 5, which shows a deep and narrow valley around 800 Hz. To sum up, the pitch-corrected LTAS method maximally eradicates direct influences of F0 without sacrificing frequency selectivity. Some indirect influences of F0 (a rising F0 with a constant glottal waveform will lead to a stretching spectral envelope, and a high F0 may involve a raised larynx, which shortens the vocal tract and may thereby raise some formants) still remain visible in the pitch-corrected LTAS.

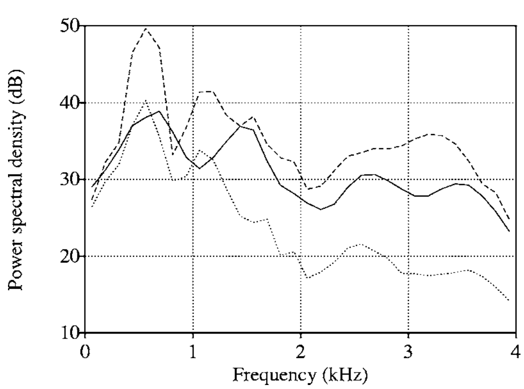

FIG. 5. Pooled pitch-corrected LTASs (—tarankanje,···klapa,---ojkanje).

III. RESULTS

A. Pooled data

Figure 5 shows the pitch-corrected LTASs for each of the three styles, pooled over all 12 singers. Several interstyle differences can be read off this figure. The most noticeable differences are related to sound levels, spectral slopes, and locations of the spectral peaks.

As far as the sound levels are concerned, Fig. 5 shows that klapa is least loud and ojkanje is loudest. This feature is noticeable through the differences between the amplitudes of the strongest spectral peaks at around 0.6 kHz where both the klapa and tarankanje styles have an approximately 10 dB lower amplitude than the ojkanje style. This corroborates ethnomusicological descriptions of the ojkanje style that state that vocal loudness is a dominant feature of this style (Ćaleta, 1999).

The second difference among the styles is in the average slope of the spectrum. The spectral slope is related to the relative speed of glottal closure (Fant, 1960, p. 270), which again is correlated to the vocal loudness mentioned in the previous paragraph [in fact, spectral slope is a main auditory cue for perceived loudness in speech (Sluijter, 1995)]. The spectral slope is indeed steepest for klapa, which also has the lowest vocal intensity; the difference of 20 dB between the low- and high-frequency regions resembles what is generally found in speech (Kuwabara and Ohgushi, 1984; Leino, 1994; Cleveland *et al.*, 2001). Likewise, the spectral slope is flatter for ojkanje, which has the largest vocal intensity; the low-high difference of 15 dB corresponds to that in very loud speech or in shouting (Nawka *et al.*, 1997; Nordenberg and Sundberg, 2003). The tarankanje style is special: it has an unusually flat slope (10 dB) but medium to loud vocal intensity. The flat slope suggests that this style employs a pressed voice (Stevens, 1998, p. 85; Bergan *et al.*, 2004, p. 311).

The third difference between the three styles is in the location and regularity of spectral peaks. The klapa and ojkanje styles have peaks around 0.6 and 1.1 kHz, which correspond to the locations found in speech (of any vocal intensity, including shouting). The tarankanje style is again very different: the location of the two peaks at 0.7 and 1.5 kHz, together with the valley at 1.1 kHz, indicates that the average spectrum corresponds to that of the nasalized low front vowel [æ̃] (Stevens, 1998, p. 311), while the strength of the second peak (almost as high as the first) is probably due both to the pressed voice quality mentioned before and to the raising of the bandwidth of the first formant as a result of nasality (Stevens, 1998, pp. 310, 312).

For the klapa and the tarankanje styles there seems to be a regular pattern of peaks appearing clearly around 2.5 and 3.5 kHz, probably reflecting the third and fourth formants (F3 and F4). This is the spectral region where one could look for the *singer's formant*, a strong resonance phenomenon at about 2.8 kHz, typical of operatic singing voices (Bartholomew, 1934; Sundberg, 1973, 1974), but none of the three spectra show such a strong peak in this area. However, the ojkanje style is characterized by a prominent broad plateau between 2.2 and 3.8 kHz whose local peak is suggestive of a *speaker's formant* (or *actor's formant*), a resonance phenom-

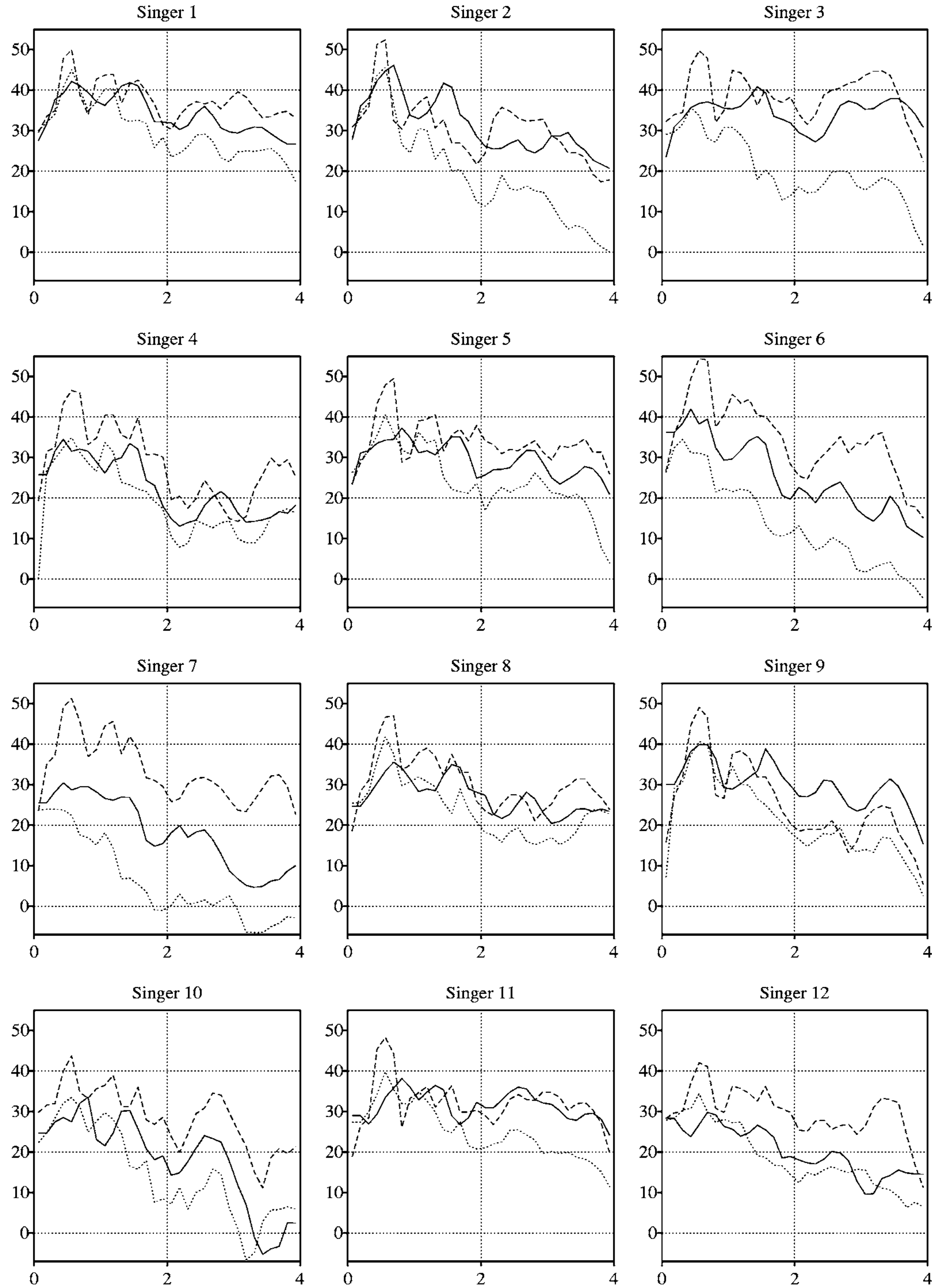

FIG. 6. Individual pitch-corrected LTASs (—tarankanje,···klapa,---ojkanje). Horizontally: frequency in kHz; vertically: power spectral density in dB.

enon that is usually associated with the speech of professional speakers such as radio announcers (Kuwabara and Ohgushi, 1984) and theater actors (Leino, 1994; Nawka *et al.*, 1997; Thunberg, 2003). These authors agree that a speaker's formant is weaker than a singer's formant and associated with a higher spectral region than the singer's formant, namely between 3.0 and 3.8 kHz (for males). In the present study, the amplitude of the peak in ojkanje is about 17 dB lower than the strongest peak of the spectrum. The naming of this peak as a speaker's formant may be problematic, though, since it is found in singing rather than in speaking. Section IV proposes an alternative term.

B. Individual data

While the average interstyle differences can be read off Fig. 5, it is also important to investigate to what extent the differences are consistent across speakers. This can be done informally by visual inspection of the individual pitch-corrected LTASs, and formally by performing a principal-component analysis or a data reduction into preestablished properties of the individual LTASs.

Figure 6 shows the individual pitch-corrected LTASs for each of the 12 singers and each of the three styles. Several properties that were noted for the pooled data of Fig. 5 also appear for many, most, or all of the individual singers: the lowest vocal intensity for klapa (all singers except perhaps singers 4 and 8), the highest vocal intensity for ojkanje (all singers except perhaps 9), the steepest (i.e., most speechlike) slope for klapa (all singers except perhaps 9), the flattest slope for tarankanje (singers 2, 3, 5, 8, 9, and 11), the two speechlike peaks around 0.6 and 1.1 kHz for klapa and ojkanje (all singers), the valley around 1.1 kHz for tarankanje (all singers except 7 and 12), and the speaker's-formant-like broad spectral peak around 3.5 kHz for ojkanje (singers 3, 4, 6, 7, 8, 9, 12, and perhaps 1 and 11). Apart from these main effects of style, the figure shows main effects of speaker (e.g., singer 6 has steeper spectral slopes than singer 1, for all three styles, and singer 10 uses a singer's-formant-like local peak in all styles) and interactions between speaker and style (e.g., singers 6 and 7 make larger loudness differences between the styles than singers 1, 8, and 11).

The main interest here must be in the detectability of the styles on the basis of the differences noted in Sec. III A, i.e., in the consistency between singers with regard to these differences. Two methods are discussed in Secs. III C and III D.

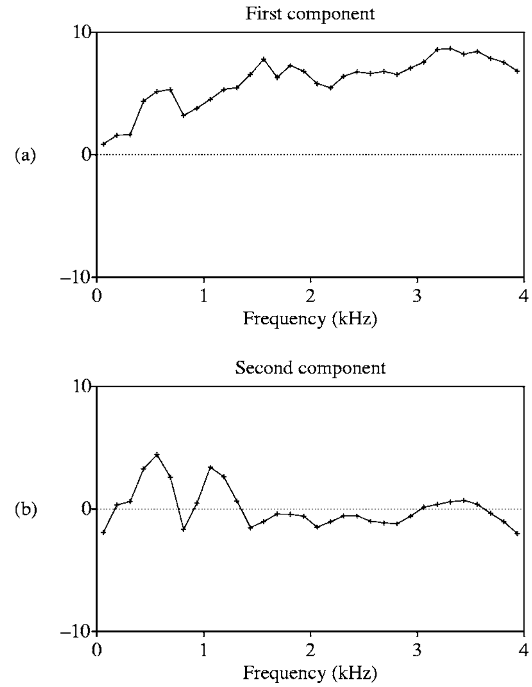

FIG. 7. The first two principal components of the 36 pitch- and speaker-corrected LTASs.

C. Style discrimination by principal component analysis

While the consistency across singers can be informally read off the individual data, there is also a formal technique that can establish this consistency. Intersubject consistency in the relation between the styles is present if, after any singer-specific properties have been factored out, the remaining variation between the 36 recordings is mainly due to the style. Precisely this can be measured by a principal component analysis. The first step is to factor out singer dependency by computing for each singer his average LTAS (i.e., averaged over the three styles) and subtracting this average from each of his three LTASs. The resulting singer-normalized LTASs of the three styles thus add up to zero for each speaker. The 36 singer-normalized LTASs can be regarded as 36 vectors in 32 dimensions (32 is the number of 125-Hz bins in the 0–4 kHz range). The first two principal components of these 36 vectors are shown in Fig. 7.

The first principal component is a slightly rising line, lying entirely above zero. This means that a performance that contains this first component to a high degree is one that combines an overall high intensity level with relatively strong high-frequency components. Since these two spectral features can both be related to loud singing, the fact that the first principal component has this particular shape shows that most of the variation between the 36 singer-normalized LTASs (in fact, 81.2%) is variation in the loudness of the voice. The second principal component (accounting for 6.3% of the variation) has positive peaks around 550 and 1100 Hz and negative peaks around 850 and 1450 Hz. The two negative peaks correspond to the regions in Fig. 5 where the tarankanje curve is higher then the ojkanje curve, and the positive peaks correspond to the regions where the tarankanje curve lies deepest below the two others. The second component, then, turns out to measure the degree to which a performance contains the tarankanje-specific coloring of the spectrum below 2 kHz.

The interesting thing now is to see where the 36 singer-normalized LTASs end up in the two-dimensional space spanned by these two principal components. The result is in Fig. 8. Each of the 12 singers occurs three times in this figure. The three marks labelled "10," for instance, represent the tarankanje, klapa, and ojkanje performances of singer 10; as a result of the singer correction, the horizontal and vertical averages of these three points are zero.

The klapa performances tend to have negative values for the first component, which was expected because Fig. 5 shows that the average klapa performance is less loud [i.e., less like Fig. 7(a)] than the average tarankanje or ojkanje performance. The tarankanje performances tend to have

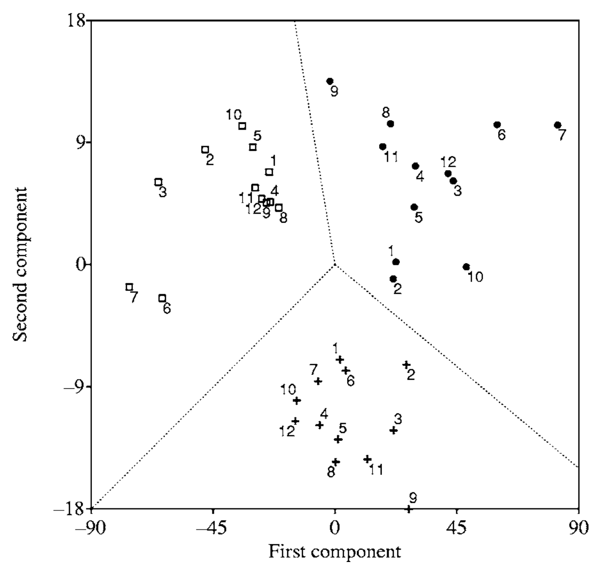

FIG. 8. Principal component analysis of the 36 performances. Pluses =tarankanje, rectangles=klapa, dots=ojkanje, 1, ..., 12=the 12 singers.

negative values for the second component, which was expected because the negative peaks in Fig. 7(b) correspond to positive peaks in Fig. 5 for tarankanje.

The main finding of Fig. 8 is that the styles form clusters without overlap: although the principal component analysis is not a clustering algorithm and does not know what style was associated with what LTAS curve, it turns out to be able to draw a perfect linear division among the 36 performances. It must be noted that, without singer normalization, this separability would not have occurred: if the 36 recordings are regarded as stemming from 36 different singers, the three clouds in Fig. 8 will come to overlap considerably. Thus, the separability of the three styles in the present PCA experiment is a direct result of the experimenters' decision to investigate the three styles with the same 12 singers.

D. Style discrimination by spectral slope and speaker's formant

Whereas the principal component analysis was allowed to take into account the overall intensity level of each performance and to define its own dimensions, the 36 performances can also be plotted in a predefined space of two dimensions that do not refer to acoustic intensity level. It seems reasonable to choose one dimension that reflects a characteristic of the voice source and one dimension that reflects a characteristic of supralaryngeal articulation (i.e., of the shape of the vocal tract).

For the voice source dimension it was decided to measure the *global spectral slope*. Spectral slope (or *spectral tilt*) measures are related to several characteristics of glottal performance, such as vocal intensity (Glave and Rietveld, 1975; Gauffin and Sundberg, 1989; Kiukaanniemi *et al.*, 1982; Sundberg, 2001, pp. 177–178) and hypo- and hyperfunctionality (Löfqvist and Mandersson, 1987). The global slope measure was defined as the difference of the average sound pressure level below 1.0 kHz and the average level between 1.0 and 4.0 kHz. The simple method of determining the spectral slope by using a pivot of 1 kHz is due to Frøkjær-Jensen and Prytz (1976), who directly computed the difference in energy above and below. However, a measure based on perceptual loudness is likely to reflect the psychoacoustic spectral slope better than an energy measure would (Zwicker and Feldtkeller, 1967) and is therefore more likely to reflect style discrimination by humans. The sound pressure levels, then, are computed in dB but mediated by sone units. The global spectral slope is then

$$10 \log_2 \frac{1}{3.0\text{kHz}} \int_{1.0\text{kHz}}^{4.0\text{kHz}} 2^{\mathrm{PSD}(f)/10}\, df - 10 \log_2 \frac{1}{1.0\text{kHz}} \int_{0}^{1.0\text{kHz}} 2^{\mathrm{PSD}(f)/10}\, df, \tag{1}$$

where PSD(f) is the power spectral density (in dB) at frequency f, as estimated from the height of the corresponding bin of the pitch-corrected LTAS.

For the articulatory dimension it was decided to measure the strength of the "speaker's formant," i.e., the strength of the spectral region between 3.0 and 3.8 kHz that "good" (cultivated, trained) speakers have been found to enhance (see Sec. III B). The speaker's formant is thought to have an articulatory correlate (Nolan, 1983, p. 151), perhaps a clustering of F4 and F5 (Leino, 1994), or just F4 alone, supported by a long closure phase (Cleveland *et al.*, 2001). One could define a global speaker's formant strength by subtracting the height of the 3.5-kHz peak from the height of a peak below 1 kHz [e.g., Nawka *et al.*, 1997; cf. the similar "resonance balance" method for the singer's formant by Schutte and Miller (1984)], but as Leino (1994, p. 209) points out, such a definition would confound this articulatory measure with the global spectral slope, which is related to the voice source (in the present case, such a subtraction would lead to low values for both tarankanje and ojkanje, but by different causes). In order for the measure of speaker's formant strength to be as independent of the spectral slope as possible, it has to be taken relative to the level of the surrounding spectral region, i.e., the regions between 2.2 and 3.0 kHz and between 3.8 and 4.6 kHz. The formula for this *local speaker's formant strength* is analogous to the formula for the global spectral slope:

$$10 \log_2 \frac{1}{0.8\text{kHz}} \int_{3.0\text{kHz}}^{3.8\text{kHz}} 2^{\mathrm{PSD}(f)/10}\, df - 10 \log_2 \frac{1}{1.6\text{kHz}} \left(\int_{2.2\text{kHz}}^{3.0\text{kHz}} 2^{\mathrm{PSD}(f)/10}\, df + \int_{3.8\text{kHz}}^{4.6\text{kHz}} 2^{\mathrm{PSD}(f)/10}\, df \right). \tag{2}$$

After the computation of the global spectral slope and the local speaker's formant strength, the values of these two numbers are normalized for each singer, analogously to the singer normalization performed before in the principal component analysis. Figure 9 plots all 36 performances in the space spanned by the two singer-normalized dimensions just defined.

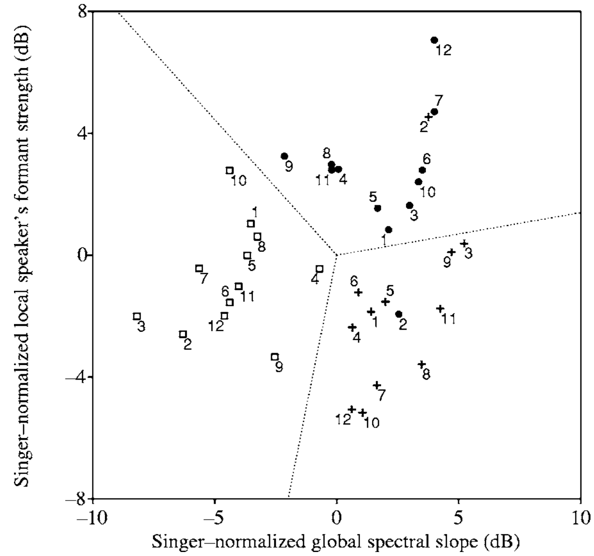

FIG. 9. The locations of the 36 performances in a space of two predefined spectral shape characteristics. Pluses=tarankanje, rectangles=klapa, dots=ojkanje, 1, … , 12=the 12 singers.

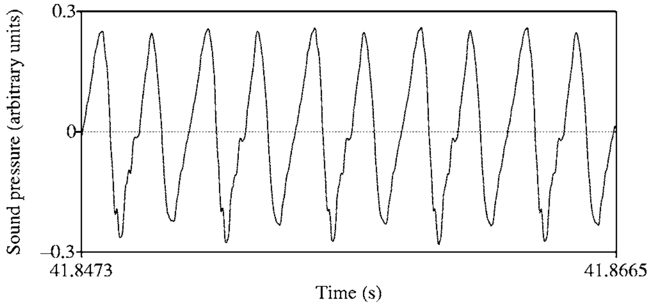

FIG. 10. Five periods of a vowel in ojkanje (singer 9).

Figure 9 shows that several differences between the styles are consistent across singers: the fact that every singer's klapa performance turns up at the left of both his tarankanje and his ojkanje performance means that every singer's spectral slope for klapa is greater than his spectral slope for either other style. The fact that all tarankanje performances show up at the right of the zero vertical shows that every singer's tarankanje performance has a flatter slope than his average performance has. The fact that 11 ojkanje performances show up above the zero horizontal shows that for every singer except singer 2 the speaker's formant for ojkanje is stronger than his average speaker's formant. Likewise, the speaker's formant seems to be absent from the tarankanje performances of all singers except singer 2.

The overall image of Fig. 9 is that a linear separation between the three styles is very well possible, i.e., given three performances of the same singer the singer-normalized measurements of spectral slope and speaker's formant are capable of predicting which performance belongs to which style (except for singer 2). This separation is shown by the dividing lines in Fig. 9. The two dimensions are probably independently controllable by the singer: the spectral slope is controlled by the voice source, whereas the speaker's formant is controlled by supralaryngeal articulation.

As in Fig. 8, the separability of the three styles in Fig. 9 is a direct consequence of investigating a single group of 12 singers on different styles, because without singer normalization the three clouds would overlap to a large degree.

E. Style discrimination by the strength of the second harmonic

There exist spectral characteristics that cannot be read off directly from the LTAS curves. One of them relates to the rather narrow peak around 600 Hz in the ojkanje style. Since the fundamental frequency of the ojkanje song investigated here is around 300 Hz, it may be worthwhile to investigate the hypothesis that ojkanje singers aim at maximizing the amplitude of the second harmonic. Figure 10 shows five periods of the ojkanje performance by singer 9. The F0 is 262 Hz, but the waveform resembles a 524 sine wave, so that it is likely that most of the spectral energy is in the second harmonic. Preliminary investigations into authentic ojkanje singers suggest that the singers aim at lending prominence not to a specific frequency, but to the second harmonic.

To investigate how well LADO singers succeed in bringing forward the second harmonic (H2), it was computed how high the energy of the second harmonic rose above the average energy of the first harmonic (i.e., F0) and the third. The measurement method can be exemplified with Fig. 3: to compute the total energy associated with H2, for instance, one just adds the energies associated with the second harmonics of the three periods in the figure, yielding $1.7+4.0+1.7=7.4\ \text{Pa}^2\ \mu\text{s}$. For the 12 LADO singers, then, the local H2 strength $[\text{i.e., H2}-(\text{H1}+\text{H3})/2]$ turned out to lie between 10.2 and 19.4 dB. These high values positively identify ojkanje, even without singer normalization, since the two other styles have lower H2 strengths (tarankanje: between −4.2 and +4.5 dB; klapa: between 0.1 and 10.1 dB). With singer normalization, the discrimination improves: every singer's H2 strength for ojkanje is much higher than his H2 strength for klapa; the difference lies between 5.3 and 14.4 dB, with a median difference of 9.2 dB.

Another way of establishing the strength of the second harmonic starts with realizing that a strong second harmonic of a male singer with an F0 of 220 to 320 Hz is likely to be the result of a tuning of the first formant. Figure 11 shows the distributions of the first four formants for each of the three styles, pooled over all 12 singers (the formants were determined with the PRAAT program, requiring five formants be-

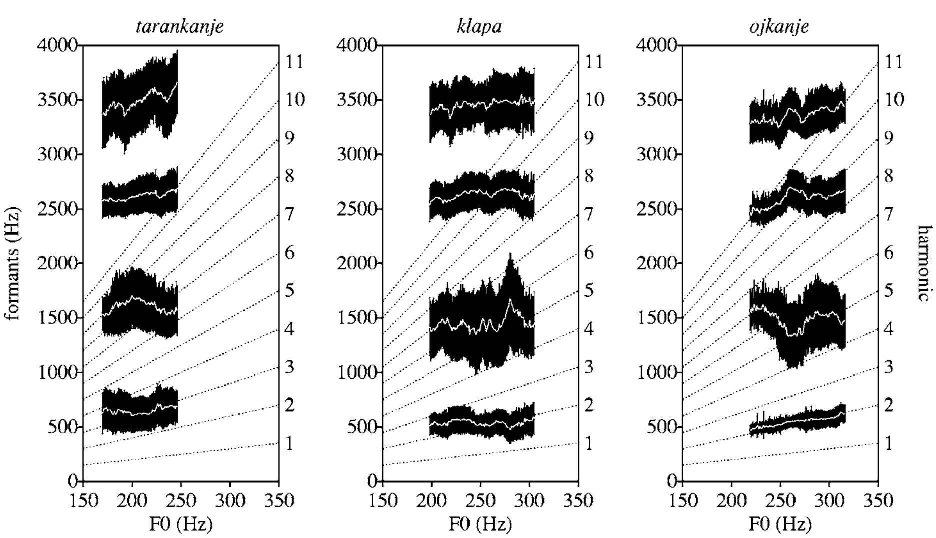

FIG. 11. Pooled formant distributions as functions of F0.

low 5000 Hz, with an analysis window length of 25 ms). The figure contains information only on those moments in the performance where F0 is between the 5th and 95th percentile of the pooled F0 distribution of the style and all four formants are simultaneously between the 5th and 95th percentiles of their pooled distributions. The white curves, then, indicate the average formants as functions of F0; these curves were computed by averaging approximately 30 000 formant values (per style per formant) in 1-Hz-wide bins. The black areas above and below the curves (they are actually vertical lines, one line per Hz) indicate the standard deviations of the formants as functions of F0. The dotted lines in the figure indicate the first to eleventh harmonics as functions of F0.

The figure confirms the hypothesis. The F1 curve for ojkanje closely follows the line of the second harmonic. This cannot be a side effect of the analysis method, because the other styles show hardly any dependence of F1 on F0. The standard deviation of F1 for ojkanje is seen to be very small. This, together with the coincidence of F1 with H2, proves that these singers tune their F1 to 2F0 (or the reverse, i.e., 2F0 to vowel height). This ojkanje song can therefore be seen as a kind of overtone singing.

A third thing that would contribute to the sharpness of the peak is the bandwidth of the first formant. The LADO singers did not utilize this possibility in this song: the median F1 bandwidth was 179 Hz, which was not so different from that of the other two styles (tarankanje 183 Hz, klapa 130 Hz).[3]

IV. DISCUSSION

In this section we want to point out that the differences between the styles, as observed in Sec. III, have perceptual goals: the acoustic differences do not just reflect single articulatory differences. Rather, for each style there are a multitude of articulatory features that synergetically act to achieve a specific psychoacoustic effect. The following subsections explain this in detail for each style.

A. Tarankanje: Whiteness

The extraordinary flatness of the spectrum in the tarankanje style seems to be implemented synergetically by settings of the voice source and by supralaryngeal settings. The energy difference between the 0–2 kHz and 2–4 kHz regions is made as small as possible by producing a pressed voice. The energy difference between the first and second spectral peaks (at 0.7 and 1.5 kHz) is made as small as possible by maintaining a nasal vocal quality reminiscent of an open nasal front vowel (Stevens, 1998, pp. 310–312). The song under investigation here seems to aim at enhancing this effect by in fact using the nasalized low front vowel [æ̃] as its most frequent vowel, although the spectrum turns out to stay very similar if all [næ̃]-like syllables are removed (which shows that the phonemic makeup of the song is largely irrelevant for the spectrum). Indeed, tarankanje singing has been described as "strong and partially through the nose, so the tone colour is nasal" (Bonifačić, 2001, p. 75).

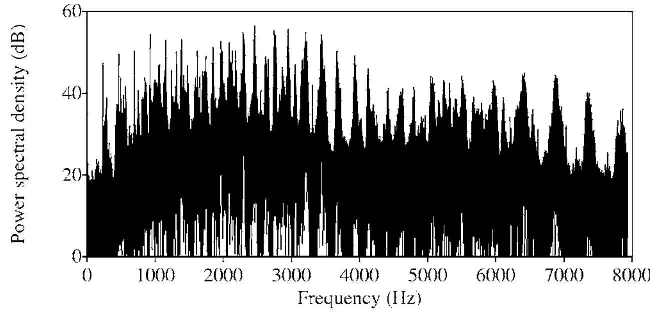

FIG. 12. The flat spectrum of two sopile instruments.

Together, these production tricks implement a spectrum that is as flat as possible, i.e., what Bergan *et al.* (2004) call a relatively "white" sound. This flat spectrum renders the sound similar to that of the wind instruments that are part of folk tradition, namely *sopile* (shawms) or *roženice*, which have been described as producing a piercing and nasal tone quality (Bonifačić, 2001). An example of a spectrum of a large and a small sopile playing together is given in Fig. 12. According to Bonifačić (1996), imitating these instruments is the very goal of tarankanje singing: if the instrument is not available, tarankanje singing can still take over its role as an accompaniment to dance.

B. Ojkanje: First-formant tuning and a speaker's (shouter's) formant

The ojkanje style is characterized by great loudness, which is reflected as a generally higher level of intensity, by the tuning of the second harmonic to F1, and by a broad spectral peak for high frequencies, reminiscent of a speaker's formant.

The tuning of the first formant as the first overtone of the fundamental frequency in ojkanje is by far the strongest peak in the spectra of all styles together. Ćaleta (1999) describes ojkanje as shoutinglike vocal production in a high register. The label of "shouting" is common and refers both to the listener's impression of this style and to the original perceptual goal of this style, which has been described as the need to communicate over a great physical distance in sparsely settled mountainous areas (Dobronić, 1915; Bezić, 1968; Marošević, 2004). The frequency range from 600 to 800 Hz is indeed optimal for achieving the special carrying power of ojkanje, because for lower frequencies the human ear becomes less sensitive (Fletcher and Munson, 1933) and for higher frequencies air absorption may hamper transmission (Evans *et al.* 1971). A very similar sharp peak has been found before in *kölning*, which is a Swedish singing style (the name is reported to derive from the verb *kalla* "call") that maids used for calling cattle and to communicate to each other, in mountainous areas (Johnson *et al.* 1985; Sundberg, 1995, p. 2). In kölning women tune their F1 to their F0 (Johnson *et al.*, 1985, p. 198). With a typical difference of an octave between women's and men's singing, it seems only natural that men would tune their F1 to 2F0, women to F0. Generalizing from the two cases of ojkanje and kölning, and following the tradition in voice research to call spectral peaks "formants," one could call this tuned F1 peak the "low shouter's formant."

The second, broader, peak lies between 3.0 and 3.8 kHz and is probably identical to the speaker's formant. Besides in "good" speakers, such a peak has been attested before in the speech and song of country singers (Cleveland *et al.*, 2001). It can be interpreted as the combined result of a vocal action and an articulatory posture. The vocal action is the observed high vocal intensity, a thing that tends to affect the level of high-frequency peaks more than the level of lower frequencies; this correlation has been shown in singing (Bartholomew, 1934; Hollien, 1983; Bloothooft and Plomp, 1986; Sundberg, 2001) as well as in speech (Nawka *et al.*, 1997; Nordenberg and Sundberg, 2003) and is present in the comparison of ojkanje with the speechlike klapa style as well (Fig. 5). The articulatory posture must be the posture that leads to a strong or close F4 and F5, hence a strong local peak (see Sec. III D). In ojkanje, the loudness in general and the high-frequency strength in particular are probably implemented synergetically by various respiratory, phonatory, and articulatory settings, namely a high subglottal pressure, a long closed phase of the vocal folds, a raised larynx, and a wide opening of the jaw [cf. Johnson *et al.* (1985), who found that kölning comes with a high subglottal pressure, a raised larynx, and a large jaw opening].[4] If future research attests the high-frequency peak in similar calling styles around the world, the term *shouter's ring* or *shouter's formant* might become appropriate.

The aforementioned proposal that the perceptual goal of ojkanje is the need to communicate over a large distance in mountainous areas (Dobronić, 1915; Bezić, 1968; Marošević, 2004) has now been confirmed by the observation of a sharply tuned resonance ($F1=2F0_{male}$) in the farthest-carrying frequency range. The proposal is also corroborated by Marošević's (2004) observation that similar vocal styles are found in mountainous regions elsewhere (Albania, Bulgaria, Greece, Turkey). And the proposal is especially strongly supported by Johnson *et al.*'s (1985) proposal that the same perceptual goal applies to a calling song (kölning) with a very similar sharply tuned resonance ($F1=F0_{female}$).

C. Klapa: Harmony

The klapa style has been described as sounding pleasant and beautiful, even cultivated (Bezić, 1979). Its perceptual goal has been described as "to achieve the best possible blend of chords" (Ćaleta, 1997, p. 135), probably coinciding with that of other love songs all over the world. Klapa singers use two tricks to achieve harmony (in the sense of overtone matching): first, they sing in a harmonically based (Western-like) musical scale; second, they use multiple larynges, i.e., klapa is always performed in a choir consisting of five to eight singers.

D. The absence of a perceptual goal: Singer's formant

The current study showed that none of the three singing styles relies on a singer's formant.[5] This finding can be compared to the findings by Burns (1986) for American country-and-western folk singing, by Ross (1992) for Estonian folk singing, and by Cleveland *et al.* (2001) for country singing, none of whom found any evidence for a singer's formant in folk singing styles. This absence is understandable in the light of Sundberg's (1972) proposal that the singer's formant originally developed in order for the singer not to be masked by the spectrum of a symphonic orchestra.[6] According to Sundberg, singers would have no need for a singer's formant if they are accompanied by instruments with lower sound levels, such as a lute. This can explain why a singer's formant never developed in ojkanje, which is always performed *a cappella*, nor in klapa, which is usually sung *a cappella* and only rarely with the accompaniment of light string instruments such as mandolins. The tarankanje style is typically performed with two complementary voices [so-called "big" and "small" (Karabaić, 1956)], which may be two human voices or one human voice and a sopile instrument. In the latter case, the human voice is not meant to overcome the spectrum of the sopile. On the contrary, the perceptual goal of the human voice seems to be "onomatopoeic imitation" (Bonifačić, 1996) to mimic the spectral whiteness of the sopile instrument.

In the case of klapa, the absence of a singer's formant can partly be explained by the fact that it is performed in a choir: since multiple voices contribute to the loudness, the singer's formant is superfluous. Rossing *et al.* (1986) indeed showed that professional singers use a singer's formant consistingly mainly when performing solo, and use it less when performing in a choir.

V. CONCLUSION

The goal of the current study, namely to find spectral differences between three styles of Croatian folk singing, was successfully reached because it turned out that spectral differences between the styles could be established for each of the 12 singers in the same way (except for singer 2's tarankanje and ojkanje in Fig. 9).

The spectral characteristics of the three styles were ultimately explained in detail by their original or current perceptual goals. The absence of the perceptual goal of overcoming an orchestra explained the lack of a singer's formant in all three styles, either negatively (because of the lack of an orchestra) or positively (because of the very desire to blend with the instrumental accompaniment). The presence of the perceptual goal of being heard across large distances in mountainous areas explained the F1 tuning in ojkanje. Finally, the presence of the perceptual goal of imitating the sound of the sopile explained the flat ("white") spectrum in tarankanje.

For all three styles the present paper investigated the same 12 singers. The objective of this was to be able to perform singer normalization, and indeed this normalization turned out to be a requirement for the near-perfect separability of the three styles observed in Figs. 8 and 9 for the professional singers of the LADO ensemble. Doing the tests of the present paper with performers randomly sampled from three hypothetical populations of singers with LADO-like interpretations of the styles is therefore expected not to yield such a good separability. It is feasible, however, to redo the tests with authentic singers with specialized skills in a single style, who may compensate for the lack of normalizability by having more outspoken style features. Informal observations indeed suggest that authentic ojkanje performers shout even higher and louder while at the same time expending less effort, authentic tarankanje performers match the sound of the sopile even more closely, and authentic klapa singers sing even more harmonically and smoothly.

ACKNOWLEDGMENTS

The research was supported by a Huygens scholarship 2002/03 awarded by Nuffic. The authors are thankful to LADO Folk Dance Ensemble of Croatia for the willingness and patience to participate in the study despite a very tight schedule at the time of data collection, and also to the staff of the Department of Electroacoustics of the Faculty of Electrical Engineering and Computing of the University of Zagreb (Croatia) for their assistance throughout the above-mentioned phase of the investigation.

[1]See EPAPS Document No. E-JASMAN-119–059603 for the sound files of the 36 performances. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

[2]The unit is Pascal-squared times microseconds. To obtain an average sound intensity in W/m^2, one divides by the duration of the period and by the acoustic impedance ρc, where ρ is the density of air and c is the speed of sound.

[3]A performance of the song *Oj djevojko* by authentic singers turned out to have an extremely low F1 bandwidth of only 10 Hz, leading to an H1-H2 difference of −30 dB in the acoustic signal. To what extent the difference between 179 and 10 Hz is due to a difference between the songs, the regions, and/or the performers must be left for future research.

[4]A direct correlation between subglottal pressure and duration of the closure phase has been reported for untrained speakers (Sundberg *et al.*, 2005) and for professional singers (Sundberg *et al.*, 1999a). However, it is possible that the correlation appears only if the goal of the performer is to vary loudness, as it is in the case of both ojkanje and the participants in Sundberg *et al.* (1999a) and Sundberg *et al.* (2005). In singing, such correlations tend not to be automatic, as the very goal of many singing styles is to be different from speech (and as a contrastive feature in language, creak can be controlled separately from loudness).

[5]Singer 10 appears to sing with a singer's formant in every style. This was not achieved by training: this singer was a relative beginner, having had only 4 years of experience in professional folk singing. This singer is simply one of the rare individuals who have a spontaneous singer's formant.

[6]The presence of this original perceptual goal does not preclude the possibility that the "ring" or "brilliance" of the singer's formant has nowadays become a desirable perceptual goal of Western classical singing in itself, and is utilized as well outside the original conditions.

Bartholomew, W. T. (**1934**). "A physical definition of 'good voice-quality' in the male voice," J. Acoust. Soc. Am. **6**, 25–33.

Bergan, C. C., Titze, I. R., and Story, B. (**2004**). "The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice," J. Voice **18**, 305–317.

Bezić, J. (**1968**). "Muzički folklor Sinjske krajine (The musical folklore of Sinjska Krajina)," Narodna Umjetnost—Croatian Journal of Ethnology and Folklore Research, Book **5–6**, 175–274.

Bezić, J. (**1979**). "Dalmatinske klapske pjesme kroz deset godina omiškog festivala (Dalmatian klapa songs throughout the ten years of the Omiš festival)," in *Zbornik dalmatinskih klapskih pjesama, izvedenih na festivalima u Omišu od 1967 do 1976 (Overview of Dalmatian klapa songs performed at the Omiš festival from 1967 to 1976)*, Omiš: Festival dalmatinskih klapa Omiš, pp. 16–23.

Bloothooft, G., and Plomp, R. (**1986**). "The sound level of the singer's formant in professional singing," J. Acoust. Soc. Am. **79**, 2028–2033.

Bloothooft, G., Bringmann, E., Van Cappellen, M., Van Luipen, J. B., and Thomassen, K. P. (**1992**). "Acoustics and perception of overtone singing," J. Acoust. Soc. Am. **92**, 1827–1836.

Boersma, P., and Weenink, D. (**2005**). "PRAAT: doing phonetics by computer (Version 4.3.31)" (Computer program), retrieved 27 November 2005, from http://www.praat.org.

Bonifačić, R. (**1996**). "*Tarankanje*: a disappearing music tradition," Narodna Umjetnost—Croatian Journal of Ethnology and Folklore Research **33**(1), 149–170.

Bonifačić, R. (**2001**). "O problematici takozvane 'istarske ljestvice' (On the topic of the so-called 'Istrian scale')," Narodna Umjetnost—Croatian Journal of Ethnology and Folklore Research **38**(2), 73–95.

Borch, D. Z., and Sundberg, J. (**2002**). "Spectral distribution of solo voice and accompaniment in pop music," Logoped. Phoniatr. Vocol. **27**, 37–41.

Burns, P. (**1986**). "Acoustical analysis of the underlying voice differences between two groups of professional singers: opera and country and western," Laryngoscope **96**, 549–554.

Ćaleta, J. (**1997**). "Klapa singing, a traditional folk phenomenon of Dalmatia," Narodna Umjetnost—Croatian Journal of Ethnology and Folklore Research **34**(1), 127–145.

Ćaleta, J. (**1999**). "The ethnomusicological approach to the concept of the Mediterranean in music of Croatia," Narodna Umjetnost—Croatian Journal of Ethnology and Folklore Research **36**(1), 183–195.

Cleveland, T. F., Sundberg, J., and Stone, R. E. (**2001**). "Long-term-average spectrum characteristics of country singers during speaking and singing," J. Voice **15**, 54–60.

Dobronić, A. (**1915**). "'Ojkanje,' prilog za proučavanje geneze naše pučke popijevke ('Ojkanje,' a contribution to the study of the origins of our folk songs)," in Zbornik za narodni život i običaje južnih Slavena (Proceedings on folk life and traditions of the Southern Slavs), Knjiga XX, svezak 1 (JAZU, Zagreb), pp. 1–25.

Doskov, D., Ivanov, T., and Boyanov, B. (**1995**). "Comparative analysis of singer's high formant in different type of singing voices," Folia Phoniatr Logop **47**, 291–295.

Estill, J., Fujimura, O., Erickson, D., Zhang, T., and Beechler, K. (**1994**). "Vocal tract contributions to voice qualities," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC 93)*, edited by A. Friberg, J. Iwarsson, E. Jansson, and J. Sundberg (Royal Swedish Academy of Music, Stockholm), pp. 161–165.

Evans, L. B., Bass, H. E., and Sutherland, L. C. (**1971**). "Atmospheric absorption of sound: theoretical predictions," J. Acoust. Soc. Am. **51**, 1565–1575.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement, and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Frøkjær-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Brüel & Kjær Tech. Rev. **3**, 3–17.

Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.

Glave, R. D., and Rietveld, A. C. M. (**1975**). "Is the effort dependence of speech loudness explicable on the basis of acoustical cues," J. Acoust. Soc. Am. **58**, 875–879.

Hollien, H. (**1983**). "The puzzle of the singer's formant," in *Vocal Fold Physiology*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 368–378.

Johnson, A., Sundberg, J., and Wilbrand, H. (**1985**). "'Kölning.' Study of phonation and articulation in a type of Swedish herding song," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC 83)*, edited by A. Askenfelt, S. Felicetti, E. Jansson, and J. Sundberg (Royal Swedish Academy of Music, Stockholm), pp. 187–202.

Karabaić, N. (**1956**). *Muzički folklor Hrvatskog primorja i Istre (Musical Folklore of the Croatian Littoral and Istria)* (Novi List, Rijeka).

Kiukaanniemi, H., Siponen, P., and Mattila, P. (**1982**). "Individual differences in the long-term speech spectrum," Folia Phoniatr. **34**, 21–28.

Klingholz, F. (**1993**). "Overtone singing: productive mechanisms and acoustic data," J. Voice **7**, 118–122.

Kuwabara, H., and Ohgushi, K. (**1984**). "Acoustic characteristics of professional male announcers' speech sounds," Acustica **55**, 233–240.

Leino, T. (**1994**). "Long-term average spectrum study on speaking voice quality in male actors," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC 93)*, edited by A. Friberg, J. Iwarsson, E. Jansson, and

J. Sundberg (Royal Swedish Academy of Music, Stockholm), pp. 206–210.
Lindestad, P.-Å., Södersten, M., Merker, B., and Granqvist, S. (**2001**). "Voice source characteristics in Mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering," J. Voice **15**, 78–85.
Löfqvist, A., and Mandersson, B. (**1987**). "Long-term average spectrum of speech and voice analysis," Folia Phoniatr. **39**, 221–229.
Marošević, G. (**1994**). "Ojkanje u izvandinarskim područjima Hrvatske (Ojkanje outside of the Dinaric region in Croatia)," Etnološka Tribina **17**, 91–102.
Marošević, G. (**2004**). "Traditional music," on *Croatian Folk Culture* (compact disc), edited by Z. Vitez (Institut za etnologiju i folkloristiku, Zagreb), pp. 408–419.
Nawka, T., Anders, L. C., Cebulla, M., and Zurakowski, D. (**1997**). "The speaker's formant in male voices," J. Voice **11**, 422–428.
Nolan, F. (**1983**). *The Phonetic Bases of Speaker Recognition* (Cambridge U.P., Cambridge, MA).
Nordenberg, M., and Sundberg, J. (**2003**). "Effect on LTAS of vocal loudness variation," in *Speech, Music and Hearing*, Q. Prog. Stat. Report 45 (Royal Institute of Technology, Stockholm), pp. 93–100.
Rapanić, Ž. (**1979**). "Deset godina Festivala dalmatinskih klapa (Ten years of the Festival of Dalmatian klapa)," in *Zbornik dalmatinskih klapskih pjesama, izvedenih na festivalima u Omišu od 1967 do 1976 (Overview of Dalmatian klapa songs performed at the Omiš festival from 1967 to 1976)*. Omiš: Festival dalmatinskih klapa Omiš, pp. 9–13.
Ross, J. (**1992**). "Formant frequencies in Estonian folk singing," J. Acoust. Soc. Am. **91**, 3532–3539.
Rossing, T. D., Sundberg, J., and Ternström, S. (**1986**). "Acoustic comparison of voice use in solo and choir singing," J. Acoust. Soc. Am. **79**, 1975–1981.
Schutte, H. K., and Miller, D. G. (**1993**). "Belting and pop, nonclassical approaches to the female middle voice: Some preliminary considerations," J. Voice **7**, 142–150.
Schutte, H. K., and Miller, R. (**1984**). "Resonance balance in register categories of the singing voice: a spectral analysis study," Folia Phoniatr. **36**, 289–295.
Sluijter, A. (**1995**). "Phonetic correlates of stress and accent," Ph.D. thesis, Leiden University (Holland Academic Graphics, The Hague).
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Stone, R. E., Jr., Cleveland, T. F., and Sundberg, J. (**1999**). "Formant frequencies in country singers' speech and singing," J. Voice **13**, 161–167.
Stone, R. E., Jr., Cleveland, T. F., Sundberg, J., and Prokop, J. (**2003**). "Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer," J. Voice **17**, 283–297.
Sundberg, J. (**1972**). "A perceptual function of the 'singing formant'," in Speech Transmission Laboratory Quarterly Progress and Status Report **2–3**, 61–63.
Sundberg, J. (**1973**). "The source spectrum in professional singing," Folia Phoniatr. **25**, 71–90.
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant'," J. Acoust. Soc. Am. **55**, 838–844.
Sundberg, J. (**1995**). "Swedish voices in music," in Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Vol. **1**, pp. 2–8.
Sundberg, J. (**2001**). "Level and center frequency of the singer's formant," J. Voice **15**, 176–186.
Sundberg, J., Andersson, M., and Hultqvist, C. (**1999a**). "Effects of subglottal pressure variation on professional baritone singers' voice sources," J. Acoust. Soc. Am. **105**, 1965–1971.
Sundberg, J., Fahlstedt, E., and Morell, A. (**2005**). "Effects on the glottal voice source of vocal loudness variation in untrained female and male voices," J. Acoust. Soc. Am. **117**, 879–885.
Sundberg, J., Cleveland, T. F., Stone, R. E., Jr., and Iwarsson, J. (**1999b**). "Voice source characteristics in six premier country singers," J. Voice **13**, 168–183.
Thalén, M., and Sundberg, J. (**2001**). "Describing different styles of singing: a comparison of a female singer's voice source in 'classical,' 'pop,' 'jazz' and 'blues,'," Logoped. Phoniatr. Vocol. **26**, 82–93.
Thunberg, G. C. (**2003**). "Spectral balance utilization in different speaking styles-a preliminary investigation," in Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 2149–2152.
Van Tongeren, M. C. (**2002**). *Overtone Singing—Physics and Metaphysics of Harmonics in East and West* (Fusica, Amsterdam).
Zwicker, E., and Feldtkeller, R. (**1967**). *Das Ohr als Nachrichtenempfänger* (Hirzel, Stuttgart).