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

Abbreviated title: Croatian folk singing

June 8, 2005

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

Abstract

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 analysed with a pitch-corrected LTAS (long-term average spectrum) method. After factoring out each singer's average, the 36 LTAS contours were reduced to a 2-dimensional representation in two ways: (1) a principal-component analysis, (2) a graphical plot of spectral slope versus speaker 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 shows a high spectral peak around 3.5 kHz, which is implemented by high vocal effort and the articulation of a *shouter's formant* and can be explained by the desire to be heard across large distances on mountain slopes. The *tarankanje* style has a very flat spectrum, which is implemented by vocal pressedness and nasality and can be explained by the desire to match the timbral characteristics of the *sopile* folk instrument.

PACS numbers: 43.75.Rs, 43.70.Gr, 43.72.Ar

I. INTRODUCTION

Although there exist many more non-classical 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 non-classical 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 & western (Burns, 1986; Stone *et al.*, 1999; Sundberg *et al.*, 1999; 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 aim of the present investigation is to add three more styles to the breadth of investigated non-classical 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, serenade-like 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 (which is called *dozivački* among the singers who participated in this study), 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 non-music (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 analysing the intra-subject 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 practising 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 formally 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).

B. Data collection

The recordings were performed in an anechoic chamber of the Department of Electroacoustics of the *Faculty of EE 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 & Kjær type 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 metres 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.

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 loudness 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. The average durations of the songs turned out to be 57.16 seconds for *klapa*, 50.75 seconds for *ojkanje*, and 43.50 seconds for *tarankanje*.

C. Acoustic analysis: pitch-corrected LTAS

In order to obtain information on the average production (phonatory and articulatory) and average spectral properties (e.g. the presence or absence of a singer's formant) of the three styles, 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, 2004). 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 Figure 1, which shows the simple LTAS for *tarankanje*, pooled over all 12 singers.

[Figure 1 around here]

The figure clearly shows a peaked spectrum, especially in the second bin (125–250 Hz; the peak is at the centre, 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 LTASes 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.

[Figure 2 around here]

To annihilate the influence of F0, a *pitch-corrected LTAS* method was designed, and this was used for all analyses in this paper. The procedure runs as follows. For the voiced parts of the recording, each pitch period (as detected by PRAAT's cycle-to-cycle waveform matching) is excised and converted to a line spectrum by Fourier transformation (data windowing is not needed, and the phase of the glottal pulse within the excised period does not influence the result). For instance, a period with a duration of 4.9 ms is transformed to a spectrum with a frequency spacing of 1/4.9 ms = 204 Hz. This spectrum therefore contains information on energies at the harmonics, i.e. at 204, 408, 612 Hz and so on. Each of these energies is put into the appropriate 125-Hz-wide bin of an LTAS (in this example the second, fourth and fifth bin). When the whole sound has been processed and all the energies have been added into their appropriate bins, the total energy in each bin is divided by the number of energies that had been put into that bin. This division is what performs the actual pitch correction: the result is an LTAS in which each bin represents the average energy of the harmonics that entered it. Figure 3 shows the resulting LTAS curve for *tarankanje*. The curve is much smoother than the non-corrected LTAS of Figure 1. Comparable improvements apply to the other two styles.

[Figure 3 around here]

III. RESULTS

A. Pooled data

Figure 4 shows the pitch-corrected LTASes for each of the three styles, pooled over all 12 singers. Several inter-style differences can be read off this figure. The most noticeable differences are related to sound levels, spectral slopes, and locations of the spectral peaks.

[Figure 4 around here]

As far as the sound levels are concerned, Figure 4 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 intensity 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 intensity 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, corresponds to the spectrum of the nasalized low front vowel [æ̃] (Stevens, 1998, p. 311), which is indeed the most frequent vowel in the *tarankanje* performances, 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. 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.

Although there is no singer's formant in any of the three styles, 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 phenomenon that is usually associated with the speech of professional speakers such as radio announcers (Kuwabara & Ohgushi, 1984) and theatre actors (Leino, 1994; Thunberg, 2003). Reportedly, this speaker's formant is weaker than the 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). 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). 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. The Discussion section proposes an alternative term.

B. Individual data

While the average inter-style differences can be read off Figure 4, 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 LTASes, and formally by performing a principal-component analysis or a data reduction into pre-established properties of the individual LTASes.

Figure 5 shows the individual pitch-corrected LTASes for each of the 12 singers and each of the three styles. Several properties that were noted for the pooled data of Figure 4 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, 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 singer 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).

[Figure 5 around here]

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

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. Inter-subject 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 LTASes. The resulting singer-normalized LTASes of the three styles thus add up to zero for each speaker. The 36 singer-normalized LTASes 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 Figure 6.

[Figure 6 around here]

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 LTASes (in fact, 81.2 percent) is variation in the loudness of the voice. The second principal component (accounting for 6.3 percent 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 Figure 4 whereas 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 colouring of the spectrum below 2 kHz.

The interesting thing now is to see where the 36 singer-normalized LTASes end up in the two-dimensional space spanned by these two principal components. The result is in Figure 7. 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.

[Figure 7 around here]

The *klapa* performances tend to have negative values for the first component, which was expected because Figure 4 shows that the average *klapa* performance is less loud (i.e. less like Figure 6a) than the average *tarankanje* or *ojkanje* performance. The *tarankanje* performances tend to have negative values for the second component, which was expected because the negative peaks in Figure 6b correspond to positive peaks in Figure 4 for *tarankanje*.

The main finding of Figure 7 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.

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.

For the voice source dimension it was decided to measure the *global spectral slope*. Spectral slope 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:

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

where *PSD*(*f*) is the power spectral density 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" (see section B). 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), 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 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:

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

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 8 plots all 36 performances in the space spanned by the two singer-normalized dimensions just defined.

Figure 8 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 eleven *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 Figure 8 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 Figure 8. 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.

[Figure 8 around here]

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 9 shows five periods of the ojkanje performance by Singer 9. The F0 is 288 Hz, but the waveform resembles a 576 sine wave, so that it is likely that most of the spectral energy is in the second harmonic. Figure 10 shows five periods from a performance of an authentic *ojkanje* singer (found on the Web) singing the song *Ej djevojko*. The vowel has an F0 of 400 Hz, but most of the energy is in the second harmonic, i.e. at 800 Hz. The fact that the resonance moves up and down with F0 suggests that the singers aim at lending prominence not to a specific frequency, but to the second harmonic. The *ojkanje* style can therefore be seen as a kind of overtone singing.

[Figure 9 around here]

[Figure 10 around here]

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. For the 12 *LADO* singers, this was between 10.3 and 19.5 dB. These high values positively identify *ojkanje*, even without singer normalization, since the two other styles have lower second-harmonic amplitude values (*tarankanje*: between -4.2 and +4.6 dB; *klapa*: between 0.1 and 9.8 dB). With singer normalization, the discrimination improves: a singer's H2 strength for *ojkanje* is between 5.7 and 14.4 dB higher than the same singer's H2 strength for *klapa*, with a median of 9.0 dB.

IV. DISCUSSION

In this section we want to point out that the differences between the styles, as observed in the Results sections, 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 acoustic 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 articulating an open nasal front vowel (Stevens, 1998, pp. 310-312). 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). 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: the shouter's formant

The *ojkanje* style is characterized by great loudness, which is reflected as a generally higher level of intensity as well as by a broad spectral peak for high frequencies, reminiscent of a speaker's formant. The peak is explainable as the result of an interaction between the voice source and the articulatory posture. Ćaleta (1999) describes *ojkanje* as shouting-like vocal production in a high register; the appropriate label for the spectral peak may therefore be *shouter's ring* or *shouter's formant*. The two acoustic effects (intensity and relative peak height) are correlated 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), i.e., an increase in intensity affects the level of high-frequency peaks more than the level of lower frequencies. In *ojkanje*, the acoustic effects 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. All these contribute to perceived loudness. The perceptual goal of this style has indeed been described as the need to communicate over a great physical distance in sparsely settled mountainous areas (Dobronić, 1915; Bezić, 1968; Marošević, 2004), as is corroborated by Marošević's (2004) observation that similar vocal styles are found in mountainous regions elsewhere (Albania, Bulgaria, Greece, Turkey).

Another thing that contributes to the carrying power of *ojkanje* is the tuning of the second formant as the first overtone of the fundamental frequency. Next to the 'high shouter's formant' around 3.5 kHz, this produces a strong 'low shouter's formant' around 600 Hz.

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. This is understandable in the light of Sundberg's (1972) proposal that the perceptual goal of a singer's formant is for the singer not to be masked by the spectrum of a symphonic orchestra. 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 accounts for the absence of a singer's formant in *ojkanje*, which is always sung a cappella, as well as 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 only when performing solo, and hardly use it when performing in a choir.

V. CONCLUSION

The aim of the current study, namely to investigate spectral differences between three styles of Croatian folk singing, was successfully reached, because it turned out that the differences between the styles could be established for all singers by factoring out all singer dependencies, which was achieved by using the same singer group for each of the three styles and by subtracting average speaker properties.

The spectral characteristics of the three styles were ultimately explained in detail by their 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 presence of low and high shouter’s formants in *ojkanje*. Finally, the presence of the perceptual goal of imitating the sound of the sopile explained the flat ('white') spectrum in *tarankanje*.

In authentic singers, the differences between the styles are expected to be higher, given the specialized skills required. Informal observations indeed suggest that authentic *ojkanje* performers shout even higher and louder, authentic *tarankanje* performers match the sound of the sopile even more closely, and authentic *klapa* singers sing even more harmonically and 'effortless'. Redoing the tests of the present paper with authentic performers, however, may still require many more participants than 12 per style, given that singer dependencies can no longer be factored out.

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 abovementioned phase of the investigation.

REFERENCES

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., 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.

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

Boersma, P., and Weenink, D. (**2004**). Praat: doing phonetics by computer [Computer program]. Retrieved from 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 Academy of Music, Stockholm), pp. 161-165.

Fant, G. (1960). *Acoustic theory of speech production* (Mouton, The Hague).

Frøkjær-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Brüel & Kjær Technical Review **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.

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 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 University Press, Cambridge Mass.).

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.

Sluijter, A. (**1995**). *Phonetic correlates of stress and accent*. PhD thesis, Leiden University (Holland Academic Graphics, The Hague).

Stevens, K. N. (**1998**). *Acoustic phonetics* (MIT Press, Cambridge Mass.).

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. (**2001**). "Level and center frequency of the singer's formant," J. Voice **15**, 176-186.

Sundberg, J., Cleveland, T. F., Stone, R. E., Jr, and Iwarsson, J. (**1999**). “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).

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

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

FIG. 3. Pooled LTAS for the 12 *tarankanje* performances, with pitch correction.

FIG. 4. Pooled pitch-corrected LTASes (— *tarankanje*, ···· *klapa*, - - - *ojkanje*).

FIG. 5. Individual pitch-corrected LTASes (— *tarankanje*, ···· *klapa*, - - - *ojkanje*). Horizontally: frequency in kHz; vertically: power spectral density in dB.

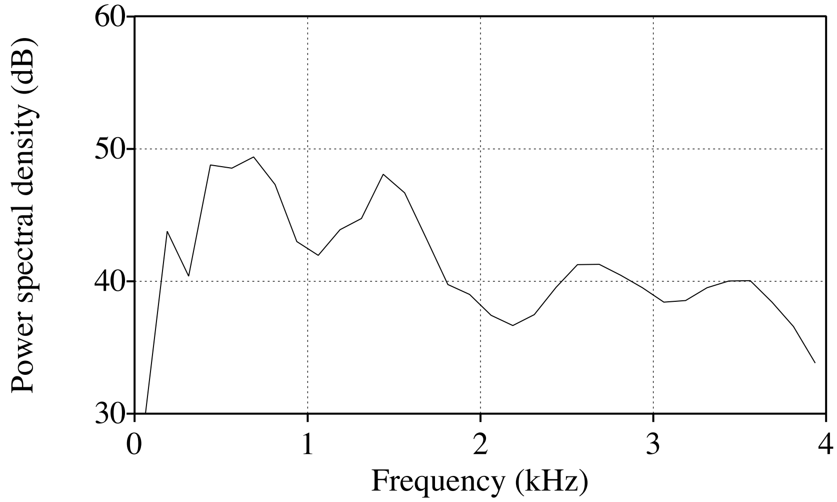
FIG. 6. The first two principal components of the 36 pitch- and speaker-corrected LTASes.

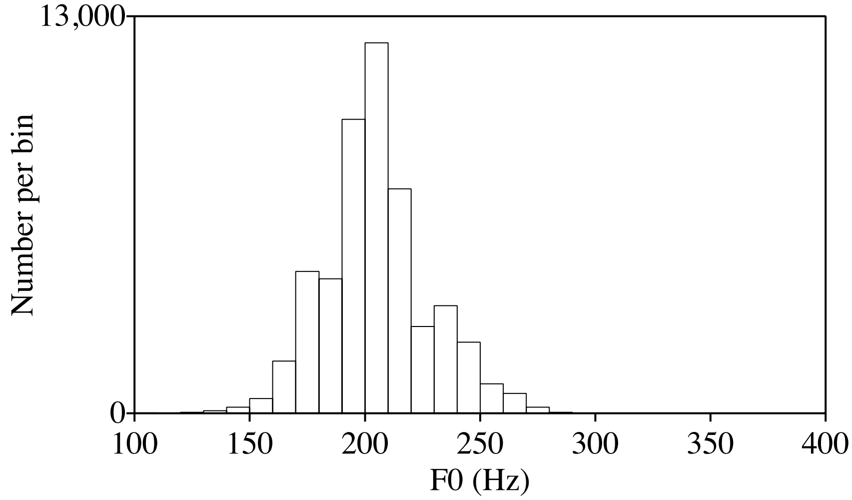
FIG. 7. Principal component analysis of the 36 performances. Pluses = *tarankanje*, rectangles = *klapa*, dots = *ojkanje*, 1...12 = the 12 singers.

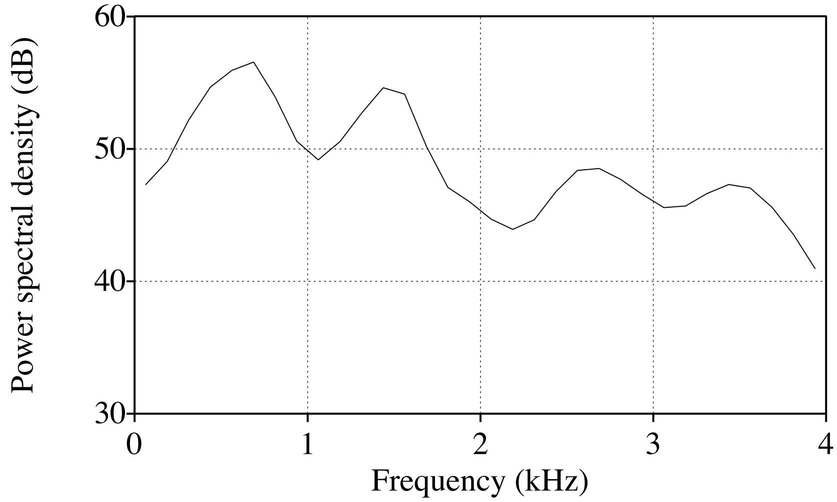
FIG. 8. The locations of the 36 performances in a space of two predefined spectral shape characteristics. Pluses = *tarankanje*, rectangles = *klapa*, dots = *ojkanje*, 1...12 = the 12 singers.

FIG. 9. Five periods of a vowel in *ojkanje* (Singer 9).

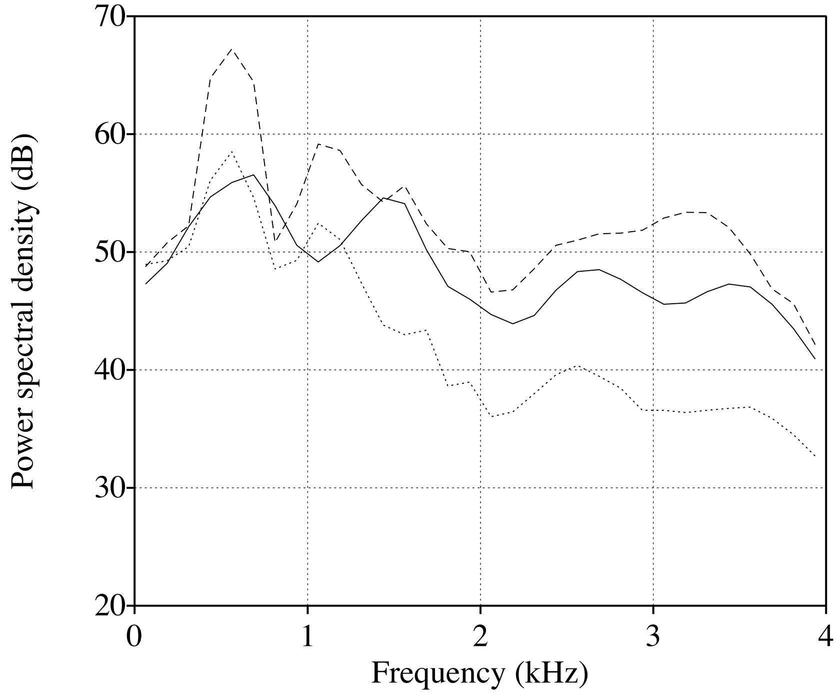
FIG. 10. Five periods of a vowel in *ojkanje* (authentic singer).

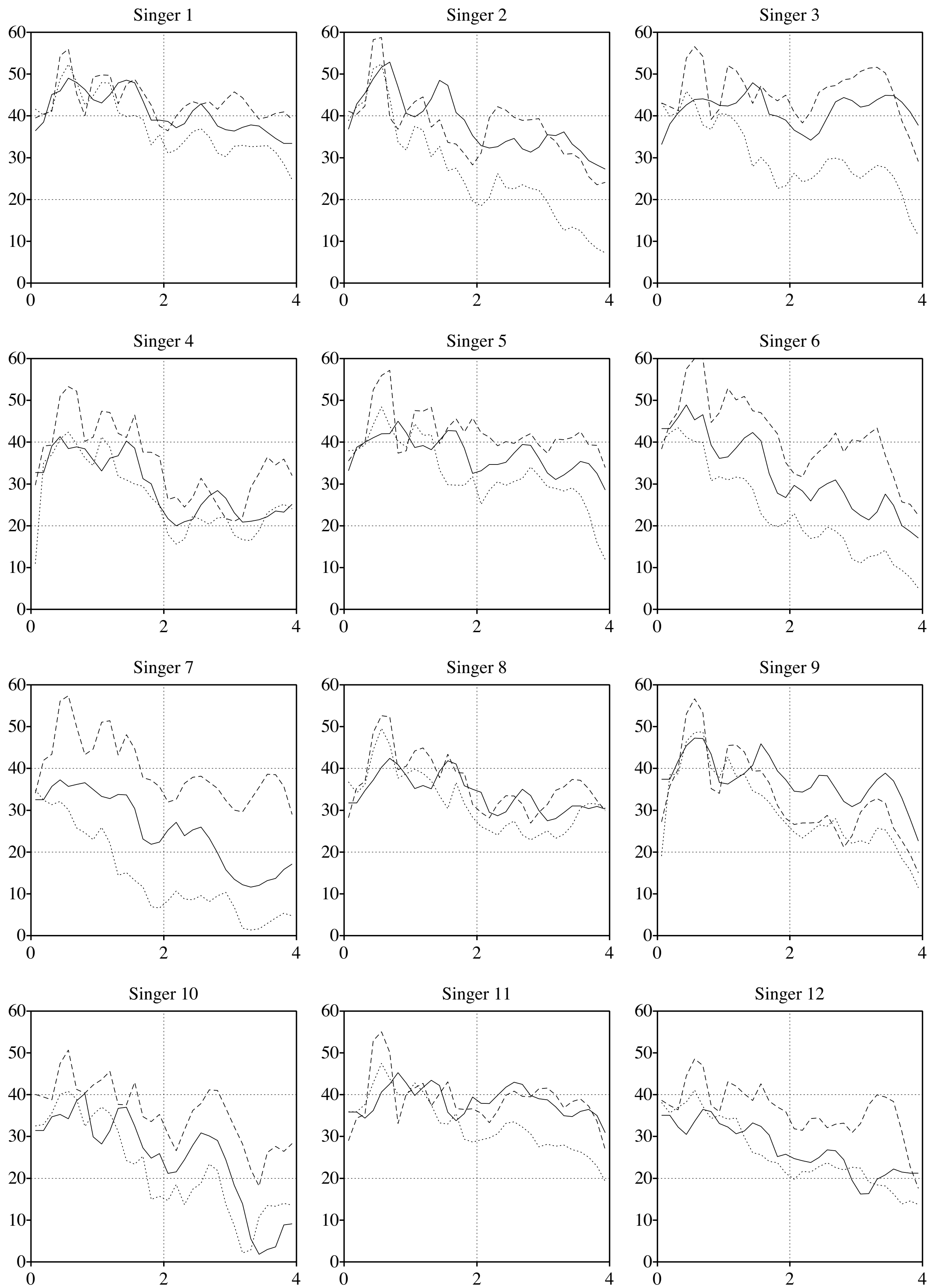

60
50
40
30
Power spectral density (dB)
0
1
2
3
4
Frequency (kHz)

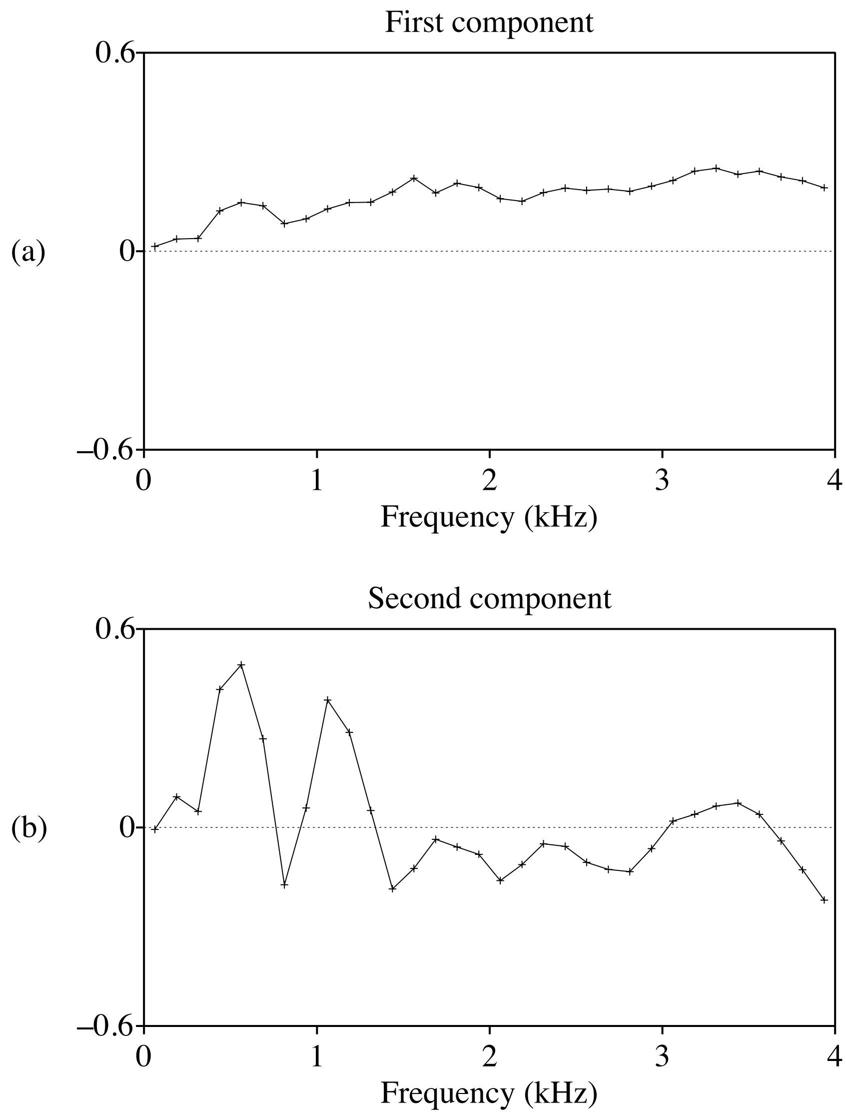

13,000
0
Number per bin
100
150
200
250
300
350
400
F0 (Hz)

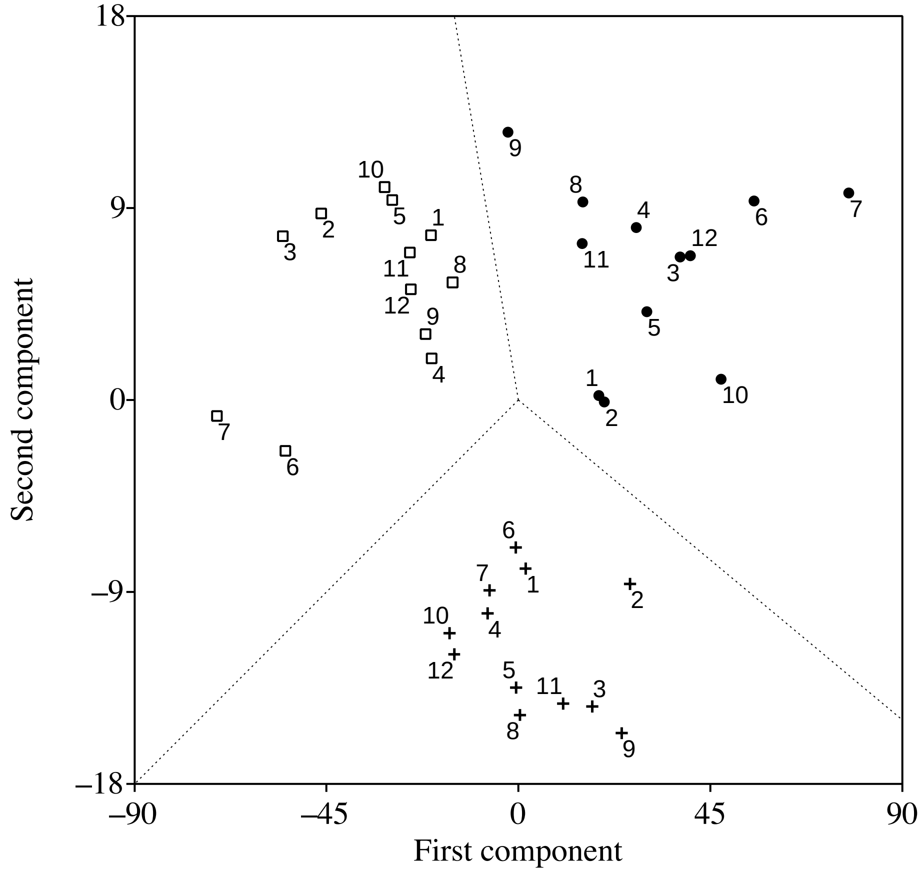


60
50
40
30
Power spectral density (dB)
0
1
2
3
4
Frequency (kHz)

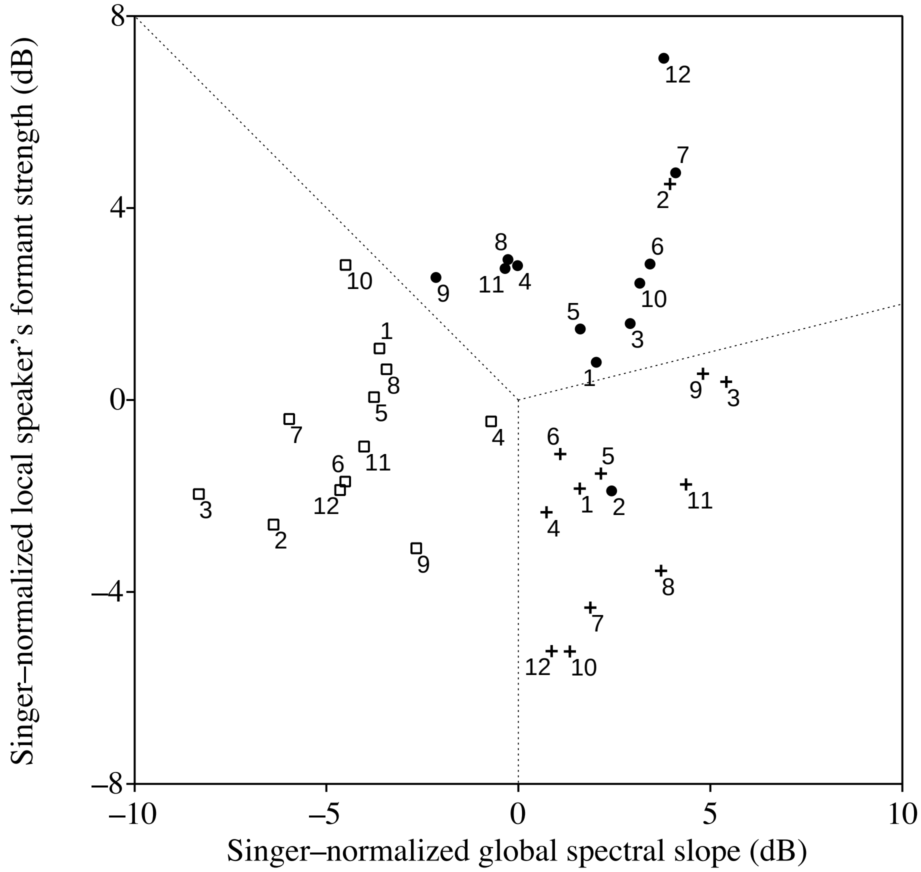

70
60
50
40
30
20
Power spectral density (dB)
0
1
2
3
4
Frequency (kHz)


Singer 1
Singer 2
Singer 3
Singer 4
Singer 5
Singer 6
Singer 7
Singer 8
Singer 9
Singer 10
Singer 11
Singer 12
0
10
20
30
40
50
60
0
2
4

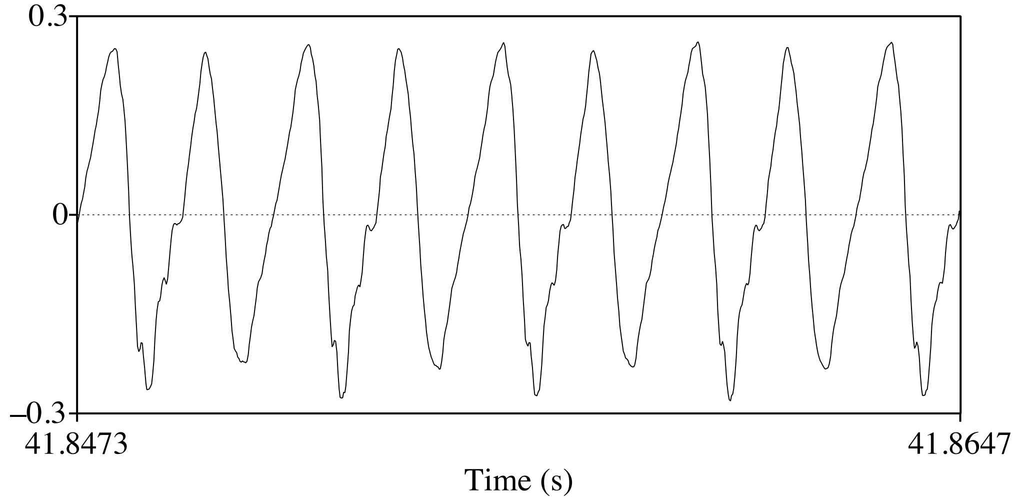

First component
0.6
(a)
0
–0.6
0
1
2
3
4
Frequency (kHz)
Second component
0.6
(b)
0
–0.6
0
1
2
3
4
Frequency (kHz)

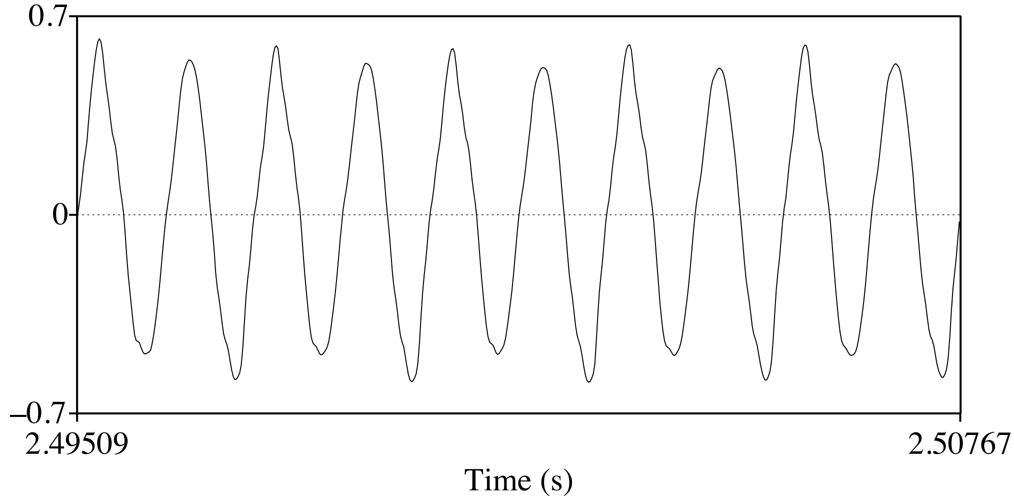


18
9
0
−9
−18
−90
−45
0
45
90
First component
Second component



Singer–normalized local speaker's formant strength (dB)
Singer–normalized global spectral slope (dB)


0.3
0
–0.3
41.8473
41.8647
Time (s)


0.7
0
–0.7
2.49509
2.50767
Time (s)