Frank Scherbaum, fs@geo.uni-potsdam.de, Universität Potsdam
Nana Mzhavanadze, mzhavanadze@uni-potsdam.de, Universität Potsdam

Svan Funeral Dirges (Zär): Language-Music Relation and Phonetic Properties

Abstract

This paper reports on the results of a follow-up study to Scherbaum and Mzhavanadze (2020) and Mzhavanadze and Scherbaum (2020), which jointly describe the acoustical and the musicological properties, respectively, of a new collection of field recordings of three-voiced Svan funeral dirges, known as *zär* in Svan and *zari* in Georgian. The focus of the present work is on the language-music relation and the phonetic properties. It was motivated by the pioneering work of Bolle Zemp (1994; 1997; 2001) who for the first time examined this topic from both an ethnomusicological and a linguistic perspective. We revisit some of Bolle Zemp's observations and assumptions with a much enlarged (by a factor of 11) data set and by using computational phonetic analysis tools not available at the time of her study. Her observation of correlations between pitch, duration, and timbre features related to the interjection of *woj* (*wai*) is consistent with our analysis for only some of the singers. Therefore, instead of assigning a general semantic meaning to these correlations, we offer an alternative interpretation as a natural consequence of the formant tuning techniques which we found to be employed in the vocal production by some of the singers.

I. Introduction

The present study was stimulated by the pioneering work of Bolle Zemp (1994; 1997; 2001) on the language-music relation and the phonetic properties of three-voiced Svan funeral dirges known as *zär* in Svan and *zari* in Georgian. We extend her study, which was focused on the audio recording of a single funeral chant from Latali (Lat'li in Svan), to the analysis of eleven different performances of six different variants of *zär*, performed by singers from different villages, which were obtained during an ethnomusicological field expedition in 2016 (Scherbaum and Mzhavanadze, 2018; Scherbaum et al., 2018; Scherbaum et al., 2019)[1]. Some of these performances were recorded in their natural

[1] All the recordings of the 2016 field expedition have been made publicly available and can be accessed either through the open access long-term archive at the University of Jena, which also hosts the field report and the meta data (https://lazardb.gbv.de/search; see Scherbaum et al., 2018 for details), or through the research repository at the University of

context at actual funerals which we were allowed to document. Based on the information obtained from our local informants during the 2016 campaign, the present collection represents more than half of the eleven different variants of *zär* believed to still be practiced today (Scherbaum and Mzhavanadze, 2020). The recording strategy employed during the 2016 field expedition makes the new corpus particularly suitable for modern state-of-the-art analysis. Whenever possible[2], one singer from each voice group was simultaneously recorded with a high quality headset microphone and a larynx microphone. In addition, the whole ensemble was recorded with a high resolution (4K) video camera on which a directional microphone was mounted, plus a conventional stereo microphone. The systematic use of larynx microphones allows the undistorted documentation of the acoustical contribution of each singer while all of them are singing together in their natural context (Scherbaum et al., 2015). In addition, larynx microphone recordings have been shown to contain essential information in relation to a singer's voice regarding pitch, intonation, timbre and voice intensity which allows the application of computer based methods to document and analyse vocal music of the oral tradition in new ways, e. g. to apply computerised pitch analysis techniques to determine the fundamental frequency (F0) trajectories and their microtonal structure, to study the tuning systems used by the singers, as well as possible interactions between singers (Scherbaum et al., 2015; Scherbaum, 2016).

Erlangen of the GVM project (https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Scherbaum). Contact the first author (fs@geo.uni-potsdam.de) for access information.

[2] For reasons of reverence we have refrained from using headset microphones for the recordings which were made at real funerals.

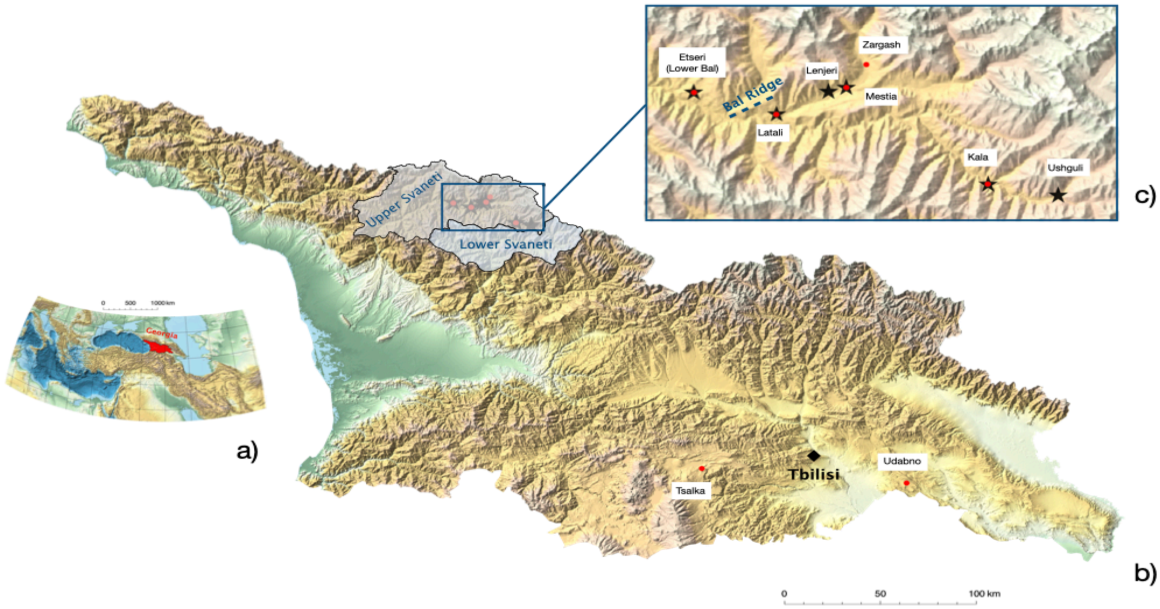

Figure 1. Geographical situation. a) The location of Georgia in its regional context.
b) Study area and recordings sites (marked by red solid circles) of the field expedition of 2016. c) Locations of recording sites (red solid circles) and locations of origin of zär (black stars) within Upper Svaneti. The Bal ridge (altitude-wise) divides the Lower Bal and Upper Bal regions of Upper Svaneti.

The recording locations and the locations of origin of the different *zär* variants are shown in Figure 1 by red solid circles and black stars, respectively. Apart from the Upper Svaneti area, the field work also covered the eco-migrant Svan communities of Didgori, Tsalka and Udabno, outside Svaneti, near the capital of Tbilisi (Figure 1).

2. Language-music relation

The quantitative investigation of language-music relations in *zär* was originally initiated by the Swiss ethnomusicologist Sylvie Bolle Zemp, who discussed the results of her studies in a sequence of papers, the core of which is in German (Bolle Zemp, 1994; 1997; 2001). What made her work so innovative at the time is that she examined this topic from both an ethnomusicological and a linguistic perspective. For her analysis, Bolle Zemp had at her disposal the recordings of a single funeral chant from Latali, which she phonetically transcribed using the symbols of the International Phonetic Alphabet (in its 1993 version). Employing the tools accessible at that time, she processed the chant through sonographic images and analysed the musical content of the verbal ‘text’; explored all the possible references of the utterances employed in *zär*, and visualised the results in the form of sonograms. Based on the results of her analysis, Bolle Zemp suggested a strong correlation between words and music. She assigned semantic importance to the core utterance *woj*, which she interpreted as an interjection of pain, equivalent to the Georgian *wai*. She argued that as an utterance of mourning connotation, it lies at the root of several vocal formulas of the chant. In her view, 'singers' emphasize human emotions such as pain, dignity, etc. by modification of the sound characteristics of spoken language, e.g. by formation of vowels and consonants in different ways, by manipulation of the interjections, and by stylisation of expressions of the spoken language through certain vocal process (velarisation, descending glissandi, nasalisation). In her view, the structure of the movement of voices, the duration of a sound, the sequence of concomitants, and the interrelationship of consonant and dissonance chords is greatly conditioned by the 'text', which coordinates the musical process (Bolle Zemp, 1997, 2001). In particular, she suspected a correlation between a sequence of high/low/high timbres, high/low/high pitches and short/long/short durations of vocal formulas. Although the research methodology of Bolle Zemp's study makes it transparent and reproducible, the analysis of only one example of *zär* limits the generalizability of her results to other variants of the chant. Besides, although the general quality of the recordings of S. Bolle-Zemp is good, the recordings of that time do not allow the analysis of the acoustical properties of individual voices because of the then unsolved problem to separate individual voices from polyphonic field recordings[3].

In this section, we are revisiting some of the issues brought up by Bolle Zemp (1997). In particular we want to test if if her hypothesis regarding possible correlations between pitch, duration, and timbre features could be tested quantitatively by using a computational acoustical phonetic analysis. Since each of the individual singers in our corpus was separately recorded, we can quantitatively exploit the acoustic properties of the individual voices as well as their interaction for this purpose. For each voice, the F0 trajectory as well as the sequence of individual note objects (containing onset time, pitch, duration and lyrics information) were determined (cf. Figure 3 in Scherbaum and Mzhavanadze, 2020). The text of each note object was identified by a native Georgian speaker (Nana Mzhavanadze) in collaboration with two native Svan speakers (Madona Chamgeliani and Ketevan Margiani) and expressed in Svan language. For this purpose, the Svan character set on the Titus website[4] was used.

In order to obtain an overview of the lyrics used in the complete collection, Figure 2 shows an overview of the relative frequencies of occurrences of the individual note lyrics in the

---

[3] This problem becomes also acute when focusing on retrospective study of the chant to reconstruct the stages of its development and changes. The quality of older archive recordings are often critically poor and sometimes does not allow even a minimal manual processing.

[4] http://titus.uni-frankfurt.de/didact/caucasus/kartlaut.htm, http://titus.uni-frankfurt.de/didact/caucasus/kaukvok.htm

form of so-called word clouds in which the size of a word corresponds to the relative frequency of its occurrence in the corresponding *zär*. This frequency is not always the same for the same *zär* variant as can be seen for example in Figure 2 a) and b), d) and e), g) and h), and i) - k). In particular, the Latali variant, which was the subject of Bolle Zemp's work, has considerably different lyrics in the realizations by the singers from Latali and from Udabno, respectively. Figure 2 indicates that overall the most common note texts in our transcriptions are {"ჟ̂ა", "ჟ̂ო", "ჲი", "იჲჰ", "ჰა"}[5].

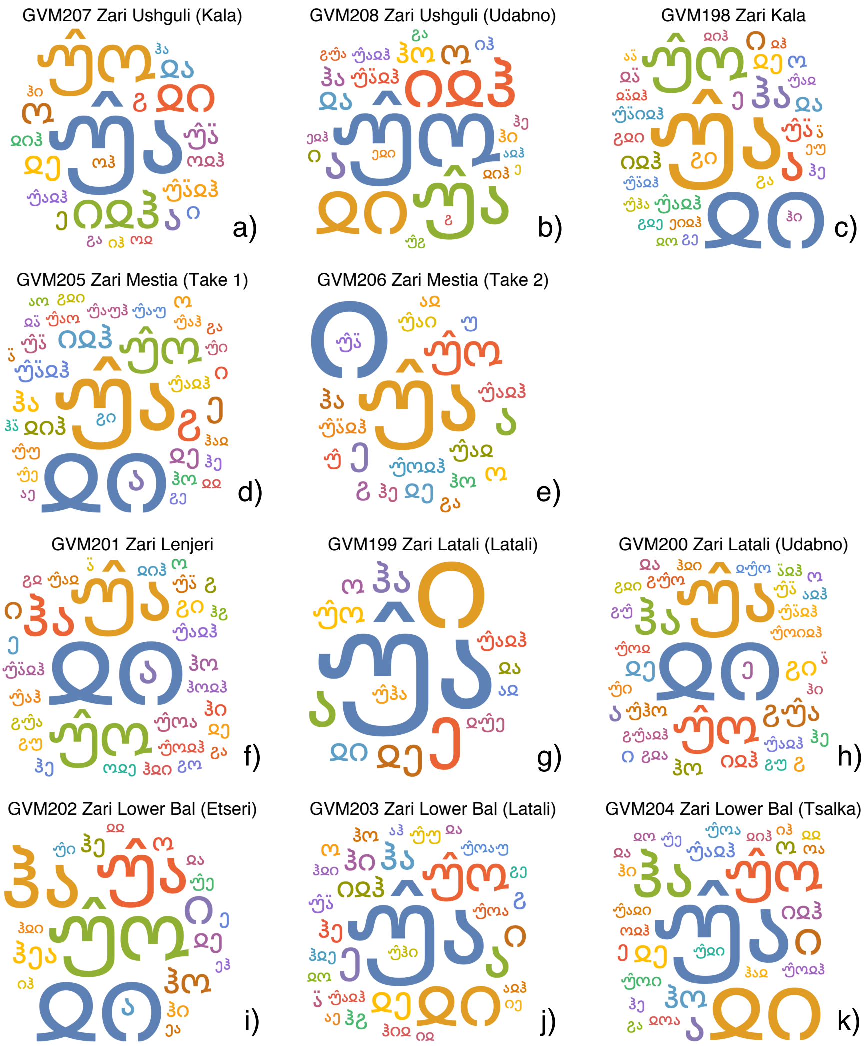

Figure 2. Word cloud derived from the set of note texts from the complete collection of *zär* recordings (see Scherbaum and Mzhavanadze, 2020). The size of a word corresponds to the relative frequency of its occurrence in the corresponding zär.

The main focus of Bolle Zemp's interpretation was on the interjection *woj*, which she assumed to be related to the Georgian interjection of pain *wai*. In Svan, these two interjections would be written as "ჟ̂აჲი" or "ჟ̂ოჲი", which Bolle Zemp perceived as a single word. In our attempts to transcribe the zär lyrics, we almost always perceived "ჟ̂ა" ("ûa")- "ჲი" ("yi") and "ჟ̂ო" ("ûo") - "ჲი" ("yi") as two different notes and transcribed

[5] In transcription: {"ûa", "ûo", "yi", "iyh", "ha"}.

them accordingly. It goes without saying that the quality of phonetic transcriptions depends strongly on the training of the transcriber. But even for well trained transcribers, as demonstrated in section 13.4 of Reetz and Jongman (2009), it is very hard not to be influenced in perception by one's mother tongue. In the present study, although done by natural speakers and experienced singers, the identification of the note lyrics was still found to be extremely challenging. For example, it was not always straight forward to distinguish between "უ̂ა" ("ûa"), and "უ̂ო" ("ûo") sounds or between the different realisations of vowel sounds involving "ი" ("i"). For the subsequent analysis, in order to overcome some of the subjectivity of the phonetic transcriptions, we therefore performed a computational acoustical phonetic analysis.

2.1. Phonetic analysis

Since each of the individual voices was separately recorded by either a larynx microphone or by a combination of a headset- and a larynx microphone, we performed a so-called acoustical phonetic analysis, based on the analysis of voice formants. Formants are resonance frequencies of the vocal tract which are related to its shape and which are controlled by the position of the tongue, the jaw, the lips and the velum, all of which are referred to as articulators (cf. Sundberg, 1987; Reetz and Jongman, 2009). The articulatory description of *placement* of consonants or *frontness* or *height* of vowels, or the description of vowels in terms of quality or color, as used by Bolle Zemp (1997), can acoustically be related to the values of the formant frequencies. The *height* of a vowel for example is related to the value of the first formant (F1), while the *frontness* (which is related to the position of the arching tongue) is related to the second formant (F2). The first two formants are essential for the identification of vowel sounds, while the higher formants (F3, F4, ..) are believed to be more related to the identity of the singer. Formants are associated with peaks in the smoothed power spectrum of the sound. In the present study, we used the Linear Predictive Coding (LPC) algorithm implemented in Mathematica (Wolfram Research, Inc., 2020), which is also used in the standard phonetic software Praat (Boersma, P., and Weenink, D. , 1992). Since formant frequencies are controlled by the shape of the vocal tract, they are believed to be independent of the fundamental frequency of the vocal fold oscillations (F0) (Reetz and Jongman, 2009).

Figure 3 illustrates the processing sequence of the phonetic analysis. Each of the note objects in Figure 3a) is defined by an onset time, a F0 value, a duration and a text string representing the lyrics. For a note of interest, the onset time and duration information is used to cut out the corresponding audio segment from the audio file. This is shown in Figure

3b) for a note with the lyrics "უ̂ა" ("ûa") and in Figure 3c) for a note with the lyrics "უ̂ო" ("ûo"). The main panels in b) and c) show the spectrograms of the audio segments, the waveforms of which are shown by the gray and brown waveform segments, respectively, on top of the main panels. The small figures to the right of each spectrogram correspond to the Short-Time Fourier Transform (STFT) of the complete audio segment of the note object. The horizontal red dashed lines indicate the first four formant frequencies (F1, F2, F3, F4) as determined by the Linear Predictive Coding (LPC) algorithm (for details see Wolfram Research, Inc., 2020). Their numerical values are displayed together with the values of the fundamental frequency of the notes in the tables in Figure 3 d) and e).

Finally, F1 and F2 are plotted on a so-called vowel map in Figure 3f) together with the average formant frequencies for the vowels of American English (in magenta) collected by Hillenbrandt et al. (1994) and for German (in blue) by Sendlmeier and Seebode (2006). The positions of the two examples with the lyrics "უ̂ა" ("ûa") and "უ̂ო" ("ûo") fall roughly on a line between the German "a:" and "u:".

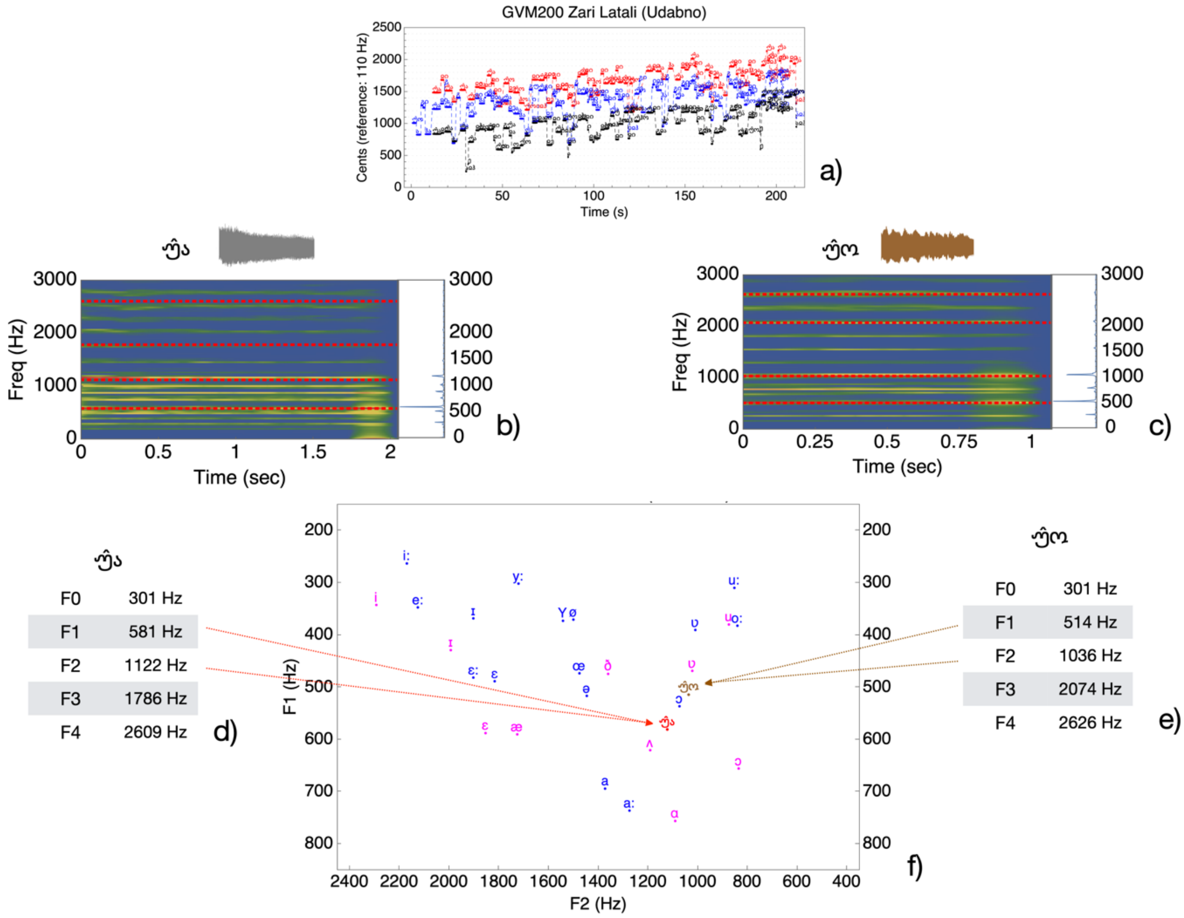

Figure 2. Illustration of the workflow of the phonetic analysis.

Figure 4 shows the vowel map for all notes with lyrics "უ̂ა" ("ûa") in red or "უ̂ო" ("ûo") in brown in the headset recordings of the Latali *zär*, sung by the singers from Udabno.

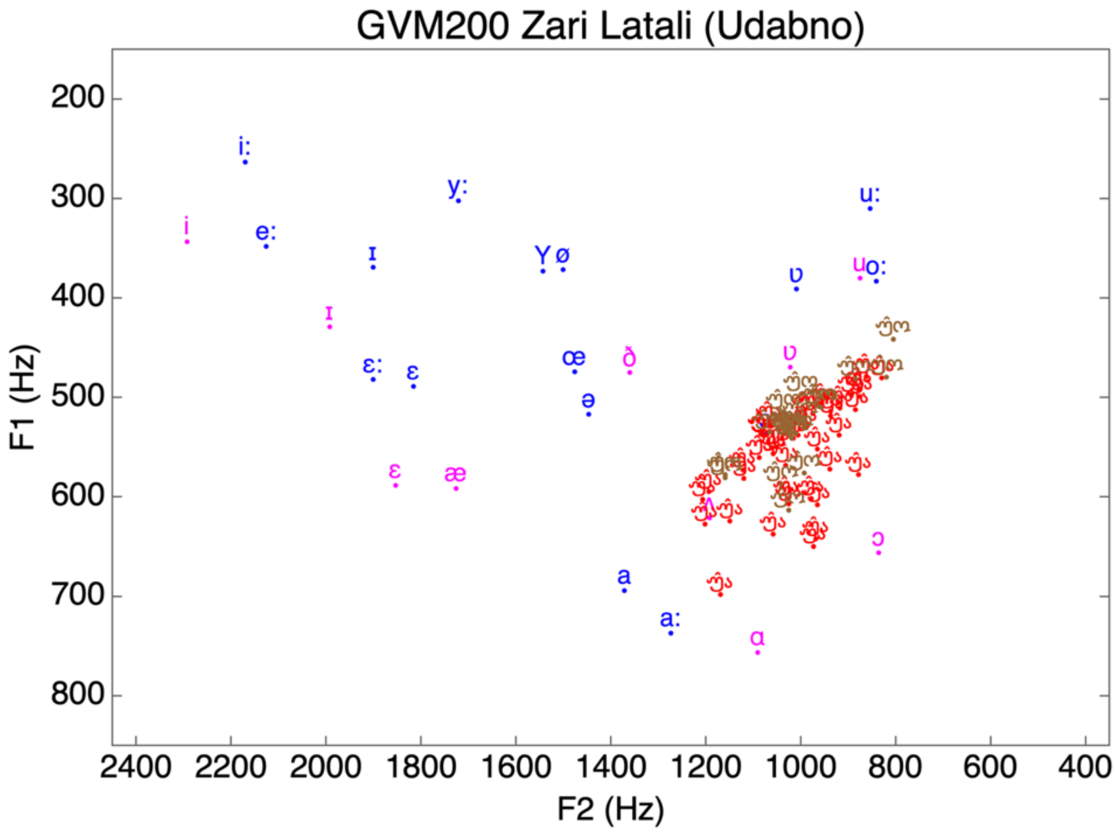

Figure 3 Vowel map for all notes with lyrics "უ̂ა" ("ûa") in red or "უ̂ო" ("ûo") in brown in the headset recordings of the Latali zär, sung by the singers from Udabno.

It can be seen that the vowel map position of the individual notes fall again between the German “a:” and “u:” sounds, but largely overlap. For once this is in line with our experience that the two sounds are sometimes difficult to distinguish. The word clouds in Figure 2 show that the syllables "უ̂ა" and "უ̂ო" are definitely very present in the phonetic soundscape of all zär variants. Comparing the durations of all notes with note lyrics "უ̂ა" or "უ̂ო" with the total duration of all notes in the particular zär reveals that in terms of durations roughly 50% of all zär lyrics in all zär variants are spent on either "უ̂ა" or "უ̂ო" . In about 1/3 of these cases, "უ̂ა" or "უ̂ო" appear in combination with either "ჲი" or "იჲჰ" as subsequent note.

As a consequence, Bolle Zemp’s conjecture regarding the importance of the interjections *woj* /*wai* seems to generalize well to all zär variants, if one assumes that they correspond to different combinations of "უ̂ა" ("ûa") or "უ̂ო" ("ûo") for the first note text and "ჲი" ("yi") or "იჲჰ" ("iyh") for the second one in our study. But does this mean, as Bolle Zemp (1997) suggested, that these syllables have a strong mourning connotation and a sematic meaning, which she suspected based on postulated correlations in changes in pitch, duration, and timbre features?

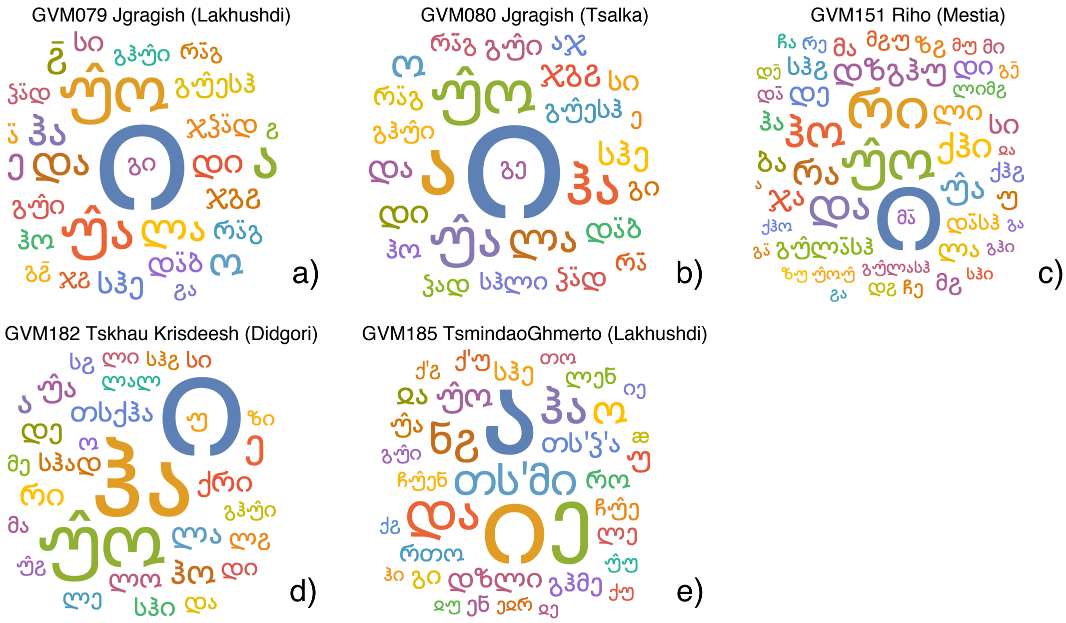

Figure 4. Word cloud derived from the set of note texts from a selection of other Svan song recordings.

As one can see in Figure 5, the syllables "჻ა" ("ûa") or "჻ო" ("ûo") are present in significant amounts in other Svan songs as well. This raises doubts regarding their connotation being limited to mourning.

Since the first formant (F1) is related to the *height* of a vowel, while the second formant (F2) is related to the position of the arching tongue and therefore to the amount of velarization, one can check for correlations between changes in pitch, duration, and timbre features by calculating pairwise correlations between F0, F1, F2 and note durations. For all three voices in the Latali zär, sung by the Udabno singers, and for all notes with lyrics "჻ა" or "჻ო", these so-called Trellis plots[6] are shown in Figure 6.

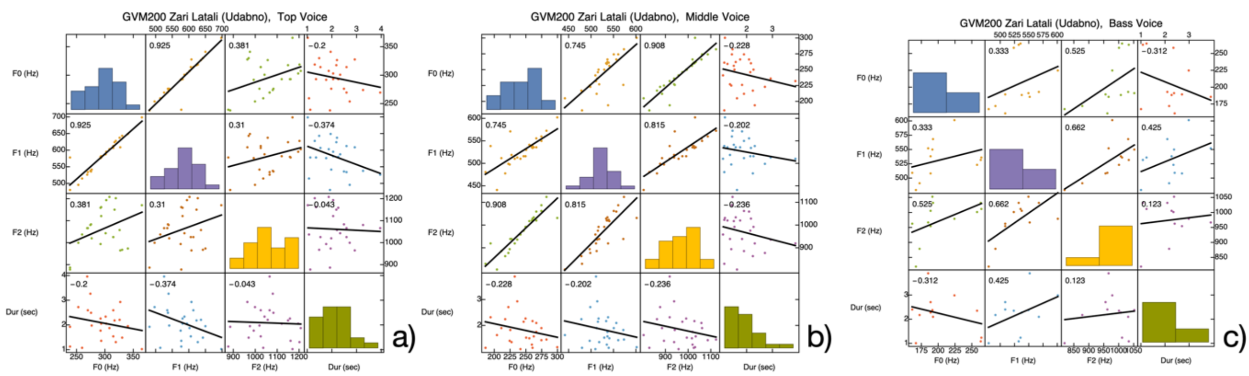

Figure 6 Trellis plot of F0, F1, F2, and note duration of notes with lyrics "჻ა" or "჻ო" in the Latali zär, sung by the Udabno singers.

[6] A Trellis plot is a group of smaller plots arranged in a grid in which each subplot is conditioned on a different variable.

With correlation coefficients above 0.9 in both cases, Figure 6 a) and b) show strong correlations between F0 (pitch) and F1 (related to vowel height) for the top voice and between F0 (pitch) and F2 (related to tongue position) for the middle voice, respectively. For the bass voice, the correlations are much weaker and hence not really convincing. Closer inspection revealed that the top voice singer shifts his first formant towards 2 x F0 (cf. dashed line in Figure 7a), while the middle voice singer shift his second formant towards the fourth harmonic of the sound spectrum which has a frequency of 4 x F0 (cf. dashed line in Figure 18b). This "formant tracking" (Bozeman, 2013) gives particular harmonics, in the present case the second (H2 at 2 x F0) and fourth (H4 at 4 x F0), an energy boost, makes them appear louder, rougher, and gives them a "ringing" quality. It is a vocalization strategy well known for both classical and non-classical singing (Bozeman, 2013; Sundberg et al., 2013 and references therein).

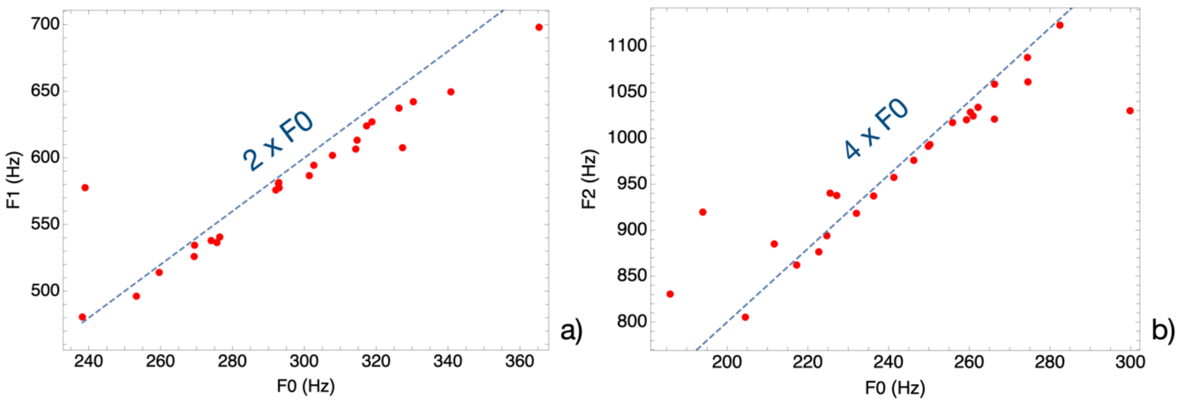

Figure 7. Evidence for formant tuning in the top voice (a) and middle voice (b) recordings of the Latali zär, sung by the Udabno singers.

Men are known in particular to tune their second formant to higher harmonics if they want a real powerful note or when their voice turns over (Bozeman and O'Connor, 2017). When harmonics pass through the first formants, one can hear this as a closed timbre ("voce chiusa"), the more harmonics are below the first formant, the more the timbre is called open ("voce aperta") (Bozeman and O'Connor, 2017)

Since in the present example the correlations in changes in pitch and timbre features (we could not detect a correlation with durations) differ in style between the different voices or are not even used at all, as in the case of the bass voice, we interpret them more as an expression of personal taste of the singers than as having a semantic meaning. In the latter case, we would expect them to be similar for all three voices.

3. Conclusions

In this note, using a computational phonetic analysis, we have revisited some of the conjectures made in the pioneering study by Bolle Zemp (1997). Her observation of the frequency of occurrence of the interjections *woj* (*wai*) in *zär* is consistent with our recordings. However, we come to a different conclusion regarding the interpretation of correlations between pitch, duration, and timbre features.

Acknowledgements

First and foremost, our gratitude goes to all the people during the 2016 field expedition who allowed us to be part of and record their rituals. We are thankful to Meinard Müller and his team for the stimulating collaboration within the framework of the project "Computational Analysis of Traditional Georgian Vocal Music (GVM)" (DFG MU 2686/13-1, SCHE 280/20-1) as well as for hosting the web-based repository of the GVM data.

References

Boersma, Paul, and David Weenink. 1992. “Praat: Doing Phonetics by Computer.” http://www.praat.org.

Bolle Zemp, Sylvie. 2001. “Khmovnebi Da Akordebi. Simghera Zemo Svanetshi (Vowels and Chords. Singing in Zemo Svaneti).(in Georgian with English Summary).” In *Sasuliero Da Saero Musikis Mravalkhmianobis Problemebi (Problems of Polyphony in Sacred and Secular Music).*, edited by Rusudan Tsurtsumia, 292–303. Tbilisi: Tbilisi State Conservatoire.

Bolle Zemp, Sylvie. 1994. “Georgia. Polyphony of Svaneti.” https://doi.org/UMR 9957.

Bolle Zemp, Sylvie. 1997. “Mehrstimmige Wehrufe.” *Georgica*, 134–48.

Bozeman, Kenneth W. 2013. *Practical Vocal Acoustics*. Vox Musica. Pendragon Press, Hillsdale, NY.

Bozeman, Kenneth W., and Karyn O’Connor. 2017. *Formant Tuning, Vowel Modification, Male and Female Passaggio, and Registration*. Youtube. https://www.youtube.com/watch?v=dxttfvX-FRs.

Hillenbrand, James, Laura A. Getty, Kimberlee Wheeler, and Michael J. Clark. 1994. “Acoustic Characteristics of American English Vowels.” *The Journal of the Acoustical Society of America* 95 (5): 2875–2875. https://doi.org/10.1121/1.409456.

Mzhavanadze, Nana, and Frank Scherbaum. 2020. “Svan Funeral Dirges (Zär): Musicological Analysis.” *Musicology* submitted.

Reetz, H., and Jongman, A. (2009). *Phonetics. Transcription, Production, Acoustics, and Perception*. Wiley-Blackwell.

Scherbaum, Frank. 2016. “On the Benefit of Larynx-Microphone Field Recordings for the Documentation and Analysis of Polyphonic Vocal Music.” *Proc. of the 6th International Workshop Folk Music Analysis,15 - 17 June, Dublin/Ireland*, 80–87.

Scherbaum, Frank, Wolfgang Loos, Frank Kane, and Daniel Vollmer. 2015. “Body Vibrations as Source of Information for the Analysis of Polyphonic Vocal Music.” In *Proceedings of the 5th International Workshop on Folk Music Analysis, June 10-12, 2015, University Pierre and Marie Curie, Paris, France*, 5:89–93.

Scherbaum, Frank, and Nana Mzhavanadze. 2020. “Svan Funeral Dirges (Zär): Musical Acoustical Analysis of a New Collection of Field Recordings.” *Musicologist* accepted for publication.

Scherbaum, Frank, and Nana Mzhavanadze. 2018. “A New Archive of Multichannel-Multimedia Field Recordings of Traditional Georgian Singing , Praying , and Lamenting with Special Emphasis on Svaneti.” LaZAR-Database. Https://Lazardb.Gbv.de/

Scherbaum, Frank, Nana Mzhavanadze, Sebastian Rosenzweig, and Meinard Müller. 2019. “Multi-Media Recordings of Traditional Georgian Vocal Music for Computational

Analysis.” In *9th International Workshop on Folk Music Analysis, 2-4 July, 2019*, accepted. Birmingham.

Scherbaum, F., Mzhavanadze, N., Arom, S., Rosenzweig, S., and Müller, M. (2020). *Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies*. (F. Scherbaum, Ed.), *Computational Analysis Of Traditional Georgian Vocal Music*.

Scherbaum, F., and Mzhavanadze, N. (2020). Svan Funeral Dirges (Zär): Musical Acoustical Analysis of a New Collection of Field Recordings. *Musicologist*, *accepted for publication*.

Sendlmeier, Walter F, and Julia Seebode. 2006. “Formantkarten Des Deutschen Vokalsystems.” Berlin. http://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf.

Sundberg, Johan. 1987. *The Science of the Singing Voice*. Northern Illinois University Press.

Sundberg, Johan, Filipa M.B. Lã, and Brian P. Gill. 2013. “Formant Tuning Strategies in Professional Male Opera Singers.” *Journal of Voice* 27 (3): 278–88. https://doi.org/10.1016/j.jvoice.2012.12.002.

Wolfram Research, Inc. 2020. *Mathematica*. Version 12. Champaign, Illinois: Wolfram Research, Inc.