Abstract. Russian lament is a vocal form used by village women to express grief that can be either spontaneous or ritualistic. The purpose of this study is to examine the patterns of vocal fold vibration as observed from high speed imaging during singing, speaking and lamenting. It was found that for lamenting, compared with singing or speaking, the vocal folds are (1) longer and thinner, (2) more tense and (3) less periodic. These observations are accompanied by (1) higher F0, (2) smaller H1-A3 values, and (3) more acoustic instability. More work is needed with additional speakers to confirm these findings.

1. Introduction

Russian lament is a vocal form used by village women to express grief that can be either spontaneous or ritualistic. Lament is a cross-cultural phenomenon, existing in a variety of cultures around the world. Its performance is a highly emotional process that combines singing, chanting, and crying. Lament is always improvised, although it is based on the culture-dependent patterns that regulate, among other performative aspects, the use of voice quality, sobbing, excited exclamations, speech interruptions, sighs, voiced breathing (both inhalations and exhalations).

Vocal fold vibration in vocal expression of sadness: lamenting, speaking and singing

Margarita Mazo\textsuperscript{1}, Ken-ichi Sakakibara\textsuperscript{2}, Hiroshi Imagawa\textsuperscript{3}, Niro Tayama\textsuperscript{4} and Donna Erickson\textsuperscript{5}

\textsuperscript{1}School of Music at The Ohio State University, \textsuperscript{2}Health Sciences University of Hokkaido, \textsuperscript{3,4}University of Tokyo, \textsuperscript{5}Showa Music University

\textsuperscript{1}\texttt{ophee-mazo.1@osu.edu}, \textsuperscript{2}\texttt{kis@hoku-iryo-u.ac.jp}, \textsuperscript{3}\texttt{imagawa@m.u-tokyo.ac.jp}, \textsuperscript{4}\texttt{ntayama@hosp.nsgm.go.jp} & \textsuperscript{5}\texttt{EricksonDonna2000@gmail.com}
the signal, whereas lamenting is slower, higher in pitch, and has irregular pitch/intensity.

Studies of listeners’ perceptions of the sadness conveyed by the different modes of vocal expression (lamenting, singing and speaking) was done with American college students with various cultural backgrounds and showed that laments were well-perceived as being more sad than singing or speaking [3,4,7]. The more recent perceptual studies including heart-rate monitoring [4] show that although the results of the heart are very complex, generally perception of lament is accompanied by increased fluctuation in rate of heart rate (BPM and SD). Figure 1 below illustrates heart beat responses to listening to the different modes of vocal expression of grief for a single male speaker (N). The figure shows the standard deviation of his BPM responses -- lamenting (triangle) and speaking (square), and singing (diamond). Notice the increasing standard deviation in lamenting during the stimulus period, while in the other two modes, the standard deviation decreases.

These results suggest that expression of grief in lamenting is perceived very differently from that in singing or speaking. We surmise that the different acoustic characteristic of lament, including the instability of pitch and amplitude, timbral changes, etc. may well affect the differences in the perception of lament.

The interpretation of the previous studies of lament suggested that the way the vocal folds produce these acoustic instabilities during laments might be as follows: For regular singing or speaking, a flexible control of vocal fold tension and a certain air pressure are required in order to sustain uninterrupted utterances without a change in voice quality. During lamenting, however, the physical conditions are sometimes at the point where these requirements can barely be satisfied. This is because that in order to produce laments, it is important for the lamenter to allow intense emotions to come into play, resulting in loss of control. It is also necessary to maintain the culturally-prescribed form of the lament--among other things, to continue to verbalize linguistically, etc.. Under this condition, any minute change, including articulation of consonants, becomes difficult. These two demands (losing control as there is a buildup of emotion vs. maintaining structure) create extreme tension, especially as the lament progresses. Under intense emotion, the whole body responds. The lamenter eventually no longer has control of her muscles, including vocal and diaphragm muscles. The result is an abnormal tension in her vocal folds that does not allow fine tuning of the vocal output. This lack of control results in instability in pitch and intensity (amplitude) during lamenting. It also could cause

![Standard Deviation for Individual N](image)

Fig. 1. Differences in perception of three modalities of verbal expression of grief, as indicated by variation of heartbeat (standard deviation of BMP) [4]
abrupt and significant changes in voice quality or "unusual" timbral phenomena, such as double phonation or amplitude/frequency modulation as reported earlier [4,5].

The purpose of this current study is to examine some of the physiological differences in the performer while producing lamenting, singing and speaking; specifically, the patterns of vocal fold vibration occurring in these three different modalities of expressing sadness and grief. The questions we ask are the following: (1) How do the vocal folds produce the acoustic instability found in laments? (2) How is the pattern of vibration different for laments, compared with speaking and singing? (3) How well do the acoustic measurements of vocal fold vibration (i.e., instability (jitter and shimmer) and glottal opening characteristics as measured from amplitudes of the harmonics) compare with the visual observations of the vocal fold vibration patterns? and (4) How do listeners with no knowledge of Russian language perceive the emotion in each of the modalities?

By examining these questions we hope to take a step forward in better understanding the physiological, acoustical and perceptual characteristics of different modalities of vocal expressions of grief.

2. Experimental methods

In order to answer the above questions, two experiments were performed: (2) high speed imaging of the vocal folds during production of the three modalities of vocal expression of grief, (2) perception testing with non-speakers of Russian to see if they could identify the emotion in each of the modalities, and also, to rate the emotional intensity of utterances. Acoustic analyses of the utterances were made for comparison with the high speed imaging.

2.1. High Speed Imaging Experiment.

High speed image recordings were made at the University of Tokyo Univeristy Hospital of the subject speaking, singing and lamenting with the same verbal materials. The subject was one of the authors of this paper, an ethnomusicologist who has published widely on lament, who also has singing experience. At the time of recording, she was under extreme personal emotional pressure. The verbal material was the Russian word for “mother (of mine) ”-- “Matushka (ty moia)”. Each modality was produced and recorded two times for a total of 6 utterances.

2.2. Perception Tests

Eleven Japanese female music university undergraduates who had no knowledge of Russian listened to the lamented, sung, and spoken utterances recorded during the highspeed imaging experiment and (a) rated how emotional the utterances were and (b) identified what emotion they heard. The ratings were from 1 to 5, with “5” indicating “extremely emotional”, “3” as “emotional”, and “1” as “not emotional”. The tests were presented through HDA200 Sennheiser headphones in a quiet room, using a Windows-based computer software from Runtime Revolution. There were a total of 12 utterances (2 randomizations of each of the 6 utterances). Each test was preceded by a practice test of 5 utterances.

2.3. Acoustic Analysis

In order to measure vocal fold instability in the
utterances, vocal jitter and shimmer were examined using the free downloadable acoustic analysis software Praat. The definitions of these according to the Praat manual are as follows: (a) Jitter (local): the average absolute difference between consecutive periods, divided by the average period; and (b) Shimmer (local): the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

In order to measure glottal opening/closing characteristics, amplitude changes in the harmonic structure of the acoustic signal were measured using the free downloadable acoustic analysis software Wavesurfer. Specifically, we measured the amplitude (dB) of the fundamental frequency (H1), the next harmonic (H2), and that of the strongest harmonic within the third formant (A3). The measurements were made at the acoustic steady-state center of the vowel. Increased values of H1-H2 and H1-A3 are said to be indicative of a more open glottis and slower speed of glottal closing, respectively (e.g., [6]). Hence, a large H1-H2 value would be associated with a more breathy voice (due to a larger glottal opening), and a small H1-A3 value, with a more tense voice (since glottal closing would be more abrupt, presumably due to stiffer vocal folds).

3. Results
3.1. Perception tests
The results of the perception tests indicated that listeners perceived lament as sad 93% of the time, and rated the laments as being extremely emotional (average rating of 4.4, where 5 indicates extremely emotional). These findings are consistent with previous studies about listeners’ perception of laments.

However, the sung and spoken version of the laments were not heard as predominantly sad nor especially emotional. In fact, singing was heard 43% of the time as happy while speaking was heard 75% of the time as angry. Neither singing nor speaking was rated as emotionally intense (ratings of 2.9, and 1.8, respectively). Why singing was heard predominantly as happy and speaking as angry are interesting questions which go beyond the scope of this paper.

3.2. Acoustic analysis
The first instance of each modality is analyzed. Figures 1 & 2 show the acoustic and EGG signals (not discussed in this paper) of the the sung and spoken versions of “matushka” spoken two times. For singing and speaking, there is clear separation of the 3 syllables—“ma”, “tush” and “ka”.

![Fig. 1. “Matushka” sung twice. The acoustic wave form is shown in the top window and the EGG signal in bottom window.](image1)

![Fig. 2 “Matushka” spoken twice.](image2)
Figure 3 shows the lamented utterance, starting with a big inhalatory sigh followed by the utterance, “matushka tymoia”, and ending with (six) exhalatory sobs.

The acoustic measurements reported below are for the initial [ma] of the lament.

**F0.** As reported in previous studies, the voice pitch (fundamental frequency, F0) is higher for lamenting than for singing or speaking. Recent studies (e.g., [1]), also show that for active grieving while speaking, F0 is high, especially compared to a “neutral” non-emotional speaking voice. In this study, the F0 for lament was highest (416 Hz), then singing (328 Hz), and then speaking (264 Hz).

**Jitter and shimmer.** The speaking mode has the largest amount of jitter (1.85%), whereas lamenting (0.63 %) and singing (0.57%) are smaller. This suggests relatively little fluctuation in the tension of the vocal folds and vocal muscles for lamenting and singing. As for shimmer, the lament mode has the largest amount (9.6%), then speaking (7.74%), and then singing (3.72 %). A large amount of shimmer suggests instability in amplitude characteristics of the glottal cycle, having to do probably with air flow/air pressure instability, rather than instability in the tension of the vocal folds.

For the initial inhalatory sigh and the (next to) final exhalatory sigh in the lament, we see extremely high values of jitter (3.44% and 4.01%, respectively) and shimmer (10.74% and 9.8%, respectively), suggesting much fluctuation in the tension of the vocal folds per vibratory cycle as well as airflow fluctuation.

**Glottal opening size (H1-H2) & speed of glottal closing (H1-A3).** Speaking has a larger H1-H2 value (5.4 dB) than singing (4.8 dB), suggesting that speaking is more breathy than singing. Interestingly, lamenting has the smallest H1-H2 value (3.8 dB), suggesting that it is least breathy. For H1-A3, lament has a decidedly lower value (6.1 dB) than either singing (30.7 dB) or speaking (27.4 dB), suggesting that the speed of glottal closing is much faster for lament than for singing or speaking. The auditory impression of this sound is more tense.

For the inhalatory and exhalatory sighs, H1-H2 is 28.3 dB and 15.9 dB, respectively, and H1-A3 is 30.3 dB for both. All values are extremely large, indicating extreme breathiness, and contrasting with the tense voice quality of the lamented verbal material.

3.2. **High speed imaging**

Observation of the high speed imaging of the vocal folds suggests interesting differences in vocal fold vibration among the three modes, with lamenting showing a great deal more instability.

**Singing &speaking**

Regular vibration with complete closure, starting at the bottom and working up, can be seen for both singing (fig. 4) and speaking (not shown). For speaking, the folds are shorter and thicker, due to the lower F0.
Lamenting (figs. 5, 6, 7).

The initial inhalatory sigh (Fig. 5) is produced with the arytenoids (at the top of the image) stretched thinly and pulled back. This contrasts with singing and speaking, in which the arytenoids are thick and forward. The pattern of vibration during this initial inhalatory sigh is such that only the bottom part of the vocals folds close (the front of the throat). The top part (near the arytenoids) remains mostly open. Also, the bottom part tends to close somewhat regularly, while the top part closes incompletely and irregularly.

“Matushka (ty moia)” (fig. 6) is produced with irregular patterns of vibration. During [ma], the middle part of the folds close, but not the top and bottom. Also, there is asymmetry in the top part and incomplete closure with the right vocal fold (left in the image) moving more than the left one.

The voiceless consonants /t/, and /shk/ are produced with a period of no pulsing-- a narrow slit between the folds--followed by ripple pulsing as the vocal folds begin to vibrate. The pattern of vibration changes again, with the top part closing first, then the middle part. This is followed by something that appears to be “parallel vibration” (side to side symmetry)—both folds vibrate first to right then to left, but with no closure. This then changes to a pattern with gradually more opening at the top, with only the bottom part vibrating.

At the very end of the lamented utterance during the exhalatory sobs (fig. 7), we see a large opening, with the arytenoids opened wide (Fig. 7). The opening is followed by asymmetric vocal fold vibration, and then another big opening. The opening appears to be slightly heart-shaped. After each opening, the vocal cords appear shorter, vibrating with incomplete closure. This is repeated for each sob, and finally we see the arytenoids bend forward almost covering the glottis.
Throughout the various progressions of the lament, we observe a generally widened epilaryngeal area, compared with singing or speaking. Also we see relatively thin folds, with the false vocal folds retracted, resulting in an increase of the size of the sinus of Morgagni (e.g., the space between the true and false vocal folds), characteristic of crying. Lament shows a progression of variations of vocal fold tension, from very tense and thin (initial inhalatory sigh) to still tense and thin, but arytenoids less tense (linguistic portion of the lament), and finally to a wide open glottis, with folds still tense (final exhalatory sobs).

3.4. Correspondence between visual, acoustic and auditory observations.

In general, the voice quality of the linguistic portion of the lament is tense and instable, which may well be a result among other things of the observed tension of the vocal folds and arytenoids.

**F0.** For the high F0 during lamented /ma/, we observed longer thinner vocal folds than those for singing or speaking.

**Jitter.** For the low jitter values during lamented /ma/, we observed more tense vocal folds and arytenoids than for singing or speaking.

**Shimmer.** For the high shimmer values during lamented /ma/, we observed more aperiodicity in the vocal folds than for singing or speaking.

**H1-H2.** For the small values of H1-H2 during lamented /ma/, we seem to have observed less open vocal folds than for singing or speaking. They are especially less open than during the initial and final sighs of the lament.

**H1-A3.** For the small values of H1-A3 during lamented /ma/, we seem to have observed more rapid closing of the vocal folds than for singing or speaking. They are especially more rapid than during the initial and final sighs.

4. Conclusion

Previous studies reported a considerable difference in the acoustic and perceptual characteristics of lamenting, as compared with singing or speaking. These findings were confirmed by this study. Lamenting, compared with singing or speaking, is characterized by increased F0 along with instability of pitch and intensity, and is perceived by listeners, even those who do not know Russian, as extremely sad.

In addition, our study uncovered some things about vocal fold vibration that may account for features of lament reported earlier. During the verbal portion of the lament, the vocal folds are (1) longer and thinner, (2) more tense, as are the arytenoids also, and (3) the vibratory cycle less periodic with less complete closure than that seen for the sung or spoken versions. The higher F0 of laments can be explained by the longer and thinner vocal folds, the smaller H1-A3 value along with the auditory impression of tenseness can be explained by the more tense vocal folds, and the instability of pitch and especially of intensity (amplitude) can be explained by the peculiarities of the vibratory cycles observed in the high speed imagings.

The lack of control reported to be experienced by the lamenter is manifested, among other things by increased vocal shimmer and rapid, incomplete closure, along with a great deal of change in vocal
fold vibration pattern throughout the entire lament. More work is needed with more speakers to substantiate these findings.

Our future goal is to incorporate our findings to date into exploration of how brain activities are affected by perception and production of emotion in laments.

Acknowledgments
We thank Albert Rilliard for his help with the Runtime Revolution program for the perception tests, T. Oomae and students at Showa Music University for participation in the perception tests. In addition, we thank Kerrie Obert for her observations about vocal fold vibration during lament. This work was supported in part by Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010):19520371 and SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

References


